JUMS

Volume 11, Issue 3, September 2026

JUNIOR MANAGEMENT SCIENCE

Editor:
DOMINIK VAN AAKEN

Advisory Editorial Board:
FREDERIK AHLEMANN
JAN-PHILIPP AHRENS
THOMAS BAHLINGER
MARKUS BECKMANN
SULEIKA BORT
ROLF BRÜHL
KATRIN BURMEISTER-LAMP
CATHERINE CLEOPHAS
NILS CRASSELT
BENEDIKT DOWNAR
KERSTIN FEHRE
MATTHIAS FINK
DAVID FLORYSIAK
GUNTHER FRIEDL
MARTIN FRIESL
FRANZ FUERST
WOLFGANG GÜTTEL
NINA KATRIN HANSEN
ANNE KATARINA HEIDER
CHRISTIAN HOFMANN
SVEN HÖRNER
STEPHAN KAISER
NADINE KAMMERLANDER
ALFRED KIESER
ALEKSANDRA KLEIN
NATALIA KLIEWER
DODO ZU KNYPHAUSEN-AUFSESS
SABINE T. KÖSZEGI
ARJAN KOZICA
CHRISTIAN KOZIOL
MARTIN KREEB
WERNER KUNZ
HANS-ULRICH KÜPPER
MICHAEL MEYER
JÜRGEN MÜHLBACHER
GORDON MÜLLER-SEITZ
J. PETER MURMANN
ANDREAS OSTERMAIER
BURKHARD PEDELL
ARTHUR POSCH
MARCEL PROKOPCZUK
TANJA RABL
SASCHA RAITHEL
NICOLE RATZINGER-SAKEL
ASTRID REICHEL
KATJA ROST
THOMAS RUSSACK
FLORIAN SAHLING
MARKO SARSTEDT
ANDREAS G. SCHERER
STEFAN SCHMID
UTE SCHMIEL
CHRISTIAN SCHMITZ
MARTIN SCHNEIDER
MARKUS SCHOLZ
LARS SCHWEIZER
DAVID SEIDL
THORSTEN SELLHORN
STEFAN SEURING
VIOLETTA SPLITTER
ANDREAS SUCHANEK
TILL TALAULICAR
ANN TANK
ANJA TUSCHKE
MATTHIAS UHL
CHRISTINE VALLASTER
PATRICK VELTE
CHRISTIAN VÖGTLIN
BARBARA E. WEISSENBERGER
ISABELL M. WELPE
HANNES WINNER
THOMAS WRONA
THOMAS ZWICK

Jonas M. Berken, Nachhaltigkeitskennzahlen im Controlling: Wie Corporate Digital Responsibility messbar wird 470

Rozmari Genova, Steuerliche Analyse von Direktinvestitionen in Bulgarien im Vergleich zu einer Geschäftsausweitung in Deutschland für mittelständische Unternehmen 488

Paul Lehmann, CEO Narcissism and Stock Market Reactions: An Empirical Analysis 513

Simon E. Liedtke, Economic Policy Uncertainty and Investment: Cross-Country Evidence 532

Jan P. Tollkühn, Breaking the Glass Ceiling: The Effects on Perceived Equality of Women Joining the Management Board 551

Felix Maximilian Kania, Toward Reliable Emission Reporting: A Comparative Study of GHGRP Self-Reports and Climate TRACE Satellite Data 561

Roya Kazimova, Experimental Analysis of Gender Bias in Hiring within STEM Fields 598

Lars Kiehne, The Influence of Perceived Undervaluation on the Duration of Share Repurchase Programmes 611

Christoph Rehrl, Firm-wide Effects of Earnings Stripping Rules 642

Thomas Zimmermann, Der Einfluss unterschiedlicher Expertengruppen auf die ertragsteuerliche Gesetzgebung in Deutschland: Eine Analyse von schriftlichen Stellungnahmen 671

Published by Junior Management Science e.V.
This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.

ISSN: 2942-1861

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Nachhaltigkeitskennzahlen im Controlling: Wie Corporate Digital Responsibility messbar wird

Jonas Michael Berken

Heinrich-Heine-Universität Düsseldorf

Abstract

Digitalization is leading to profound changes in business models, companies, and entire industries. In light of these changes, the traditional understanding of corporate responsibility is increasingly being challenged. This is primarily due to the new, far-reaching, and partly unpredictable effects arising from the use of digital technologies by companies. Corporate Digital Responsibility (CDR) addresses these effects by extending corporate responsibility to areas such as data protection, digital diversity, and cybersecurity. Nevertheless, linking CDR with economic success remains a major challenge for companies. Controlling can make a significant contribution in this regard if performance in CDR-specific areas of responsibility can be measured through appropriate sustainability metrics. Against this backdrop, this thesis examines existing sustainability controlling instruments and metrics for CDR-specific performance measurement based on a literature review. Building on these findings, a CDR Balanced Scorecard is developed that illustrates cause-and-effect relationships between CDR and economic success and demonstrates a strategic management approach for steering CDR-specific areas of responsibility.

Zusammenfassung

Die Digitalisierung führt zu tiefgreifenden Veränderungen von Geschäftsmodellen, Unternehmen und ganzen Branchen. Im Lichte dieser Veränderungen stößt das herkömmliche Verständnis von Unternehmensverantwortung zunehmend an seine Grenzen. Grund hierfür sind die neuen, weitreichenden und teils unvorhersehbaren Auswirkungen, die vom unternehmerischen Einsatz digitaler Technologien ausgehen. Corporate Digital Responsibility (CDR) adressiert diese Auswirkungen und erweitert die Verantwortung von Unternehmen um Bereiche, wie etwa Datenschutz, digitale Vielfalt oder Cyber-Sicherheit. Gleichwohl stellt die Verknüpfung von CDR mit wirtschaftlichem Erfolg für Unternehmen eine große Herausforderung dar. Das Controlling kann hierzu einen wesentlichen Beitrag leisten, wenn es gelingt, Leistungen in CDR-spezifischen Verantwortungsbereichen mit geeigneten Nachhaltigkeitskennzahlen messbar zu machen. Ausgehend davon untersucht diese Arbeit auf Basis einer literaturgestützten Analyse bestehende Instrumente des Nachhaltigkeitscontrollings sowie Kennzahlen zur CDR-spezifischen Performance-Messung. Darauf aufbauend wird eine CDR-Balanced Scorecard entwickelt, die Ursache-Wirkungsbeziehungen zwischen CDR und wirtschaftlichem Erfolg darlegt und einen strategischen Ansatz zur Steuerung CDR-spezifischer Verantwortungsbereiche veranschaulicht.

Keywords: corporate digital responsibility; nachhaltigkeitscontrolling; sustainability balanced scorecard; nachhaltigkeitskennzahlen

1 Einleitung

Apple, Microsoft, Nvidia, Alphabet und Amazon sind, sofern die Marktkapitalisierung als Kriterium betrachtet wird, die fünf wertvollsten Unternehmen der Welt.[1] Und alle haben eines gemeinsam. Es handelt sich um Unternehmen mit digitalen Geschäftsmodellen. Im Zuge der Digitalisierung sind es solche Unternehmen, denen die World Bank Group (WBG) ein überdurchschnittlich gutes Wachstum, Innovationskraft und Millionen von neu geschaffenen Arbeitsplätzen attestiert (vgl. World Bank Group, 2024, S. XVIII).

Parallel dazu wächst das gesellschaftliche Interesse, dass Unternehmen ihrer Verantwortung gerecht werden und nachhaltige Geschäftsmodelle entwickeln (vgl. Schmidpeter, 2015, S. 1-20). Nach dem Brundtland-Bericht der World Commission on Environment and Development (WCED) ist eine nachhaltige Entwicklung dadurch gekennzeichnet, dass die Bedürfnisse der Gegenwart befriedigt werden, ohne künftige Generationen darin zu gefährden, ihre Bedürfnisse

Mein Dank gilt Frau Prof. Dr. Barbara Weißenberger und Herrn Armando Schrödter für die Betreuung dieser Arbeit sowie die sehr hilfreichen fachlichen Anregungen während des Entstehungsprozesses.

[1] Vgl. Companies Market Cap (2024): https://companiesmarketcap.com/, letzter Zugriff am 20.12.2024

DOI: https://doi.org/10.5282/jums/v11i3pp470-487

© The Author(s) 2026. Published by Junior Management Science.
This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.

zu befriedigen (vgl. World Commission on Environment and Development, 1987, S. 16). Für Unternehmen bedeutet dies, ihr Handeln nicht nur an finanziellen Kriterien auszurichten, sondern ebenso ökologische und soziale Aspekte einzubeziehen (vgl. Sailer, 2022, S. 18ff.). Auf dieser Grundlage können negative Externalitäten, also Auswirkungen unternehmerischen Handelns, die nachteilig für unbeteiligte Dritte sind und im Marktpreis keine Berücksichtigung finden, minimiert oder gänzlich vermieden werden (vgl. Dheur 2015, S. 346).

Die negativen Externalitäten, die Unternehmen durch den Einsatz digitaler Technologien erzeugen, sind neu, weitreichend und teils unvorhersehbar (vgl. Lobschat et al., 2021, S. 876f.). Exemplarisch hierfür sind soziale Netzwerke zu nennen, die in großem Umfang personenbezogene Daten sammeln und zu Werbezwecken unrechtmäßig an Dritte weitergeben (vgl. Lehmkemper, 2023, S. 19).

Angesichts dieser Externalitäten ist das herkömmliche Verständnis von Unternehmensverantwortung um eine Corporate Digital Responsibility (CDR) zu erweitern (vgl. Lautermann & Frick, 2023, S. 29). Ausgehend davon ist es im Interesse der Gesellschaft, dass Unternehmen den Einsatz digitaler Technologien, auch unter Einbezug ökologischer und sozialer Aspekte, nicht nur bewerten, sondern auch optimieren, um so nachhaltige Geschäftsmodelle zu entwickeln.

Das Controlling kann hierzu einen wesentlichen Beitrag leisten, indem es geeignete Nachhaltigkeitskennzahlen zur Messung von Corporate Digital Responsibility entwickelt. Es gilt als evident, dass Unternehmen, die ihr Geschäft zahlenmäßig erfassen und steuern, finanziell erfolgreicher sind (vgl. Lingle & Schiemann, 1996, S. 57-60). Analog ist es wünschenswert, das Gleiche mit Blick auf CDR und den Nachhaltigkeitserfolg zu erreichen.

Während Kennzahlen zur Messung des innerbetrieblichen Geschäfts schon heute ohne großen Aufwand aus IT-Systemen entnommen werden können, erweist sich die Entwicklung und Implementierung geeigneter Nachhaltigkeitskennzahlen noch als schwierig. Ursächlich hierfür ist, dass Nachhaltigkeit oft von weichen, qualitativen Faktoren geprägt ist, die mit Kennzahlen nicht eindeutig gemessen werden können (vgl. Colsman, 2016, S. 36-51). Hinzu kommt mit Blick auf CDR, dass digitale Technologien neue und teils unvorhersehbare Externalitäten erzeugen, zu deren Messung es in der Praxis bisher an geeigneten Nachhaltigkeitskennzahlen mangelt (vgl. Lobschat et al., 2021, S. 876f.)

Daher stellt sich die Frage, wie CDR im Rahmen des Controllings durch geeignete Kennzahlen messbar wird. Und wie CDR darüber hinaus in einen Steuerungsansatz integriert werden kann, der sich dadurch auszeichnet, schon vorlaufend diejenigen Leistungen zu messen und zu steuern, die die Ursache dafür sind, dass CDR im Ergebnis erst möglich wird. Diese Fragen zu beantworten, ist sogleich die Zielsetzung dieser Arbeit. Hierzu ist die im Folgenden beschriebene Vorgehensweise gewählt worden.

So wird zunächst im zweiten Kapitel untersucht, wie sich Unternehmen und deren Verantwortung im Zuge der Digitalisierung verändern und weshalb das herkömmliche Verständnis von Unternehmensverantwortung um CDR erweitert werden muss. Anschließend erfolgt im dritten Kapitel eine Betrachtung des heutigen Verständnisses von Controlling und Nachhaltigkeitscontrolling in Abgrenzung zum klassischen Controlling. Darauf aufbauend werden dann Instrumente des Nachhaltigkeitscontrollings - wie die Wesentlichkeitsanalyse und der Stakeholder-Dialog, aber auch Nachhaltigkeitskennzahlen und Indikatoren sowie die *Sustainability Balanced Scorecard* (SBSC) - beleuchtet, um deren Beitrag zur Lösung der Problemen in der Nachhaltigkeitssteuerung aufzuzeigen.

Vor dem Hintergrund der gewonnen Erkenntnisse wird im vierten Kapitel zunächst dargestellt, wie CDR in das Nachhaltigkeitscontrolling integriert werden kann und welche Anforderungen die praktische Controllerarbeit hierbei im Allgemeinen erfüllen muss. Ferner wird eine Vielzahl von Nachhaltigkeitskennzahlen aus Literatur, Nachhaltigkeitsberichten und Berichtsstandards auf ihre Eignung für das Nachhaltigkeitscontrolling geprüft und im Hinblick auf die Performance-Messung in CDR-spezifischen Verantwortungsbereichen optimiert. Abschließend wird gezeigt, wie diese Nachhaltigkeitskennzahlen in die vier Perspektiven einer Balanced Scorecard überführt werden können, sodass ein integrierter Steuerungsansatz in Form einer *CDR-Balanced Scorecard* entsteht, der CDR ganzheitlich mit einer Verbesserung des Finanzergebnisses verbindet.

2 Unternehmen und deren Verantwortung im Zuge der Digitalisierung

2.1 Einsatz und Auswirkungen digitaler Technologie im Unternehmenskontext

Zunächst bedarf es eines grundlegenden Verständnisses von Digitalisierung, digitaler Technologie und ihrer Auswirkungen auf Geschäftsmodelle, um anschließend die Unternehmensverantwortung im Zuge der Digitalisierung untersuchen zu können.

Hans Corsten und Stefan Roth definieren Digitalisierung als einen Prozess, der basierend auf Cyber-physischen Systemen sowie dem Internet der Dinge zur Erschließung von wertschöpfenden Diensten innerhalb von Ökosystemen führt (vgl. Corsten & Roth, 2022, S. 15). Digitale Technologien umfassen alle Technologien, welche hierzu das Erstellen, Verarbeiten, Übertragen und Nutzen von digitalen Gütern ermöglichen (vgl. Loebbecke, 2006, S. 360).

Um Cyber-physische Systeme besser zu verstehen, empfiehlt sich ein Blick auf den allgemeinen Systembegriff. Der Wirtschaftswissenschaftler Hans Ulrich beschreibt Systeme als "*eine geordnete Gesamtheit von Elementen, zwischen denen irgendwelche Beziehungen bestehen oder hergestellt werden können* (Ulrich, 1975, S.33)."

Hinsichtlich der Elemente eines Cyber-physischen Systems sind zunächst einmal die eingebetteten Systeme zu nennen. Diese eingebetteten Systeme dienen ausschließlich der Funktion des physischen Geräts, in das sie eingebettet wurden und bestehen aus einer Software sowie einem Prozessor

zur Ausführung dieser Software. Derartige eingebettete Systeme verhelfen Haushaltsgeräten, wie etwa Waschmaschinen zu ihrer Funktion. Cyber-physische Systeme entstehen, sobald eingebettete Systeme über das Internet der Dinge miteinander verbunden und so koordiniert zur Erfüllung eines übergeordneten Ziels in der realen Welt gesteuert werden. Infolgedessen entfällt bei Cyber-physischen Systemen die klare Systemgrenze, die in ein jeweiliges Gerät eingebettete Systeme auszeichnet. Das Internet der Dinge ist ein Netzwerk, welches diverse Geräte bzw. Maschinen über den Austausch von Daten in Echtzeit miteinander kommunizieren lässt (vgl. Koch & Grimm, 2022, S. 316-326).

Die Funktionalität von Cyber-physischen Systemen ist hingegen nicht auf Kommunikation beschränkt. Das Fraunhofer-Institut nennt in Bezug auf deren Funktionalität zunächst das Vorhandensein von Sensoren zur kognitiven Wahrnehmung der realen Welt. Software erfüllt die Funktion, das Wahrgenommene, also die in Form von Daten erfassten Informationen, auszuwerten und im Anschluss daran Entscheidungen zu treffen. Diese werden durch Aktoren zielgerichtet ausgeführt, ohne dass ein menschliches Eingreifen erforderlich ist. Digitale Informationstechnologie (IT) und physische Maschinen werden so zu einer symbiotischen Einheit. Aufgrund der Fähigkeit, ohne menschliches Eingreifen fundierte Entscheidungen zu treffen und auszuführen, spricht man auch von intelligenten Systemen (vgl. Drossel et al., 2018, S. 198-215).

Um zu veranschaulichen, was der Einsatz dieser Systeme im Unternehmenskontext bewirkt, eignet sich ein Blick auf die sogenannte “Industrie 4.0”. Der Begriff hat sich im Zuge der Digitalisierung inflationär verbreitet und steht für die vierte industrielle Revolution. Damit ist die zunehmende Vernetzung von industrieller Infrastruktur gemeint. All die am Wertschöpfungsprozess beteiligten intelligenten Maschinen bilden über das Internet kommunizierende Cyber-physische Systeme. Auf diese Weise entwickelt sich eine intelligente Automatisierung industrieller Infrastruktur, wodurch die Effizienz des Wertschöpfungsprozesses erheblich gesteigert werden kann (vgl. Obermaier, 2019, S. 3-43).

Überdies ermöglicht das Internet der Dinge die Zusammenführung von Daten, die Sensoren in zunehmender Menge in jeglichen Bereichen der Wertschöpfung generieren und den Unternehmen als hilfreiche Entscheidungsgrundlage dienen (vgl. Seufert & Oehler, 2016, S. 75-79). Datenbanken, wie z. B. “SAP HANA” speichern und verarbeiten die generierten Daten in Echtzeit, sodass Algorithmen, die sich statistischer Methodik bedienen, eine zielgerichtete Analyse der Daten vollziehen können. Die Entwicklung, dass die Datenmenge aufgrund steigender Kapazitäten von digitaler Informationstechnologie stetig zunimmt, wird terminologisch als “Big Data” bezeichnet (vgl. Van der Aalst, 2016, S. 3-13). Eine spezifische Form von Datenbank, die durch den Aufstieg von Kryptowährungen an Popularität gewonnen hat, ist die sogenannte Blockchain. Eine Blockchain speichert Daten anonym in chronologisch verknüpften Blöcken. Im Gegensatz zu traditionellen Datenbanken werden Daten dezentral auf allen beteiligten Rechnern synchronisiert gespeichert (vgl. Adam, 2022, S. 2-21).

Die Ressourcen für die Nutzung diverser digitaler Technologien werden durch Cloud- Computing bereitgestellt. Gemäß der anerkannten Definition einer Behörde aus den USA umfasst der Begriff Cloud-Computing alle Rechenkapazitäten, darunter Netze, Server, Speicherkapazität und weitere Dienste, die bedarfsgerecht bei minimalem Aufwand zur Verfügung gestellt werden können (vgl. Mell & Grance, 2011, S. 1f.).

Damit Software dazu befähigt wird, Aufgaben ohne menschliches Eingreifen zu erfüllen, muss sie mithilfe von großen Datenmengen lernen können. Wenn eine solche Lernfähigkeit attestiert ist, spricht man von einer Künstlichen Intelligenz (KI) (vgl. Buxmann & Schmidt, 2019, S. 7f.). Der Lernprozess, in dem Algorithmen darauf trainiert werden, bestimmte Zusammenhänge in einer großen Menge von Daten zu erkennen, wird “Maschinelles Lernen” genannt (vgl. Murphy, 2012, S. 1). Im Ergebnis soll KI repetitive Aufgaben erfüllen und große Mengen von Daten in kurzer Zeit analysieren können, wobei dem Menschen eine anleitende und überwachende Rolle zukommt (vgl. Leukert et al., 2019, S. 43f.).

Das Potenzial von digitaler Technologie im Unternehmenskontext zeigt sich vor allem mit Blick auf die Börse. Die wertvollsten börsennotierten Unternehmen verfolgen ein digitales Geschäftsmodell.[2] Nach Osterwalder und Pigneur beschreibt ein Geschäftsmodell die Art und Weise, in der Unternehmen Werte schaffen, vermitteln und erfassen und besteht aus den folgenden Elementen: Kundensegmente, Nutzenversprechen, Kanäle, Kundenbeziehungen, Einnahmequellen, Schlüsselressourcen, Schlüsselaktivitäten, Schlüsselpartnerschaften und Kostenstruktur (vgl. Osterwalder & Pigneuer, 2010, S. 14-23). Wenn diese Elemente im Wesentlichen auf digitaler Technologie basieren, liegt ein digitales Geschäftsmodell vor (vgl. Bozem & Nagl, 2021, S. 36).

Apple, Facebook und viele weitere Unternehmen mit digitalen Geschäftsmodellen zeichnen sich dadurch aus, dass sie über eine digitale Plattform verfügen. Eine solche digitale Plattform stellt die digitale Infrastruktur zur Verfügung, die nötig ist, um den Austausch von Informationen oder Services zwischen externen Parteien zu ermöglichen. Zusammengenommen bildet sich ein Ökosystem. Alle profitieren hierbei von niedrigen Transaktionskosten sowie einem Phänomen, das als Netzwerkeffekt bezeichnet wird. Je mehr externe Parteien zu Nutzern der Plattform werden, desto besser kann die Plattform adäquate Austauschbeziehungen herstellen. Dadurch wiederum wird die Plattform für ihre Nutzer immer wertvoller. Im Stil einer zirkulären, positiven Rückkopplungsschleife steigt stetig der Wert der Plattform bedingt durch die laufende Akquirierung neuer Nutzer. Auf diese Weise hat sich Facebook als soziales Netzwerk etablieren können, das Milliarden von Menschen auf der ganzen Welt miteinander verbindet. Apple wiederum erschuf über sein eigens ent-

[2] Vgl. Companies Market Cap (2024): https://companiesmarketcap.com/, letzter Zugriff am 20.12.2024

wickeltes Betriebssystem einen zweiseitigen Markt für den Austausch zwischen App-Entwicklern und Nutzern (vgl. Van Alstyne et al., 2016, S. 22-31).

Das Geschäftsmodell mit digitalen Plattformen unterscheidet sich fundamental von traditionellen Geschäftsmodellen, die prinzipiell darauf bedacht sind, eine effiziente Allokation von Ressourcen bei der Herstellung eines physischen Produkts zu ermöglichen. Stattdessen steht man als Plattformbetreiber in der Verantwortung für die strukturelle Entwicklung eines Ökosystems mit der Absicht, Netzwerkeffekte zu realisieren und Daten zu generieren. Die preisgegebenen Daten der Nutzer dienen den Plattformbetreibern als Grundlage dafür, wertschöpfende datengetriebene Services zu entwickeln und im Ökosystem zu etablieren (vgl. Pflaum & Klötzer, 2019, S. 65-71).

Das macht Daten zur wichtigsten Ressource von Plattformbetreibern. Traditionelle Geschäftsmodelle hingegen sind auf den Besitz von kapitalintensiver, physischer Infrastruktur angewiesen. Daher sind digitale Geschäftsmodelle im Vergleich wesentlich günstiger skalierbar und im Ergebnis profitabler (vgl. Van Alstyne et al., 2016, S. 22-31).

Aus diesem Grund ergänzen Industrieunternehmen ihr traditionelles Geschäftsmodell zunehmend durch digitale Geschäftsmodelle. Das Fraunhofer-Institut hat in einer Studie Industrieunternehmen untersucht und festgestellt, dass sich die Verfolgung digitaler Geschäftsmodelle, komplementär zum ursprünglichen Hauptgeschäftsmodell, positiv auf Produktivität und Erfolg auswirkt. Im Ergebnis kann dem Kunden über sogenannte produktbegleitende Dienstleistungen ein individueller Nutzen geboten werden, der über die eigentlichen Produktfunktionen hinausgeht (vgl. Lerch & Maloca, 2020, S. 2-9). Eine Studie des Massachusetts Institute of Technology (MIT) stellte fest, dass Geschäftsmodell-Innovationen branchenübergreifend wesentlich häufiger bei Unternehmen auftreten, die digitale Technologie nutzen und diese infolgedessen deutlich profitabler sind als ihre Konkurrenten (vgl. Westerman et al., 2012, S. 9-20).

Welchen Einfluss die Digitalisierung auf Branchen hat, die nicht primär für ihre Technologieorientierung bekannt sind, lässt sich an der Automobilindustrie beispielhaft zeigen. Das Konzept der *Software-Defined Vehicles* steht für die Vision von einem intelligenten Auto, dessen Komponenten, die ehemals als eingebettete Systeme agierten, zu einem vernetzten Cyber-physischen System werden. Über Software-Updates lassen sich so Fernreparaturen realisieren und neue Funktionen freischalten (vgl. Liu et al., 2022, S. 180ff.). Bosch, der größte Automobilzulieferer der Welt, prognostiziert, dass Autos auch nach der Fertigstellung durch Software-Updates während des gesamten Lebenszyklus verbessert werden können (vgl. Bosch, 2023).

Resümierend lässt sich festhalten, dass sich Geschäftsmodelle, Unternehmen und ganze Branchen im Zuge der Digitalisierung strukturell verändern. Traditionelle Geschäftsmodelle, die ihren Wert aus der Herstellung physischer Produkte schöpfen, werden zunehmend von digitalen Geschäftsmodellen mit datengetriebenen Services verdrängt (vgl. Van Alstyne et al. 2018, S. 22-31). Um die Innovationskraft digitaler Technologie terminologisch zu erfassen, ist in Anlehnung an Clayton Christensen auch von “digitaler Disruption” die Rede (vgl. Matzler et al., 2016, S. 81ff.). Die Digitalisierung führt im Unternehmenskontext zu einer “digitalen Transformation”, in der Unternehmen die Art und Weise ihrer bisherigen Wertschöpfung hinterfragen müssen, um langfristig wettbewerbsfähig zu bleiben (vgl. Vial, 2019, S. 15-19; Westerman et al., 2012, S. 9-20).

2.2 CDR als Erweiterung von CSR

Im Folgenden wird untersucht, was Corporate Social Responsibility (CSR) bedeutet und warum CSR im Zuge der digitalen Transformation um digitale Verantwortung zu erweitern ist. Wenn Unternehmensverantwortung Gegenstand des öffentlichen Diskurses wird, ist zumeist von CSR die Rede. Ferner wird der Begriff Nachhaltigkeit eingebracht, der im Unternehmenskontext untrennbar mit CSR zusammenwächst und daher in dieser Arbeit synonym verwendet wird (vgl. Schneider, 2015, S. 27-30).

Gleichwohl hat die Wissenschaft den Begriff CSR bislang nicht einheitlich definieren können (vgl. Crane et al., 2013, S. 3f.). Um trotzdem eine klare Grundlage für die fortlaufende Untersuchung zu schaffen, wird die weit verbreitete Definition der Europäischen Kommission aus dem Jahr 2011 verwendet. Gemäß der Europäischen Kommission bedeutet CSR, dass Unternehmen in Absprache mit ihren Stakeholdern soziale, ökologische und ethische Aspekte in ihre Strategie integrieren, um auch über rechtliche Pflichten hinaus ihrer Verantwortung vollumfänglich gerecht werden zu können. Das übergeordnete Ziel soll darin bestehen, negative Externalitäten zu minimieren und den *Shared Value* zu maximieren (vgl. Europäische Kommission, 2011, S. 6).

Das Konzept des *Shared Value* ist im Jahr 2006 von Michael Porter und Mark Kramer vorgestellt worden und sieht vor, dass Unternehmen einen Mehrwert schaffen, der sowohl Shareholdern als auch der Gesellschaft zugutekommt. Es adressiert den bestehenden Interessenskonflikt zwischen CSR im traditionellen Sinne und dem Gewinnanspruch von Unternehmen. Porter und Kramer vertreten die Ansicht, dass das Interesse der Gesellschaft an Unternehmensverantwortung nicht zwingend dem Gewinnanspruch der Unternehmen zuwider ist. Stattdessen wird von einer wechselseitigen Abhängigkeit ausgegangen. Unternehmen können demnach auch wirtschaftlich davon profitieren, wenn sie ihrer Verantwortung gerecht werden und damit einhergehend negative Externalitäten ihres Handelns minimieren. Ein effizienter Gebrauch limitierter Ressourcen ist beispielsweise nicht nur ökologisch sinnvoll, sondern auch kostensparend (vgl. Kramer/Porter 2011, S. 145ff.).

Bereits 2001 veröffentlichte die Europäische Kommission eine Definition von CSR. In ihrer ursprünglichen Definition hieß es lediglich, dass CSR ein Konzept zur freiwilligen Integration sozialer und ökologischer Aspekte sei (vgl. Europäische Kommission, 2001, S. 6). Der bestehende Interessenskonflikt zwischen Unternehmen und Gesellschaft blieb damals also gänzlich unberücksichtigt. Das stieß auf Kritik. In-

folgedessen revidierte die Europäische Kommission im Jahr 2011 ihre ursprüngliche Definition und bezog das zwischenzeitlich vorgestellte Konzept des *Shared Value* mit ein. Dadurch wird deutlich, dass die Umsetzung von CSR auch im Einklang mit dem Gewinnanspruch eines Unternehmens gelingt (vgl. Wojcik, 2016, S. 39-42).

Unternehmen, die ihrer Verantwortung nicht gerecht werden, riskieren ggf. den Verlust ihrer *License to Operate*, also der von der Gesellschaft gewährten Möglichkeit zur wirtschaftlichen Betätigung (vgl. Deegan, 2006; Weißenberger und Marrocco, 2022, S. 43). Demgegenüber kann sich die Übernahme von Verantwortung langfristig positiv auf die Kunden- und Mitarbeiterzufriedenheit oder die Reputation eines Unternehmens auswirken, was indirekt auch dem finanziellen Erfolg zugutekommt (vgl. Lai et al., 2010, S. 460-466).

Indem Unternehmen die Maximierung des *Shared Value* zum strategischen Ziel erklären, können derartige Risiken minimiert, Chancen realisiert und CSR erfolgreich umgesetzt werden. Allerdings hängt das laut Porter und Kramer auch von Produktinnovationen und einer optimierten Wertschöpfung ab (vgl. Kramer/Porter 2011, S. 157ff.).

Die Digitalisierung bietet Chancen dazu. Im Unternehmenskontext führt sie ohnehin zu mehr Wertschöpfung, Geschäftsmodell-Innovationen und disruptiven Technologien (vgl. Matzler et al., 2016, S. 81-87; Vial, 2019, S. 15-19; Van Alstyne et al., 2016, S. 22-31). Carsharing-Dienste, die als spezifische Form der “Sharing-Economy” gelten, bieten über digitale Plattformen eine bedarfsgerechte, ressourcensparende und zugleich flexible Mobilität (vgl. Baedeker et al., 2018, S. 6f.). Die intelligente Vernetzung und Automatisierung industrieller Infrastruktur, das Schlüsselelement der Industrie 4.0, kann die Effizienz steigern und ökologische Ressourcen innerhalb des Wertschöpfungsprozesses einsparen (vgl. Obermaier, 2019, S. 43; Kirmes, 2017, S. 14ff.). Software-Updates verbessern Autos auch nach ihrer Fertigstellung, was zu mehr Komfort und Nachhaltigkeit beiträgt (vgl. Bosch, 2023).

Andererseits sind durch den unternehmerischen Einsatz digitaler Technologie bisher in erster Linie neue, weitreichende negative Externalitäten präsent geworden. Viele Unternehmen greifen bei der Verfolgung ihres digitalen Geschäftsmodells in die Privatsphäre von Nutzerinnen und Nutzern ein, indem sie deren personenbezogene Daten zu Werbezwecken verarbeiten und an Dritte weitergeben (vgl. Lehmkemper, 2023, S. 19). Digitale Endgeräte werden des Öfteren bereits nach einer kurzen Gebrauchsdauer stillgelegt und verschrottet, was zu einem erhöhten Ressourcenverbrauch führt (vgl. Köhler et al., 2018, S. 66). Eine zunehmend digitalisierte Arbeitswelt erfordert lebenslanges Lernen, damit Beschäftigte, deren Berufsbilder verschwinden, neue Tätigkeiten ausüben können (vgl. IHK Berlin, 2017, S. 4f.). Der Ausbau von Cloud-Computing und dazu benötigten Daten- und Rechenzentren führt zu einem erheblichen Energiebedarf, der durch Effizienzgewinne bislang nur eingeschränkt kompensiert werden kann (vgl. Hintemann, 2018).

Damit Unternehmen digitale Technologie verantwortungsvoll einsetzen und die sich bietenden Chancen zur Maximierung des *Shared Value* nutzen, ist es erforderlich, die herkömmliche CSR um digitale Verantwortung, also Corporate Digital Responsibility (CDR) zu erweitern (vgl. Lautermann & Frick, 2023, S. 29). Digitale Geschäftsmodelle gewinnen im Zuge der Digitalisierung an Bedeutung und erzeugen neue, teils unvorhersehbare Externalitäten. Das Konzept der CDR verfolgt die Absicht, Unternehmen, die digitale Technologie nutzen bzw. entwickeln, frühzeitig für derartige Externalitäten zu sensibilisieren (vgl. Lobschat et al., 2021, S. 876f.).

Speziell das Attribut der Frühzeitigkeit gewinnt angesichts der Verbreitung agiler Entwicklungsmethoden an Bedeutung. Innerhalb kürzester Zeit werden neue Produkte iterativ, in Absprache mit dem Kunden entwickelt und verbessert, wobei vor allem sichtbare Produkteigenschaften, also technische Funktionen und Benutzerfreundlichkeit im Vordergrund stehen. Negative Externalitäten, zu deren Kompensation noch in der Entwick-lung Gegenmaßnahmen erforderlich gewesen wären, werden oft erst nach der Markteinführung sichtbar (vgl. Spiekermann, 2016, S. 163-166; Kenning und Weißenberger, 2019).

Digitale Verantwortung ist wie CSR bislang nicht einheitlich definiert worden. In einer Initiative vom Bundesministerium der Justiz und für Verbraucherschutz (BMJV) heißt es, CDR ist freiwillig und Bestandteil einer umfassenden Unternehmensverantwortung, die darauf abzielt, die digitale Transformation nachhaltig und zum Vorteil der Gesellschaft zu gestalten (vgl. BMJV, 2021, S. 2). Obwohl zu CDR bislang auch keine explizite Berichtspflicht besteht, sind unter den Europäischen Nachhaltigkeitsberichtsstandards, die zum 01.01.2024 in Kraft getreten sind, auch Themen mit CDR-Bezug, wie etwa Datenschutz (vgl. Europäische Union, 2024, S. 28).

Nach Barbara Weißenberger und Armando Marrocco lässt sich die Nachhaltigkeitsstrategie eines Unternehmens im Kontext von CDR in folgende vier Handlungsfelder segmentieren: Finanzen, Umwelt, Soziales und Governance (vgl. Weißenberger & Marrocco, 2022, S.45f.). Über das erstgenannte Handlungsfeld wird auch das Interesse der Shareholder an finanziellen Erfolgen anerkannt. Die drei letztgenannten Handlungsfelder bilden zusammen die ESG-Kriterien, die sich als Konzept zur Bewertung des Nachhaltigkeitserfolgs von Unternehmen etabliert haben (vgl. Helmold et al., 2024, S. 8).

Angesichts des Stellenwerts der digitalen Transformation und der sich bietenden Chancen auf nachhaltige Produktinnovationen und Wettbewerbsvorteile ist es sinnvoll, digitale Verantwortung fest in die Unternehmensstrategie zu integrieren. Folglich ist von einer Umsetzung lediglich auf Ebene des Compliance-Managements abzuraten (vgl. Weißenberger & Marrocco, 2022, S. 46ff.). Was das betrifft, so besteht Grund zur Skepsis, da eine auf Regeleinhaltung reduzierte Denkweise der Idee eines strategischen Steuerungsansatzes, der finanziellen Erfolg mit Nachhaltigkeit verbindet, eher entgegensteht (vgl. Reimer & Schäffer, 2024, S. 9).

Daher gilt für die Integration von CDR in das Nachhaltigkeitscontrolling, die im vierten Kapitel zum Thema wird, dass digitale Verantwortung ein fester Bestandteil der Unternehmensstrategie sein muss.

Zur inhaltlichen Konkretisierung der vier oben genannten CDR-Handlungsfelder eignen sich die von Saskia Dörr definierten CDR-spezifischen Verantwortungsbereiche (vgl. Weißenberger & Marrocco, 2022, S.45f.). Anhand dieser Verantwortungsbereiche können Unternehmen in Zusammenarbeit mit ihren Stakeholdern prüfen, wo Handlungsbedarf besteht, Ziele festlegen und adäquate Maßnahmen ergreifen. Dörr definiert insgesamt 15 verschiedene CDR-spezifische Verantwortungsbereiche. Im Hinblick auf eine praxeologisch leichtere Handhabbarkeit wird diese Anzahl auf sieben Verantwortungsbereiche verdichtet, die im Folgenden den vier CDR-Handlungsfeldern zugeordnet sind:

- Handlungsfeld Umwelt: Digitaler Ressourcenverbrauch, Digitaler CO2-Fußabdruck;
- Handlungsfeld Soziales: Digitale Vielfalt, Digitale Arbeitswelt;
- Handlungsfeld Governance: Datenschutz, Cyber-Sicherheit;
- Handlungsfeld Finanzen: Entwicklung innovativer CDR-basierter Geschäftsmodelle (vgl. Weißenberger und Marrocco, 2022, S.46; Dörr, 2020, S. 94-97; Kempkes et al., 2021, S. 31ff.).

Vor diesem Hintergrund werden im Weiteren diese sieben CDR-spezifischen Verantwortungsbereiche behandelt und messbar gemacht.

3 Nachhaltigkeitscontrolling

3.1 Abgrenzung zum klassischen Controlling

Um die Integration von CDR in das Nachhaltigkeitscontrolling, die vierten Kapitel behandelt wird, auf eine fundierte wissenschaftliche Basis zu stellen, werden zunächst das heutige Verständnis von Controlling und Nachhaltigkeitscontrolling in Abgrenzung zum klassischen Controlling beleuchtet.

Gemäß der Grundsatzposition der International Group of Controlling (IGC) und des Internationalen Controller Vereins (ICV) ist das Controlling eine Führungsarbeit, die im Zusammenwirken von Managern und Controllern stattfindet. Folglich ist die Ausübung dieser Tätigkeit nicht allein den Controllern vorbehalten. Dem Management obliegt die Unternehmensführung, da es die Entscheidungen trifft. Controller sind mitverantwortlich, da sie das Management zu einer rationalen Unternehmensführung befähigen, indem sie es mit entscheidungs- und steuerungsrelevanten Informationen versorgen. Das “Controller-Leitbild” von IGC und ICV besagt, dass Controller und Manager in den Managementprozess, bestehend aus Zielfindung, Planung und Steuerung, partnerschaftlich involviert sind (vgl. Bieniek et al., 2023, S. 2-7).

Das von Barbara Weißenberger und der IGC entwickelte *House of Controlling* greift dieses Leitbild auf und bietet einen hilfreichen Überblick zur praktischen Arbeit von Controllern. Da im vierten Kapitel das *House of Controlling* dazu dient, die Anforderungen an die Controllerarbeit zur Integration von CDR in das Nachhaltigkeitscontrolling zu analysieren, wird es im Folgenden kurz vorgestellt und behandelt. Es segmentiert die Controllerarbeit in originäre und derivative Aktionsfelder. Zu den originären Aktionsfeldern gehören die Bereiche Planung, Berichtswesen und Performance-Messung, die zusammen den etablierten Steuerungskreislauf des Controllings bilden (vgl. IGC & Weißenberger, 2006, S. 20). Das Gelingen der Unternehmenssteuerung, die terminologisch synonym zum Controlling verwendet wird, hängt in erster Linie davon ab, dass diese drei Bereiche effektiv ineinandergreifen (vgl. Weber & Schäffer, 2022, S. 34).

Zu den derivativen Aktionsfeldern gehören die Organisation der Controllerarbeit und die Gestaltung von Vorsystemen, die eine Steuerung erst ermöglichen (vgl. IGC & Weißenberger, 2006, S. 20). Bei diesen Vorsystemen handelt es sich heutzutage vor allem um IT Systeme, die große Datenmengen in Echtzeit verarbeiten und speichern können (vgl. Schön, 2022, S. 351-382).

Im Planungsprozess werden alle verfügbaren Informationen, die sich auf die Zukunft beziehen und demnach planungsrelevant sind, verarbeitet, um Ziele zu definieren sowie die Mittel zu ihrer Erreichung. Je nach Ausmaß und Zeithorizont des Geplanten wird zwi-schen strategischer und operativer Planung differenziert (vgl. Horvath et al., 2023, S. 71f.). Das Berichtswesen dient der Informationsversorgung, also der Übermittlung unternehmensinterner Informationen (vgl. Weber & Schäffer, 2022, S. 233-237).

Unter Performance-Messung versteht man das Entwickeln und Einsetzen von geeigneten Kennzahlen zur Bewertung von Leistungen in diversen Organisationseinheiten eines Unternehmens, um entscheidungs- und steuerungsrelevante Informationen gewinnen und daraus Vorschläge für adäquate Maßnahmen ableiten zu können. Das dient der Unternehmenssteuerung insofern, als dass eine konkrete Zielformulierung und -erreichung in allen Bereichen eines Unternehmens sichergestellt werden kann (vgl. Gleich, 2021, S. 17ff.).

Ursprünglich ist die Unternehmenssteuerung mit rein monetären Größen betrieben worden. Das klassische Controlling stammt aus dem internen Rechnungswesen und verstand sich vorwiegend als Dienstleistungsbereich, der finanzielle Zahlen aus der Buchhaltung aufbereitet und der Unternehmensführung zur Verfügung stellt. Daher entstand das Rollenbild des Controllers als “Erbsenzähler”. Dieses Rollenbild erscheint heute hingegen, aufgrund höherer Anforderungen und der engen Verzahnung von Unternehmensführung und Controlling obsolet (vgl. Becker & Ulrich, 2016, S.3ff.). Die Entwicklung begann in den 1990er-Jahren, in denen der Controller anfing, sich von der vergangenheitsorientierten Buchhaltung zu lösen. Stattdessen rückte

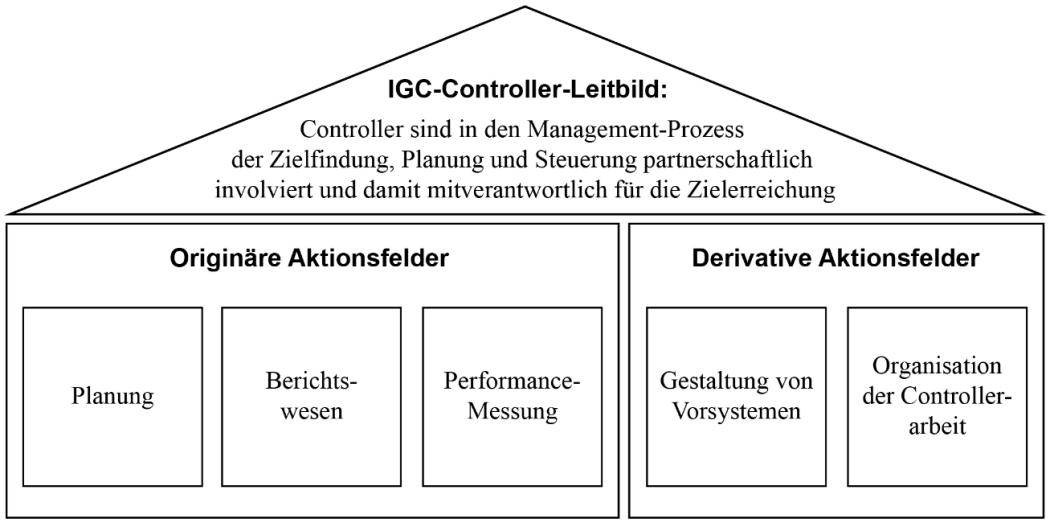


Abbildung 1: House of Controlling

(Quelle: leicht verändert entnommen aus IGC und Weißenberger, 2006, S. 20)

die Messung nicht-finanzieller Kennzahlen mehr in den Vordergrund (vgl. Horvath et al., 2023, S.20).

Folglich kommen in der Performance-Messung nach heutigem Verständnis mehrdimensionale Kennzahlen zum Einsatz, die neben finanzieller Ergebnistransparenz auch Leistungstransparenz schaffen. So können durch nicht-finanzielle Kennzahlen auch die Leistungen innerhalb der Geschäftstätigkeit selbst, die sich nicht oder nur eingeschränkt monetarisieren lassen und aus denen letztlich das finanzielle Ergebnis hervorgeht, bewertet werden. Das befähigt den Controller zur Generierung entscheidungs- und steuerungsrelevanter Informationen mit Zukunftsbezug (vgl. Gleich, 2021, S. 3ff.). Im Zuge dessen ist der Controller in die strategische Planung und Entscheidungsfindung involviert worden und hat sich zum "Business Partner" der Unternehmensführung entwickelt (vgl. Horvath et al. 2020, S. 23f.).

Mit Blick auf den finanziellen Erfolg von Unternehmen zeigt sich, dass diejenigen, die ihre Geschäftstätigkeit mit nicht-finanziellen Kennzahlen erfassen und steuern, bessere Ergebnisse erzielen (vgl. Lingle & Schiemann, 1996, S. 57-60). Analog ist es angesichts des zunehmenden Interesses der Gesellschaft an Unternehmensverantwortung wünschenswert (vgl. Schmidpeter, 2015, S. 1-20), das Gleiche im Hinblick auf den Nachhaltigkeitserfolg zu erreichen.

Das Nachhaltigkeitscontrolling knüpft an die Abkehr von rein monetären Größen an und macht einen weiteren Entwicklungsschritt, da eine triviale Berücksichtigung von betriebswirtschaftlichen Größen nicht-finanzieller Dimension nicht mehr genügt. Die Ausweitung der Unternehmenmenssteuerung um Nachhaltigkeitsaspekte führt dazu, dass die bisherige Handlungsmaxime des Controllers "*What you cannot measure, you cannot manage*" überdacht werden muss. Stattdessen gilt es, auch Faktoren zu berücksichtigen, die weich, qualitativ und nur indirekt messbar sind, wie es bei vielen Nachhaltigkeitsaspekten der Fall ist (vgl. Colsman, 2016, S. 44ff.).

So kann der individuelle Gewinnanspruch eines Unternehmens um die ESG-Kriterien ergänzt und die Maximierung des *Shared Value*, also eines Mehrwerts für Unternehmen und Gesellschaft gemeinsam angesteuert werden (vgl. Schmidpeter und Günther, 2013, S. 27; Porter/Kramer 2011, S. 145ff.).

Probleme des Nachhaltigkeitscontrollings sind vor allem eingeschränkte Messbarkeit und unklare Ursache-Wirkungszusammenhänge, die zu scheinbar existierenden Zielkonflikten führen (vgl. Colsman, 2016, S. 26-42).

Um Zielkonflikte lösen zu können, ist der Dialog mit den Stakeholdern unerlässlich. Er ist elementar für das Instrument der Wesentlichkeitsanalyse, das Nachhaltigkeitsaspekte, zu denen Stakeholder ihr Interesse bekundet haben, auf Wesentlichkeit für die externe Berichterstattung prüft. Im Nachhaltigkeitscontrolling ist die Wesentlichkeitsanalyse relevant, da sie eine Voraussetzung dafür ist, dass in der strategischen Planung geschäftsspezifisch die Nachhaltigkeitsaspekte priorisiert werden können, die wesentlich sind (vgl. Sailer, 2022, S. 121f.). Zudem kann der Stakeholder-Dialog bei der Bewertung von Handlungsoptionen in Anbetracht von Realisierbarkeit und Akzeptanz helfen (vgl. Colsman, 2016, S. 60).

Falls zu einem wesentlichen Nachhaltigkeitsaspekt ein Zielkonflikt besteht, lautet die Entscheidungsregel, das Interesse der Stakeholder, das bei Erfüllung etwa in Form von Mitarbeiterzufriedenheit langfristig zum Unternehmenserfolg beiträgt, vor das Interesse der Shareholder an kurzfristigen Profiten zu stellen. Controller können sich rationalitätssichernd in den Planungsprozess einbringen, indem sie das Management bei etwaigen Abwägungen mit entscheidungsrelevanten Informationen versorgen, die den langfristigen Beitrag von Nachhaltigkeit zum Unternehmenserfolg darlegen. Im Detail bedeutet das, soweit möglich,

die Kalkulation von Kosten- und Risikoeffekten eines nicht nachhaltigen Handelns, aber auch die Kalkulation von Kosten-, Reputations- und Umsatzeffekten nachhaltigen Handelns (vgl. Reimer & Schäffer, 2024, S. 9-13).

Damit das gelingt, muss der Controller Nachhaltigkeit vollumfänglich in seine originären Aktionsfelder, also den Steuerungskreislauf des Controllings integrieren. Ebenso ist eine adäquate Gestaltung der Vorsysteme erforderlich, damit die Erhebung und Analyse von Daten in den ESG-Handlungsfeldern realisiert werden kann (vgl. Colsman, 2016, S. 36-47; IGC und Weißenberger, 2006, S. 20).

3.2 Nachhaltigkeitskennzahlen und Sustainability Balanced Scorecard

Wie oben erwähnt, sind eine eingeschränkte Messbarkeit und unklare Ursache-Wirkungszusammenhänge Gründe dafür, dass scheinbar existierende Zielkonflikte bestehen, die Unternehmen davon abhalten, nachhaltig zu agieren. Durch die Entwicklung geeigneter Nachhaltigkeitskennzahlen und einer Sustainability Balanced Scorecard können Controller diese Gründe beseitigen und dem Management rationalitätssichernd zur Seite stehen.

Mit Kennzahlen werden betriebswirtschaftliche Sachverhalte in einer komplexen Welt zahlenmäßig und damit in konzentrierter Form erfasst, damit anschließend schnell und prägnant berichtet werden kann (vgl. Weber und Schäffer, 2022, S. 179; Reichmann et al., 2017, S. 39).

Vor diesem Hintergrund lassen sich Kennzahlen nach Art und Funktion unterscheiden. Nach der Art der Kennzahlen kann differenziert werden in finanzielle und nicht-finanzielle Kennzahlen. Finanzielle Kennzahlen berichten über das Finanzergebnis und die finanzielle Lage eines Unternehmens. Durch die Ergänzung finanzieller Kennzahlen um nicht-finanzielle Kennzahlen lässt sich geschäftsspezifisch vermitteln, wie das Finanzergebnis zustande kommt und wie es positiv beeinflusst werden kann. Daher wird auch zwischen Ergebniskennzahlen und vorlaufenden Kennzahlen unterschieden, mithilfe derer ein Unternehmen proaktiv gesteuert werden kann. Ferner wird die Art der Kennzahlen über die statistische Form unterschieden in relative und absolute Kennzahlen, wobei relativen Kennzahlen aufgrund ihrer Bezugsgröße eine höhere Aussagekraft attestiert wird (vgl. Weber & Schäffer, 2022, S. 179-199).

Darüber hinaus können Kennzahlen nach den Funktionen unterschieden werden, die sie im Rahmen des Controllings erfüllen. Im Zuge ihrer Vorgaben- und Operationalisierungsfunktion werden Kennzahlen dazu verwendet, Ziele in einen konkreten Zahlenwert zu übersetzen und den Erfüllungsgrad von Zielen messbar zu machen. Ausgehend von ihrer Anregungsfunktion sensibilisieren Kennzahlen die Verantwortlichen für Auffällig-keiten und regen ein bestimmtes Verhalten an, das dem Erreichen des definierten Ziels dienen muss. Durch ihre Steuerungsfunktion können Kennzahlen die komplexe vorlaufende Geschäftstätigkeit vereinfacht wiedergeben und auf potenzielle Stellhebel zur Verbesserung des Finanzergebnisses aufmerksam machen (vgl. Weber & Schäffer, 2022, S. 181). In ihrer Kontrollfunktion ermöglichen Kennzahlen die Durchführung von Abweichungsanalysen, wobei Planwerte mit ermittelten Ist-Werten verglichen und dabei identifizierte Abweichungen auf ihre Ursachen untersucht werden (vgl. Gleich, 2021, S. 27f.).

Für die Performance-Messung von Nachhaltigkeit innerhalb der ESG-Handlungsfelder sind vor allem Kennzahlen nicht-finanzieller Art erforderlich. Eine hilfreiche Orientierung zur Auswahl geeigneter Nachhaltigkeitskennzahlen bieten die Richtlinien der Global Reporting Initiative (GRI), nach denen die meisten Unternehmen ihre Nachhaltigkeitsberichte erstellen (vgl. Sailer, 2022, S. 219). Die GRI-Standards enthalten Nachhaltigkeitskennzahlen zu Themen, wie etwa Materialien, Energie, Beschäftigung, Diversität oder auch Datenschutz und adressieren damit alle ESG-Handlungsfelder (vgl. GRI, 2024, S. 3f.).

Allerdings handelt es sich bei den zu Berichtszwecken erhobenen Daten häufig um einfache Durchschnittsgrößen, die nicht präzise genug sind, um als Kennzahl im Controlling eine Steuerungsfunktion zu erfüllen. Damit die Genauigkeit den Anforderungen des Controllings genügt, gilt es, in Abhängigkeit von der Steuerungsrelevanz zusätzlich Daten zu erheben, um Materialverbräuche, Emissionen oder Krankenstände differenziert nach einzelnen Standorten, Bereichen oder Kostenstellen messen zu können (vgl. Colsman, 2016, S. 51-54).

Die fortschreitende Erfassung und Vernetzung von Daten über Sensoren in jeglichen Bereichen der Wertschöpfung schafft neue Informationsquellen. Der Controller kann sich diese Datenquellen erschließen, indem er die Nutzung von Tools, die auch unter dem Begriff *Business Intelligence* (BI) bekannt sind, als Methodenkompetenz seiner Arbeit versteht (vgl. Seufert & Oehler, 2016, S. 75-79). Bezeichnend hierfür ist der Einsatz von KI, der es ermöglicht, eine Vielzahl von Datenquellen - auch solche, die speziell für die ESG-Handlungsfelder relevant sind - zu berücksichtigen, um ausdifferenzierte Nachhaltigkeitskennzahlen mit Steuerungsfunktion nutzen zu können (vgl. Weißenberger, 2021, S. 13).

Im Zuge der Integration von Nachhaltigkeit in die Unternehmenssteuerung ist es erforderlich, dass Controller sich mit weichen Faktoren befassen, diese messen und den positiven Einfluss von Nachhaltigkeit auf das Finanzergebnis darlegen. Wenn ein Faktor weich, qualitativ und nur indirekt messbar ist, können Indikatoren stellvertretend aushelfen. Diese sind weniger valide, aber leichter zu messen und ein guter Anhaltspunkt zur Einschätzung des eigentlich relevanten Aspekts. Zum Beispiel können Unternehmen zur Messung der Artenvielfalt in ihrer natürlichen Umgebung auf Indikatoren zurückgreifen, die über Schadstoffbelastung oder Ausgaben für Luftfilteranlagen informieren (vgl. Gladen, 2014, S. 10f. Küpper et al., 2024, S. 422). Zur Messung sozialer Faktoren, wie beispielsweise Mitarbeiter- und Kundenzufriedenheit, Arbeitsklima oder Qualifizierungsbedarf, können Indikatoren aus Befragungen gewonnen werden (vgl. Kobi, 2021, S. 368-371). Auf dieser Grundlage lassen sich auch weiche Faktoren in die Performance-Messung integrieren.

Des Öfteren sind es unklare Ursache-Wirkungsbeziehun-

gen bzw. scheinbar existierende Zielkonflikte, die sich kurzfristig nicht beseitigen lassen und das Management daran hindern, nachhaltig zu agieren (vgl. Colsman, 2016, S. 26-42). In Anbetracht dessen erfüllen Indikatoren auch eine Anregungsfunktion, indem sie die langfristigen Auswirkungen von nachhaltigkeitsbezogenen Maßnahmen auf das Finanzergebnis eher widerspiegeln als rein finanzielle Kennzahlen. So wird das Management dazu angeregt, nachhaltig zu agieren (vgl. Merchant/Van der Stede 2017, S. 443).

Ferner bietet sich das Kennzahlensystem der *Sustainability Balanced Scorecard* (SBSC) an, um Klarheit zu schaffen, was die Ursache-Wirkungsbeziehungen zwischen Nachhaltigkeitszielen und Finanzergebnis betrifft. Dabei handelt es sich um eine Weiterentwicklung der *Balanced Scorecard* (BSC), die von Robert Kaplan und David Norton in den 1990er Jahren entwickelt wurde (vgl. Sailer, 2022, S. 167f.).

Die BSC ist ein Planungs- und Steuerungsinstrument, das finanzielle Kennzahlen durch nicht-finanzielle Kennzahlen und Indikatoren ergänzt, um neben Ergebnistransparenz auch Leistungstransparenz zu schaffen. Im Rahmen der BSC wird das Geschäft eines Unternehmens in gleichrangige, interdependente Perspektiven unterteilt, für die strategische Ziele, Zielwerte, Kennzahlen und Maßnahmen definiert werden. Im klassischen Sinne sind das die Perspektiven Finanzen, Kunden, interne Geschäftsprozesse sowie Lernen und Entwicklung. Da keine dieser Perspektiven isoliert bearbeitet wird, ist es üblich, diese in Form einer *Strategy Map* anzuordnen, wodurch die Perspektiven bzw. die darin festgelegten Ziele als Ursache-Wirkungsketten abgebildet werden können. Die Finanzperspektive steht dabei hierarchisch ganz oben und stellt das Finanzergebnis dar, das durch die geschäftsspezifisch festgelegten Ziele und Maßnahmen in den übrigen vorlaufenden Perspektiven positiv beeinflusst werden soll (vgl. Kaplan & Norton, 2024, S. 7-30).

Die SBSC integriert Nachhaltigkeit in die BSC und kann dadurch dabei unterstützen, Zielkonflikte zu lösen und Nachhaltigkeitsziele zu identifizieren, die Treiber für ein langfristig besseres Finanzergebnis sind (vgl. Hansen/Schaltegger 2014, S. 205). Dabei lassen sich zwei Ansätze zur Integration von Nachhaltigkeit in die BSC unterscheiden. Zum einen kann Nachhaltigkeit additiv über eine fünfte Perspektive integriert werden. Das ist vor allem dann sinnvoll, wenn ein Unternehmen gerade erst damit beginnt, sich parallel zu finanziellen Zielen auch um die Erfüllung von Nachhaltigkeitszielen zu bemühen. Zum anderen kann Nachhaltigkeit integrativ in allen vier Perspektiven der BSC berücksichtigt werden. Sobald ein Unternehmen mit der Steuerung von Nachhaltigkeit vertraut ist, empfiehlt sich dieser integrative Ansatz, um die Idee eines strategischen Steuerungsansatzes, der finanziellen Erfolg mit Nachhaltigkeit verbindet, konsequent zu verwirklichen (vgl. Sailer, 2022, S. 167ff.).

4 Integration von CDR in das Nachhaltigkeitscontrolling

4.1 Anforderungen an die praktische Controllerarbeit

Das Nachhaltigkeitscontrolling steht dem Management rationalitätssichernd zur Seite, was die Übernahme von Verantwortung betrifft. Damit das auch im Zuge der digitalen Transformation gelingt, ist CDR in das Nachhaltigkeitscontrolling zu integrieren. Die damit einhergehenden Anforderungen an die praktische Controllerarbeit sind Thema dieses Kapitels. Zur Modellierung der praktischen Controllerarbeit dienen die Aktionsfelder des *House of Controlling*, die im vorherigen Kapitel erläutert wurden.

Wenn ein Unternehmen beabsichtigt, CDR in die Unternehmensstrategie zu integrieren, und mit der Planung zur Verwirklichung dieser Idee beginnt, dann ist es im ersten Schritt sinnvoll, mit Stakeholdern in den Dialog zu treten (vgl. Sailer, 2022, S. 121f.).

Im Hinblick auf CDR sind die Interessen eines Stakeholders dann wesentlich, wenn dieser in besonderem Maße von den Externalitäten der digitalen Technologien betroffen ist, die ein Unternehmen entwickelt oder einsetzt. Dazu zählen auch die Interessen von Stakeholdern, die nur indirekt betroffen sind, wie etwa nachgelagerte Kunden. Denn auch solche verfügen bei Missachtung ihrer Interessen über die Macht, den langfristigen Erfolg eines Unternehmens negativ zu beeinflussen (vgl. Spiekermann, 2016, S. 170ff.). Demnach ist es ratsam, bei der Identifikation wesentlicher Interessen umsichtig zu sein. Das Controlling ist dazu aufgefordert, mithilfe einer Wesentlichkeitsanalyse zu unterstützen.

Als Anhaltspunkt für einen wirksamen Stakeholder-Dialog zur Durchführung einer Wesentlichkeitsanalyse sind die von Saskia Dörr definierten CDR-spezifischen Verantwortungsbereiche gut, da sie eine große Vielfalt von Externalitäten berücksichtigen. Ausgehend davon können systematische Befragungen der Stakeholder stattfinden (vgl. Dörr, 2020, S. 94-97). Darauf aufbauend können Controller die strategische Bedeutung von CDR für ihr Geschäft analysieren und in Zusammenarbeit mit dem Management festlegen, welche Bereiche wirklich wesentlich sind und daher im weiteren Zielfindungs- und Planungsprozess priorisiert werden sollten.

Die Rolle von digitaler Technologie in Unternehmen variiert je nach Geschäft. Für ein rein digitales Geschäftsmodell, das datengetriebene Services anbietet, ist digitale Technologie unerlässlich. Demgegenüber stehen traditionelle Geschäftsmodelle aus der Industrie, in denen digitale Technologie tendenziell eher eine ergänzende Rolle einnimmt (vgl. Van Alstyne et al., 2016, S. 22-31; Lerch und Maloca, 2020, S. 2-9). Welche Bereiche von CDR wesentlich sind und daher priorisiert werden sollten, wird demnach von Unternehmen zu Unternehmen variieren.

Gleichwohl ist digitale Technologie mit steigender Tendenz in nahezu allen Branchen bedeutsam und ein entscheidender Wettbewerbsfaktor, wie eine Studie des MIT belegt (vgl. Westerman et al., 2012, S. 9-20). Folglich nimmt auch die strategische Bedeutung von CDR zu, weshalb es

aus Unternehmenssicht ratsam ist, sich mit einer Vielzahl von CDR-spezifischen Verantwortungsbereichen dezidiert auseinanderzusetzen.

Die Wissenschaftliche Hochschule für Unternehmensführung (WHU) hat in einer Studie die Kompetenzen untersucht, die Controller brauchen, um der Digitalisierung gewachsen zu sein. Vor allem ein Wissen um digitale Technologie und Trends sowie die Erfolgsfaktoren digitaler Geschäftsmodelle soll zukünftig zu den wichtigsten Kompetenzen des Controllers gehören (vgl. Schäffer & Brückner, 2019). Analog gilt dies auch für CDR, um die Externalitäten digitaler Technologien im Hinblick auf das eigene Geschäft und im Interesse der Stakeholder verstehen zu können. Erst auf dieser Grundlage gelingt eine sachgerechte Informationsversorgung, über die Controller die Rolle des Business Partners einnehmen und das Management bei Zielfindung, Planung und Steuerung von CDR unterstützen können.

Ferner geht es auch darum, über die klassische Performance-Messung hinaus Akzente zu setzen. Digitale Geschäftsmodelle sind oftmals weniger auf eine effiziente Ressourcenallokation angewiesen (vgl. Pflaum & Klötzer, 2019, S. 65-71). Zwar sollte eine effiziente Allokation, insbesondere vor dem Hintergrund von Nachhaltigkeit, weiterhin akribisch verfolgt werden. Aber darüber hinaus ist auch ein rationales Denken in digitalen Geschäftsmodellen erforderlich (vgl. Schönbohm & Egle, 2021, S. 212).

Im Zuge der digitalen Transformation werden traditionelle Geschäftsmodelle, die ehemals sehr erfolgreich waren, hinterfragt, ergänzt oder verdrängt durch digitale Geschäftsmodelle (vgl. Van Alstyne et al. 2018, S. 22-31). Eine Vielzahl dieser digitalen Geschäftsmodelle, z. B. das der "Sharing-Economy" oder die "On-Demand-Fertigung" in der In-dustrie, beruhen auf digitaler Technologie und haben das Potenzial, zu mehr Ressour-ceneffizienz und Nachhaltigkeit beizutragen (vgl. Dörr, 2020, S. 152-156).

In ihrer Rolle als "Change Agent" sind Controller dazu berufen, das Management bei der strategischen Planung und Durchführung von Geschäftsmodell-Innovationen zu begleiten, indem sie Risiken in Form von hohen Investitionen und Chancen in Form von Erfolgspotenzialen kalkulieren (vgl. Schönbohm & Egle, 2021, S. 205-213). Hier kann das Nachhaltigkeitscontrolling ansetzen, indem es zudem das Nachhaltigkeitspotenzial digitaler Geschäftsmodelle kalkuliert. Auf diese Weise kann sich die Prophezeiung von Michael Porter und Mark Kramer erfüllen, die Produktinnovationen und eine optimierte Wertschöpfung als Schlüssel für die Maximierung des Shared Value betrachten (vgl. Porter/Kramer 2011, S. 157ff.).

Auch für die Produktentwicklung ist ein gut funktionierender Stakeholder-Dialog elementar, da Unternehmen ihre Handlungsoptionen so frühzeitig auf Realisierbarkeit und Akzeptanz prüfen können (vgl. Driessen/Hildebrand 2012, S. 365-374; Colsman, 2016, S. 60). Wie bereits im zweiten Kapitel festgestellt werden konnte, ist das Attribut der Frühzeitigkeit angesichts der zunehmenden Verbreitung agiler Entwicklungsmethoden von entscheidender Bedeutung. Neue Produkte werden innerhalb kürzester Zeit iterativ in Absprache mit dem Kunden entwickelt, ohne das negative Externalitäten, die im späteren Gebrauch entstehen und dann kaum mehr reparabel sind, vorab thematisiert und präventiv beseitigt werden. Zwar können so spezifische Funktionen möglichst benutzerfreundlich entwickelt werden (vgl. Spiekermann, 2016, S. 163-166; Kenning und Weißenberger, 2019). Jedoch fehlt den Teams in der Produktentwicklung häufig die Zeit, um sich mit CDR-spezifischen Verantwortungsbereichen zu befassen, denen das spätere Produkt im Interesse der Stakeholder gerecht werden muss (vgl. Spiekermann, 2016, S. 172).

Bei maschinellen Lernverfahren werden daher beispielsweise vermehrt Datensätze verwendet, die Ungleichheiten enthalten, wodurch die Algorithmen, die damit trainiert werden, diese Ungleichheiten anschließend reproduzieren. Das kann zu einer algorithmischen Entscheidungsfindung führen, die bestimmte Personengruppen systematisch benachteiligt (vgl. Köchling et al., 2020, S. 54).

Um negativen Externalitäten vorzubeugen, ist es erforderlich, bereits in der Produktentwicklung Maßnahmen zu ergreifen. Es ist ratsam, sich von vornherein seiner CDR-spezifischen Verantwortung bewusst zu sein und im Dialog mit den Stakeholdern zu klären, welche Verantwortungsbereiche im Hinblick auf denkbare produktspezifische Externalitäten wesentlich sind. Im Anschluss sollten die Stakeholder auch in alle weiteren Phasen der Produktentwicklung einbezogen werden (vgl. Spiekermann, 2016, S. 170-174).

Innerhalb eines solchen Dialogs kann es jedoch schnell zu Konflikten kommen, die dem Interesse des Managements an einer strikten Einhaltung des Budgets, einem unterschiedlichen Verständnis von Moral oder unterschiedlichen Interessen unter den Stakeholdern geschuldet sind. Darum muss eine rationale Entscheidungsfindung sichergestellt werden (vgl. Spiekermann, 2016, S. 175). Controller könnten den Dialog rationalitätssichernd begleiten, indem sie, soweit möglich, denkbare Externalitäten sowie Kosten und Wirksamkeit von Handlungsoptionen prognostizieren und dadurch eine objektive Entscheidungsgrundlage schaffen. Dazu muss das Controlling mit dem Tempo agiler Entwicklungsmethoden Schritt halten und sich ebenso agil organisieren.

Im Zuge der Digitalisierung befinden sich Unternehmen in einem Wettbewerbsumfeld, das an Dynamik und Geschwindigkeit gewonnen hat. In der Konsequenz muss die Informationsversorgung an Agilität gewinnen, um Unternehmen in einem solchen Wettbewerbsumfeld erfolgreich steuern zu können. Dafür ist auch der Einsatz von BI-Tools als Vorsysteme der Controllerarbeit unverzichtbar, um Entlastung zu schaffen, Informationsquellen schneller erschließen und eine gute Entscheidungsgrundlage schaffen zu können (vgl. Seufert und Oehler, 2016, S. 75-81; Weißenberger, 2021, S. 13).

Ferner ist die Einhaltung einer strikten Budgetvorgabe zu hinterfragen, die im klassischen Controlling zu Beginn eines Entwicklungsprojekts gemacht wird, aber häufig zu Konflikten führt, wenn neue oder sich ändernde Interessen der Stakeholder zu einer Budgetüberschreitung führen (vgl. Spie-

kermann, 2016, S. 164; Bogsnes, 2023, S. 176). Eine strikte Budgetierung widerspricht auch dem Grundgedanken einer agilen Produktentwicklung, der eine kontinuierliche, in kurzen Iterationen stattfindende flexible Entwicklung und Bereitstellung von Funktionen vorsieht (vgl. Bogsnes, 2023, S. 176).

Die Entwicklung eines neuen Produkts ist ein Prozess, der von hoher Unsicherheit geprägt ist, weshalb Erfolge nur schwer planbar sind. Eine strikte Budgetvorgabe beruht jedoch fälschlicherweise auf der Annahme, dass die Zukunft vorhersehbar, also der Entwicklungsprozess eben doch planbar ist (vgl. Bogsnes, 2023, S. 174ff.). Außerdem kann eine strikte Budgetvorgabe in der Produktentwicklung dazu verleiten, kurzfristige Profite gegenüber langfristigen Investitionen vorzuziehen (vgl. Andersen & Börsch, 2016, S. 59).

Agile Entwicklungsmethoden tragen dem Rechnung, aber eine erfolgreiche Umsetzung setzt ein Budget voraus, dass sich bedarfsgerecht an sich ändernde Anforderungen anpasst, die im Verlauf des Entwicklungsprozesses ersichtlich werden (vgl. Bogsnes, 2023, S. 176).

Damit eine agile Entwicklung digitaler Produkte gelingt, die im Interesse der Stakeholder auch nachhaltig sind, dürfen kurzfristige Zusatzkosten zur Prävention negativer Externalitäten nicht vermieden werden (vgl. Spiekermann, 2016, S. 175). Vor diesem Hintergrund sollte man sich an den Prinzipien des sogenannten *Beyond Budgeting* orientieren, die unter anderem für die Abkehr von strikten Budgetvorgaben und eine flexible bedarfsgerechte Ressourcenallokation stehen. Und das heißt, anstatt kurzfristige Profite zu verfolgen, stärker einem langfristigen Zweck dienen zu wollen (vgl. Bogsnes, 2023, 54f.).

Mit Blick auf CDR besteht dieser langfristige Zweck darin, digitale Technologie verantwortungsvoll einzusetzen, kurzfristige Zusatzkosten nicht zu scheuen und die sich bietenden Chancen zur langfristigen Maximierung des *Shared Value* zu nutzen. Die *CDR-Balanced Scorecard*, die in Kapitel 4.3 genauer erläutert wird, stellt CDR als Werttreiber zur Erreichung dieses Zwecks dar und bietet damit Anregungen für einen Steuerungsansatz, der finanziellen Erfolg mit CDR verbindet.

4.2 Performance-Messung in CDR-spezifischen Verantwortungsbereichen

Sobald die strategische Bedeutung von CDR analysiert und strategische Ziele gefunden sind, ist der Controller im Rahmen der Performance-Messung dazu aufgefordert, adäquate Kennzahlen zu entwickeln und einzusetzen. Das heißt, auch und vor allem die vorlaufenden Leistungen, die das Erreichen von anvisierten CDR-Zielen erst möglich machen, in allen relevanten Bereichen eines Unternehmens mess- und steuerbar zu machen.

Hierzu werden im Folgenden diverse Nachhaltigkeitskennzahlen aus Literatur, Nachhaltigkeitsberichten und Berichtsstandards, die den im zweiten Kapitel festgelegten CDR-spezifischen Verantwortungsbereichen zugeordnet werden können, vorgestellt und auf ihre Eignung für das Nachhaltigkeitscontrolling geprüft. Darauf aufbauend werden fundierte Vorschläge zur Optimierung dieser Kennzahlen für eine Performance-Messung von CDR gemacht. Diesbezüglich gilt, dass die Auswahl geeigneter Kennzahlen in der Praxis stets vom jeweiligen Anwendungskontext abhängt und dieses Kapitel lediglich ein Ansatzpunkt für eine breite Vielfalt von Unternehmen und deren Bereichen ist.

Zunächst soll es - mit Blick auf die drei ESG-Handlungsfelder - um das erste Handlungsfeld, das der Umwelt, und die dazu festgelegten CDR-spezifischen Verantwortungsbereiche “Digitaler Ressourcenverbrauch” und “Digitaler CO2-Fußabdruck” gehen. Zu diesen Bereichen messen und berichten Unternehmen bereits vielfach, jedoch selten aus einer CDR-spezifischen Perspektive (vgl. Kempkes et al., 2021, S. 31).

Die zur Nutzung digitaler Technologie benötigten Rechenkapazitäten werden von Daten- und Rechenzentren zur Verfügung gestellt (vgl. Mell & Grance, 2011, S. 1f.). Eine solche IT-Infrastruktur hat jedoch einen hohen Energiebedarf und verursacht, je nach Zusammensetzung des Strommix, einen erheblichen CO2-Fußabdruck (vgl. Bahr et al., 2023, S. 7-11).

Aufgrund der Expansion von KI-Anwendungen ist zu befürchten, dass Energiebedarf und CO2-Fußabdruck zukünftig weiter wachsen. Die Energieeffizienz der Rechenzentren zu steigern, ist ein Stellhebel, um dem entgegenzuwirken. Sie kann über die *Power-Usage-Effectiveness* (PUE) gemessen werden, eine Kennzahl, die den Energiebedarf des gesamten Rechenzentrums mit der Menge an Energie ins Verhältnis setzt, die von der IT des Rechenzentrums tatsächlich benötigt wird (vgl. Hintemann, 2018). Das Softwareunternehmen SAP hat die PUE in seinem Nachhaltigkeitsbericht als standortübergreifende Durchschnittsgröße veröffentlicht. Zur Steigerung der Energieeffizienz seiner Rechenzentren setzt SAP auf ein intelligentes Energiemanagementsystem. Außerdem kommt in der deutschen Firmenzentrale ein System zur Kühlung zum Einsatz, das die Verwendung überschüssiger Energie ermöglicht (vgl. SAP, 2023, S. 128).

Nach der DIN EN 50600-2-2 des Deutschen Instituts für Normung kann die PUE, je nach Anzahl und Platzierung von Messpunkten innerhalb eines Rechenzentrums, auf drei sogenannten Granularitätsniveaus gemessen werden. Das gewählte Niveau entscheidet darüber, ob damit genau, also auch an einzelnen Teilsystemen oder Komponenten gemessen wird, um Einsparpotenziale bestmöglich erkennen zu können (vgl. Bahr et al., 2023, S. 9). Zu Berichtszwecken ist es ausreichend, die PUE als standortübergreifende Durchschnittsgröße zu messen. Damit die PUE im Controlling eine Steuerungsfunktion erfüllen kann und die Wirksamkeit etwaiger Maßnahmen oder Einsparpotenziale erkannt wird, ist es sinnvoll, an allen Standorten mit einer Vielzahl von Messpunkten genau zu messen.

Das Öko-Institut hat vor dem Hintergrund von Energie- und Ressourcenverbräuchen im Zuge der Digitalisierung eine Vielzahl von Studien ausgewertet und stellt fest, dass vor allem die kurze Gebrauchsdauer digitaler Endgeräte ein Problem darstellt. Ein Phänomen, das auch als “digitale Obsoleszenz” bezeichnet wird und in einem erhöhten Bedarf an

Ressourcen sowie Aufkommen an Elektroschrott resultiert (vgl. Köhler et al., 2018, S. 66). Und das greift sogleich den CDR-spezifischen Verantwortungsbereich "Digitaler Ressourcenverbrauch" auf, der im Weiteren behandelt werden soll.

Das Konzept der Kreislaufwirtschaft adressiert dieses Phänomen, indem es darauf abzielt, gebrauchte Produkte und deren Bestandteile kostengünstig und energieeffizient in Material- und Wirtschaftskreisläufe zurückzuführen (vgl. Wilts & Berg, 2017, S. 135f.). Zwecks dessen bietet die Deutsche Telekom ihren Kundinnen und Kunden die Rücknahme gebrauchter Endgeräte an, um diese anschließend ordnungsgemäß zu recyceln oder, falls möglich, wiederzuverkaufen. Die Wirksamkeit dieser Maßnahme wird über die Rücklaufquote gemessen, eine relative Kennzahl, die darstellt, wie viele der insgesamt in Umlauf gebrachten Geräte anteilig zurückgenommen werden konnten (vgl. Deutsche Telekom, 2024, S. 16).

Falls ein Wiederverkauf nicht in Frage kommt, braucht es ein Recyclingsystem, dass die Rückgewinnung von Metallen und seltenen Erden aus ausrangierten Endgeräten ermöglicht, damit diese für die Herstellung von Neuprodukten wiederverwertet werden können (vgl. Häusler & Prinzler, 2020, S. 12) Um zu überprüfen, inwieweit das gelingt, misst das Technologieunternehmen Apple, differenziert nach Produkt, Bauteil und Materialart, den Anteil recycelter Materialien in Neuprodukten (vgl. Apple, 2023, S. 36ff.). So werden nicht nur die Ergebnisse eines nachhaltigen Recyclingsystems ersichtlich, sondern es wird auch dazu angeregt, mehr recyceltes Material für die Herstellung von Neuprodukten zu verwenden.

Eine detaillierte Dokumentation von Materialien, die in Neuprodukten verbaut werden, ist entscheidend, um das generelle Problem des Recyclings - den Mangel an Informationen über die materielle Beschaffenheit von Produkten - zu lösen. Technologien der Industrie 4.0, also Cyber-physische Systeme und ihre Sensorik, können die benötigten Daten bereitstellen und so Transparenz über produktbezogene Materialverbräuche schaffen, um Recycling und Wiederverwertung zu verbessern. Das kommt letztlich nicht nur der Umwelt, sondern auch dem Finanzergebnis zugute, da recycelte Materialien häufig günstiger sind als eine Beschaffung neuer Primärmaterialien (vgl. Wilts & Berg, 2017, S. 137f.).

Eine Vielzahl der Rohstoffe, die in IT-Hardware zum Einsatz kommen, sind toxisch und nur schwer recycelbar. Daher ist es ratsam, den Rohstoffeinsatz vorab strategisch und nachhaltigkeitsbewusst zu planen (vgl. Köhler et al., 2018, S. 66f.). Das Nachhaltigkeitscontrolling könnte die Recyclingfähigkeit eingesetzter Materialien mithilfe von Indikatoren bewerten, die Kosten, Energieeffizienz und Praktikabilität des späteren Recyclings präventiv einschätzen.

Um die Wirksamkeit etwaiger Maßnahmen zur Einsparung von Energie und Ressourcen und den damit einhergehenden ökologischen Zusatznutzen den Kunden bzw. Stakeholdern gegenüber transparent zu machen, empfiehlt sich eine Lebenszyklusanalyse. So können sämtliche Energie- und Ressourcenverbräuche erfasst, der Absatzmenge gegenübergestellt und zur Berechnung eines daraus resultierenden ökologischen Fußabdrucks - einschließlich der CO2-Emissionen - aggregiert werden (vgl. Sailer, 2022, S. 210ff.).

Im nächsten Schritt werden das zweite ESG-Handlungsfeld, das des Sozialen, sowie die dazugehörigen CDR-spezifischen Verantwortungsbereiche "Digitale Vielfalt" und "Digitale Arbeitswelt" betrachtet.

Die digitale Transformation stellt Unternehmen und Beschäftigte vor Herausforderungen. In einer zunehmend digitalen Arbeitswelt ist lebenslanges Lernen entscheidend, damit Beschäftigte, deren Berufsbilder durch die intelligente Automatisierung von Prozessen obsolet werden, neue Tätigkeiten übernehmen können. Vor allem ältere Mitarbeiter, die häufig weniger Erfahrung im Umgang mit digitalen Technologien haben, sind auf ein umfassendes Aus- und Weiterbildungsangebot angewiesen. Unternehmen wiederum sind auf Personal mit digitalen Kompetenzen angewiesen, um nachhaltig wettbewerbsfähig sein zu können (vgl. IHK Berlin, 2017).

Ein CDR-spezifisches Aus- und Weiterbildungsprogramm, das auf den individuellen Personalbedarf eines Unternehmens zur nachhaltigen Bewältigung der digitalen Transformation abgestimmt ist, könnte gezielt Abhilfe schaffen. So werden Mitarbeiterinnen und Mitarbeiter beim digitalen Kompetenzaufbau gefördert und zugleich auf die Relevanz von digitaler Verantwortung aufmerksam gemacht. Auf dieser Grundlage kann die digitale Transformation, gemäß der CDR-Definition des BMJV, zum Vorteil der Gesellschaft gestaltet werden (vgl. BMJV, 2021, S. 2).

Um Transparenz in der Zielerreichung zu schaffen, können Indikatoren mithilfe von Befragungen erhoben werden. Mercedes-Benz befragt die Mitarbeiterinnen und Mitarbeiter z. B. danach, wie zufrieden sie mit der Aus- und Weiterbildung sind (vgl. Mercedes Benz Group, 2024, S. 132). Nach dem GRI-Berichtsstandard zum Thema Aus- und Weiterbildung messen Unternehmen die durchschnittliche Anzahl der Stunden, die Mitarbeiter innerhalb eines Geschäftsjahres mit Aus- und Weiterbildung verbringen (vgl. GRI, 2024, S. 905f.). Allein mit Durchschnittsgrößen ist eine gezielte Steuerung des Aufbaus digitaler Kompetenzen jedoch nicht möglich. Daher muss präziser, also differenziert nach einzelnen Kompetenzbereichen, wie etwa Maschinelles Lernen oder Cyber-Sicherheit, gemessen werden. Um am Ende eines Geschäftsjahres die erzielten Fortschritte transparent zu machen, sollte zudem die Anzahl derer, die Aus- und Weiterbildungsprogramme erfolgreich absolviert haben, dokumentiert werden.

Mit Blick auf den Verantwortungsbereich "Digitale Vielfalt" ist ein besonderes Augenmerk auf KI zu legen. Wie bereits in Kapitel 4.2 erwähnt, kommen bei maschinellen Lernverfahren oftmals Datensätze zum Einsatz, die Ungleichheiten enthalten, wodurch die Algorithmen, die damit trainieren, solche Ungleichheiten reproduzieren. Dies kann zu einer algorithmischen Entscheidungsfindung führen, die verzerrt und systematisch nachteilig für bestimmte Personengruppen ist (vgl. Köchling et al., 2020, S. 54). Daher prüft die Deutsche Telekom vor der praktischen Anwendung, ob ein KI-Algorithmus fair agiert (vgl. Deutsche Telekom, 2023, S. 11).

Die Definition von Fairness hängt hierbei vom Anwendungskontext, also den individuellen Vorstellungen des Anwenders sowie der betroffenen Stakeholder ab. Auf dieser Grundlage können Anforderungen und Fehlertoleranzen festgelegt werden, die in einer anschließenden Testphase als Metriken zur Messung der Fairness verwendet werden können. So kann anhand messbarer Größen bewertet werden, ob die Entscheidungsfindung der KI den Anforderungen genügt und eingesetzt werden kann (vgl. Strasser & Niedermayer, 2021, S. 127-133).

Da Verzerrungen jedoch nie gänzlich ausgeschlossen werden können und die auf diese Weise gemessene Fairness lediglich als subjektiver Indikator dient, der weder objektiv definiert noch vollständig erreicht werden kann, sollten Beschwerden, die auf unfaire KI-Entscheidungen zurückzuführen sind, zahlenmäßig erfasst werden (vgl. Dörr, 2020, S. 198). Diese Beschwerden können dann systematisch untersucht werden, um potenzielle Verzerrungen zu identifizieren und die Fairness weiter zu verbessern (vgl. Strasser & Niedermayer, 2021, S. 127-133).

Das dritte ESG-Handlungsfeld, das der Governance umfasst die CDR-spezifischen Verantwortungsbereiche “Datenschutz” und “Cyber-Sicherheit”.

Wie eine Umfrage der Abteilung Wirtschafts- und Sozialpolitik der Friedrich-Ebert-Stiftung zeigt, sind es im Hinblick auf CDR diese beiden Bereiche, die aus Sicht der Gesellschaft besonders relevant sind (vgl. Thorun et al., 2018, S. 3). Einerseits steht der Schutz personenbezogener Daten im Mittelpunkt, andererseits gewinnt angesichts der zunehmenden Anzahl von Cyberangriffen die Sicherstellung einer robusten und sicheren IT-Infrastruktur immer mehr an Bedeutung (vgl. Kempkes et al., 2021, S. 34). Dass eine solche Infrastruktur auch personenbezogene Daten vor unbefugtem Zugriff schützt, zeigt wie beide Verantwortungsbereiche - Datenschutz und Cyber-Sicherheit - eng miteinander verknüpft sind.

Der GRI-Berichtsstandard zum Thema Datenschutz verlangt von Unternehmen, die Gesamtzahl der begründeten Beschwerden über Datenschutzverletzungen zu erfassen. Darüber hinaus soll auch die Anzahl der Cyberangriffe, die zu Verlusten von Kundendaten geführt haben, dokumentiert werden (vgl. GRI, 2024, S. 1060). Beide sind Ergebniskennzahlen, welche widerspiegeln, ob Maßnahmen zum Schutz von Daten und IT-Infrastruktur erfolgreich waren. Jedoch braucht es auch vorlaufende Kennzahlen, mithilfe derer Datenschutz und Cyber-Sicherheit proaktiv gesteuert werden können.

Die Deutsche Telekom befragt daher alle zwei Jahre ihre 50.000 Mitarbeiter stichprobenartig, wie es um das Datenschutz- und Sicherheitsbewusstsein im Unternehmen bestellt ist (vgl. Deutsche Telekom, 2024, S. 101). Ferner können fachspezifische Indikatoren herangezogen werden, mithilfe derer die Leistungsfähigkeit eines Security Operations Center (SOC) bewertet wird. Ein SOC dient in Unternehmen als zentrale Einheit zur Überwachung und Abwehr von Cyberangriffen. Zu den relevanten Indikatoren gehören etwa die Anzahl erfolgreich abgewehrter Angriffe sowie die der identifizierten Schwachstellen innerhalb der IT-Infrastruktur (vgl. Onwubiko & Onwubiko, 2019, S. 3). Auf dieser Grundlage wird ersichtlich, welche vorlaufenden Leistungen zu den oben beschriebenen Ergebniskennzahlen geführt haben. Ferner erfüllt der letztgenannte Indikator eine wichtige Anregungsfunktion, indem er die Mitarbeiter dazu motiviert, die IT-Infrastruktur eines Unternehmens präventiv zu schützen und somit das Risiko erfolgreicher Cyberangriffe bereits im Vorhinein weitestgehend auszuschließen.

Nachdem sich die Performance-Messung hier bislang ausschließlich auf die ESG-Handlungsfelder bezog, soll es im Folgenden um ein viertes Handlungsfeld, das der Finanzen, und den dazu festgelegten CDR-spezifischen Verantwortungsbereich “Entwicklung innovativer CDR-basierter Geschäftsmodelle” gehen.

Das zentrale Element eines Geschäftsmodells ist das Wertangebot, das auf den Produkten basiert, die ein Unternehmen seinen Kunden anbietet. Das Wertangebot kann innovativ und damit neu für den Kunden sein. Andererseits kann es den bestehenden Angeboten auf dem Markt auch sehr ähneln (vgl. Osterwalder & Pigneuer, 2010, S. 26). Aufgrund der Innovationskraft digitaler Technologien soll der Fokus auf der Entwicklung eines Geschäftsmodells mit einem innovativen Wertangebot liegen. Darüber hinaus soll sich das Wertangebot im Rahmen dieser Untersuchung dadurch auszeichnen, dass es auf nachhaltigen digitalen Produkten basiert.

Zielkonflikte zwischen Nachhaltigkeit und Gewinnstreben entstehen häufig, da der Fokus auf Gewinn kurzfristig ausgerichtet ist, während Nachhaltigkeit eine eher langfristige Perspektive erfordert (vgl. Colsman, 2016, S. 26). Die Entscheidungsregel zur Lösung dieses Konflikts lautet daher, die Interessen der Stakeholder an Nachhaltigkeit, deren Erfüllung langfristig zum Unternehmenserfolg beiträgt, vor das kurzfristige Gewinnstreben zu stellen (vgl. Reimer & Schäffer, 2024, S. 9-13).

Vor diesem Hintergrund ist es für die Entwicklung eines innovativen CDR-basierten Geschäftsmodells unerlässlich, dass eine CDR-spezifische Investitionsbereitschaft besteht. Viele Innovationen scheitern bereits an Unterinvestition, da kurzfristige Profite langfristigen Erfolgen vorgezogen werden (vgl. Andersen & Börsch, 2016, S. 59). Mit Blick auf CDR kommt hinzu, dass nachhaltiges Handeln häufig mit kurzfristigen Zusatzkosten verbunden ist. Da Nachhaltigkeit jedoch langfristig zum Unternehmenserfolg beitragen kann (vgl. Reimer und Schäffer, 2024, S. 9-13; Lai et al., 2010, S. 462-466), sollten diese Kosten - etwa für den Aufbau eines effizienten Recyclingsystems oder für eine fairere Entscheidungsfindung von KI-Algorithmen - als Investition in die Zukunft betrachtet werden. Dasselbe gilt für Zusatzkosten, die in der Entwicklung digitaler Produkte anfallen und der Prävention negativer Externalitäten dienen, die im späteren Gebrauch des Produkts entstehen würden (vgl. Spiekermann, 2016, S. 175). Für die Deckung dieser Kosten braucht es die Bereitschaft des Managements, in CDR zu investieren. Zur aussagekräftigen Messung und Förderung einer CDR-spezifische Investitionsbereitschaft kann der Anteil CDR-spezifischer Investitionen am Investitionsbudget ermit-

telt werden.

Ob die agile Entwicklung eines digitalen Produkts gelingt, das neben Benutzerfreundlichkeit auch durch Nachhaltigkeit überzeugt, hängt maßgeblich von einem gut funktionierenden Stakeholder-Dialog ab. Ein solcher Dialog kann jedoch scheitern, falls das Management nicht dazu bereit ist, von einem strikten Budget abzuweichen oder es anderweitig Interessenskonflikte zwischen den Beteiligten gibt (vgl. Spiekermann, 2016, S. 175). Die Stakeholder-Zufriedenheit ist ein Indikator dafür, inwieweit verschiedene Interessen, auch unter den Stakeholdern selbst, in Einklang gebracht werden konnten (vgl. Wehrum & Mainoo, 2013, S. 63). Überdies kann sie im Verlauf der Produktentwicklung ein Indikator dafür sein, ob das spätere Endprodukt auch wirklich nachhaltig sein wird. Daher empiehlt es sich, die Zufriedenheit der Stakeholder über Befragungen zu messen und zum Ziel zu erklären.

Ferner ist es sinnvoll, dass Image über Befragungen zu messen. Ein CDR-basiertes Geschäftsmodell zielt darauf ab, über ein besseres Image das Finanzergebnis zu steigern. Der Zusammenhang zwischen einem besseren Image, das aus der Übernahme von Verantwortung resultiert, und einem besseren Finanzergebnis, ist plausibel und wird von Studien bestätigt (vgl. Lai et al., 2010, S. 460-466). Um aus den Befragungen einen aussagekräftigen Indikator zu erhalten, sollten überwiegend Gruppen von Personen oder Organisationen befragt werden, die dem angestrebten Kundensegment angehören, das erreicht und bedient werden soll.

Um den finanziellen Erfolg des CDR-basierten Geschäftsmodells transparent machen zu können, braucht es finanzielle Ergebniskennzahlen. In diesem Zusammenhang ist es auch von Interesse, erzielte Kosteneinsparungen, die z. B. über die Verwendung recycelter Materialien oder eine Steigerung der Energieeffizienz von IT-Infrastruktur realisiert worden sind, zu messen.

Darüber hinaus könnten durch eine Berechnung des Return on Investment (ROI) von CDR-spezifischen Investitionen die positiven Effekte von CDR auf das Finanzergebnis dargelegt werden. Das Center for Sustainable Business der New York University hat die Return on Sustainability Investment-Methode entwickelt, die es mitunter auch auf Basis von Schätzungen ermöglichen soll, den finanziellen Nutzen von nachhaltigkeitsorientierten Investitionen zu bewerten (vgl. Atz et al., 2020, S. 303-309). Auf dieser Grundlage könnte auch der ROI von CDR-spezifischen Investitionen kalkuliert werden. Gleichwohl ist das bislang nur eine theoretische Überlegung, deren Umsetzbarkeit es in einem praktischen Anwendungskontext zu überprüfen gilt.

Auf der Basis der in diesem Kapitel für die Performance-Messung besprochenen sieben CDR-spezifischen Verantwortungsbereiche in den vier Handlungsfeldern wird im kommenden Kapitel 4.3 eine *CDR-Balanced Scorecard* vorgestellt.

4.3 Entwurf einer CDR-Balanced Scorecard

Ein Steuerungsansatz, der CDR ganzheitlich mit einer Verbesserung des Finanzergebnisses verbinden soll, erfordert mit Blick auf die BSC mehr als die einfache additive Integration einer fünften CDR-spezifischen Perspektive. Stattdessen empfiehlt sich der integrative Ansatz, über den CDR in allen vier klassischen Perspektiven der BSC Berücksichtigung findet.

Auf dieser Grundlage können auch, sofern die hierarchische Anordnung einer Strategy Map befolgt wird, in den der Finanzperspektive vorlaufenden Perspektiven - Lernen und Entwicklung, Interne Prozesse und Kunden - CDR-Ziele festgelegt werden. Deren Erfüllung trägt dann dazu bei, dass Unternehmen im Ergebnis ihrer CDR-spezifischen Verantwortung nachkommen und zugleich das finanzielle Ergebnis steigern. Erst dadurch wird die *CDR-Balanced Scorecard* zu einem ganzheitlichen Planungsinstrument, das die Identifizierung langfristig wegtreibender CDR-Ziele ermöglicht.

Solche CDR-Ziele lassen sich aber nur erreichen, wenn geeignete Nachhaltigkeitskennzahlen entwickelt werden, die in Anbetracht dieser Ziele eine Steuerungs- und Anregungsfunktion erfüllen. So können Leistungen, die der langfristigen Zielerreichung dienen, bewertet und im Weiteren auch gesteigert werden. Und zwar von Perspektive zu Perspektive im Sinne eines zielorientierten Ursache-Wirkungs-Handelns bis hin zum Finanzergebnis.

Eine solche *CDR-Balanced Scorecard* als Strategy Map, die einen strategischen Steuerungsansatz zur Verfolgung CDR-spezifischer Verantwortungsbereiche veranschaulicht, wird in der nachfolgenden Abbildung gezeigt. Dabei ist zu erwähnen, dass Unternehmen diese allgemeine *CDR-Balanced Scorecard* im Hinblick auf ihren praktischen Anwendungskontext im jeweiligen Anwendungsbereich ihres Unternehmens individualisieren, also anpassen sollten.

In der untersten Perspektive dieser *CDR-Balanced Scorecard* - also Lernen und Entwicklung - geht es um die CDR-Ziele “CDR-spezifische Mitarbeiterkompetenzen” und “CDR-spezifische Investitionsbereitschaft”, die grundlegend dafür sind, dass eine “Agile Entwicklung nachhaltiger digitaler Produkte” sowie der “Einsatz digitaler Technologien zur nachhaltigen Prozessoptimierung” gelingen.

Wie bereits in Kapitel 4.2 erläutert, ist ein verantwortungsvolles Handeln zunächst mit Zusatzkosten verbunden, die sich erst langfristig rentieren und zu deren Deckung es eben einer Bereitschaft des Managements, in CDR zu investieren, bedarf. Eine solche Bereitschaft lässt sich exemplarisch anhand des Anteils CDR-spezifischer Investitionen am Investitionsbudget messen. Ferner sind CDR-spezifische Mitarbeiterkompetenzen erforderlich, deren Aufbau durch ein adäquates Aus- und Weiterbildungsprogramm gefördert werden sollte. Dadurch können Unternehmen ihren individuellen Personalbedarf decken und die digitale Transformation nachhaltig bewältigen.

Auf dieser Grundlage können in der folgenden Perspektive - die der internen Prozesse - die CDR-Ziele “Agile Entwicklung nachhaltiger digitaler Technologien” und “Einsatz digitaler Technologien zur nachhaltigen Prozessoptimierung” erreicht werden. Die Erreichung des letztgenannten Ziels, das sich mit Blick auf die IT-Infrastruktur eines Unternehmens exemplarisch über die PUE messen lässt, wirkt sich unmittelbar

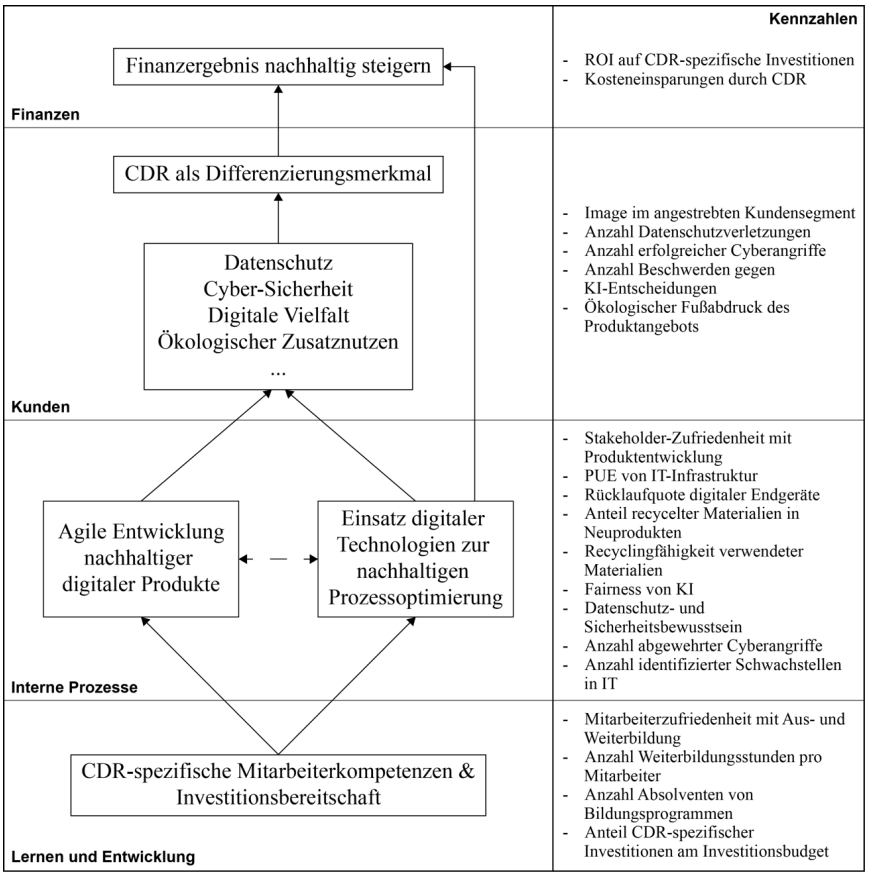


Abbildung 2: CDR-Balanced Scorecard

positiv auf das Finanzergebnis aus. Neben der PUE sind auf der zweiten Ebene weitere Kennzahlen zielführend einsetzbar, wie etwa die Fairness von KI, die noch in der Entwicklungsphase als Indikator für die spätere Nachhaltigkeit eines KI-Algorithmus genutzt werden kann oder die Anzahl identifizierter Schwachstellen in IT, die als Prozess-Indikator für die Leistungsfähigkeit eines SOC dient.

In der dritten Perspektive - die der Kunden - sind basierend auf den zuvor erreichten Zielen der ersten und zweiten Perspektive die CDR-Ziele “Datenschutz, Cyber-Sicherheit, digitale Vielfalt und ökologischer Zusatznutzen” zu nennen. Diese spiegeln die CDR-spezifisch getragene Verantwortung eines Unternehmens im Ergebnis wieder und können beispielsweise über die Anzahl erfolgreicher Cyberangriffe oder den ökologischen Fußabdruck des Produktangebots gemessen werden.

Sofern die eben genannten CDR-Ziele erreicht werden, kann sich CDR im Anschluss als Differenzierungsmerkmal eines Unternehmens etablieren, was sogleich das finale Ziel ist, um darauf aufbauend in der vierten Perspektive das Finanzergebnis nachhaltig steigern zu können. Als Kennzahlen zur Messung dieses Ziels sind der ROI auf CDR-spezifische Investitionen sowie die erzielten Kosteneinsparungen, die vor allem durch eine nachhaltige Prozessoptimierung erreicht werden sollen, zu nennen.

5 Fazit und Ausblick

Zusammenfassend lässt sich festhalten, dass die Vielzahl an Nachhaltigkeitskennzahlen, die in Literatur, Nachhaltigkeitsberichten und Berichtsstandards ausfindig gemacht werden konnte, einen hilfreichen Anhaltspunkt für die Performance-Messung in CDR-spezifischen Verantwortungsbereichen bietet.

Damit diese Kennzahlen im Rahmen des Controllings eine Anregungs- und Steuerungsfunktion erfüllen können, sind sie in Anbetracht des jeweiligen praktischen Anwendungskontextes zu präzisieren. Weiterhin sind Controller – wie bei der Integration von CSR bzw. Nachhaltigkeit in die Unternehmenssteuerung bereits üblich – darin gefordert, sich mit weichen Faktoren zu befassen, zu deren Messung Indikatoren unerlässlich sind. Darüber hinaus erfordert die Performance-Messung von CDR, dass Controller ein Verständnis für digitale Technologien, deren Auswirkungen auf Geschäftsmodelle sowie für die negativen und teils unvorhersehbaren Externalitäten entwickeln, die sich aus dem unternehmerischen Umgang mit digitalen Technologien ergeben können.

Jedoch gibt es CDR-spezifische Verantwortungsbereiche, deren Performance-Messung nur über eine interdisziplinäre Zusammenarbeit, etwa mit dem Fachpersonal eines SOC oder KI-Entwicklern gelingt. Ferner kann es mit einem großen Aufwand verbunden sein, die technischen Voraussetzungen für eine genaue Messung zu schaffen. So ist

es zur Messung der Energieeffizienz von Rechenzentren erforderlich, eine hohe Anzahl von dezentralen Messpunkten zu installieren, um eine präzisere Gesamtbeurteilung ermöglichen zu können. Hier muss das Controlling abwägen, ob und inwieweit dieser Aufwand eine bessere Entscheidungsgrundlage schafft und damit zur Erreichung langfristig werttreibender CDR-Ziele beiträgt.

Sofern ein Unternehmen bereit ist, einen Steuerungsansatz zu nutzen, der CDR ganzheitlich mit einer Verbesserung des Finanzergebnisses verbindet, empfiehlt sich die BSC als Planungs- und Steuerungsinstrument, bei dem CDR im Sinne eines integrativen Ansatzes in allen vier klassischen Perspektiven der *Balanced Scorecard* Berücksichtigung finden kann. Durch das weitere Voranschreiten der digitalen Transformation könnte CDR für immer mehr Unternehmen eine derart hohe strategische Bedeutung erlangen, dass die Vorteile eines strategischen Steuerungsansatzes den Nachteil des dazu betriebenen Aufwands bei Weitem überkompensieren.

Literatur

Adam, K. (2022). *Blockchain-Technologie für Unternehmensprozesse. Sinnvolle Anwendungen der neuen Technologien in Unternehmen* (2. Aufl.).

Andersen, N., & Börsch, A. (2016). Wie Controlling Innovationen richtig fördert. *Controlling und Management Review*, *60*, 58–63.

Apple. (2023). Fortschrittsbericht zum Umweltschutz. Für das Geschäftsjahr 2022.

Atz, U., Van Holt, T., Douglas, E., & Whelan, T. (2020). The Return on Sustainability Investment: Monetizing Financial Benefits of Sustainability Actions in Companies. In R. B. Swain & S. Sweet (Hrsg.), *Sustainable Consumption and Production – Volume II* (S. 303–354).

Baedeker, C., Bienge, K., Suski, P., & Themann, P. (2018). Sharing Economy: eine nachhaltige Konsumalternative. *Geographische Rundschau*, *70*(10), 22–28.

Bahr, T., Bassermann, B., Beuse, M., Kopton, T., Lehfeldt, A., Schauder, M., Trottmann, L., & Wegmann, T. H. (2023). *Management von CO2-Emissionen in hybriden IT-Umgebungen* (Techn. Ber.). Bitkom e.V.

Becker, W., & Ulrich, P. (2016). Einführung in das Controlling. In W. Becker & P. Ulrich (Hrsg.), *Handbuch Controlling* (S. 3–5).

Bieniek, J., Binder, C., Gänßlen, S., Keindorf, C., Losbichler, H., Petrin, P., Schäffer, U., Seidermann, P., Schuberth, K., Weber, J., & Zillmer, D. (2023). *Grundsatzposition* (Techn. Ber.). Internationaler Controller Verein/International Group of Controlling.

BMJV. (2021). Corporate Digital Responsibility-Kodex. Freiwillige Selbstverpflichtung mit Bericht.

Bogsnes, B. (2023). *Implementierung von Beyond Budgeting. Mehr Business-Agilität durch innovative Führungsprinzipien und Managementprozesse*.

Bolwin, L., Hönig, T., Kempermann, H., Kestermann, C., Klink, H., Kuttler, F., & Van Baal, S. (2023). *Der digitale Faktor. Wie Deutschland von intelligenten Technologien profitiert* (Techn. Ber.). IW Consult GmbH.

Bosch. (2023). Presse-Information. IAA Mobility 2023: Bosch wächst mit Lösungen und Technik für das Software-definierte Fahrzeug.

Bozem, K., & Nagl, A. (2021). *Digitale Geschäftsmodelle erfolgreich realisieren. Business Model Building mit Checklisten und Fallbeispielen* (2. Aufl.).

Buxmann, P., & Schmidt, H. (2019). Grundlagen der Künstlichen Intelligenz und des Maschinellen Lernens. In P. Buxmann & H. Schmidt (Hrsg.), *Künstliche Intelligenz. Mit Algorithmen zum wirtschaftlichen Erfolg* (S. 3–20).

Colsman, B. (2016). *Nachhaltigkeitscontrolling* (2. Aufl.).

Corsten, H., & Roth, S. (2022). Gedanken zur Digitalisierung. In S. Roth & H. Corsten (Hrsg.), *Handbuch Digitalisierung* (S. 3–18).

Crane, A., Matten, D., & Spence, L. J. (2013). *Corporate Social Responsibility. Readings and Cases in a Global Context*. Routledge.

Deegan, C. (2006). Legitimacy Theory. In Z. Hoque (Hrsg.), *Methodological Issues in Accounting Research* (S. 161–182).

Deutsche Telekom. (2023). CDR-Kodex: Maßnahmenbericht.

Deutsche Telekom. (2024). Corporate Responsibility Bericht 2023.

Dörr, S. (2020). *Praxisleitfaden Corporate Digital Responsibility. Unternehmerische Verantwortung und Nachhaltigkeitsmanagement im Digitalzeitalter*.

Drossel, W., Ihlenfeldt, S., Langer, T., & Dumitrescu, R. (2018). Cyber-Physische Systeme. Forschen für die digitale Fabrik. In R. Neugebauer (Hrsg.), *Digitalisierung. Schlüsseltechnologien für Wirtschaft und Gesellschaft* (S. 197–222).

Europäische Kommission. (2001). Grünbuch. Europäische Rahmenbedingungen für die soziale Verantwortung der Unternehmen.

Europäische Kommission. (2011). Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen.

Europäische Union. (2024). Amtsblatt der Europäischen Union: Berichtigung der Delegierten Verordnung (EU) 2023/2772 der Kommission vom 31. Juli 2023.

Gladen, W. (2014). *Performance Measurement. Controlling mit Kennzahlen* (6. Aufl.).

Gleich, R. (2021). *Performance Measurement. Konzepte, Fallstudien, Empirie und Handlungsempfehlungen* (3. Aufl.).

GRI. (2024). Konsolidierte GRI-Standards.

Hansen, E. G., & Schaltegger, S. (2016). The Sustainability Balanced Scorecard: A Systematic Review of Architectures. *Journal of Business Ethics*, *133*, 193–221.

Häusler, K., & Prinzler, J. (2020). *Econsense-Blueprint zur Umsetzung digitaler Verantwortung in Unternehmen* (Techn. Ber.). Econsense – Forum Nachhaltige Entwicklung der deutschen Wirtschaft e.V.

Heidbrink, L. (2017). Definitionen und Voraussetzungen der Verantwortung. In L. Heidbrink, C. Langbehn & J. Loh (Hrsg.), *Handbuch Verantwortung* (S. 5–6).

Helmold, M., Treu, J., Fritz, J., & Hummel, F. (2024). *ESG, CSR und SDG als langfristiger Wettbewerbsvorteil. Nachhaltigkeit durch innovative Konzepte, Methoden und Tools*.

Hintemann, R. (2018). *Effizienzgewinne reichen nicht aus: Energiebedarf der Rechenzentren steigt weiter deutlich an* (Techn. Ber.). Borderstep Institut für Innovation und Nachhaltigkeit.

Horvath, P., Gleich, R., & Seiter, M. (2023). *Controlling* (15. Aufl.).

IGC & Weißenberger, B. E. (2006). *Controller und IFRS: Konsequenzen für die Controlleraufgaben durch die Finanzberichterstattung nach IFRS* (Techn. Ber.). International Group of Controlling.

IHK Berlin. (2017). Digitalisierung und Diversity – Vielfalt als Erfolgsfaktor nutzen.

Kaplan, R. S., & Norton, D. P. (2024). *Balanced Scorecard. Strategien erfolgreich umsetzen* (1. Aufl.) [1. Auflage (1997)].

Kempkes, E., Kempkes, J. A., Suprano, F., & Wömpener, A. (2021). Corporate Digital Responsibility. Ausgestaltung und Steuerung der unternehmerischen Verantwortung im digitalen Zeitalter. *Controlling – Zeitschrift für erfolgsorientierte Unternehmenssteuerung*, *33*(6), 29–35.

Kenning, P., & Weißenberger, B. E. (2019). Digitale Risiken und Nebenwirkungen [26.08.2019]. *Frankfurter Allgemeine Zeitung*, 16.

Kirmes, S. (2017). *Ressourceneffizienz durch Industrie 4.0. Potenziale für KMU des verarbeitenden Gewerbes. Kurzzusammenfassung* (Techn. Ber.). VDI Zentrum Ressourceneffizienz GmbH.

Kobi, J. (2021). Personal-Controlling: Eine verkannte Controlling-Dimension ganzheitlich angehen. In R. Eschenbach, J. Baumüller & H. Siller (Hrsg.), *Funktions-Controlling. Praxishandbuch für Unternehmen, Non-Profit-Organisationen und die öffentliche Verwaltung* (2. Aufl., S. 365–388).

Koch, J., & Grimm, C. (2022). Cyber-Physikalische Systeme. In S. Roth & H. Corsten (Hrsg.), *Handbuch Digitalisierung* (S. 316–334).

Köchling, A., Riazy, S., Wehner, M. C., & Simbeck, K. (2020). Highly Accurate, But Still Discriminatory: A Fairness Evaluation of Algorithmic Video Analysis. *Business and Information Systems Engineering*, *63*(1), 39–54.

Köhler, A. R., Gröger, J., & Liu, R. (2018). *Kurzgutachten: Energie- und Ressourcenverbräuche der Digitalisierung. Kurzgutachten* (Techn. Ber.). Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen.

Küpper, H., Friedl, G., Hofmann, C., Hofmann, Y. E., & Pedell, B. (2024). *Controlling: Konzeption – Aufgaben – Instrumente* (7. Aufl.).

Lai, C., Chiu, C., Yang, C., & Pai, D. (2010). The Effects of Corporate Social Responsibility on Brand Performance: The Mediating Effect of Industrial Brand Equity and Corporate Reputation. *Journal of Business Ethics*, *95*, 457–469.

Lautermann, C., & Frick, V. (2023). Corporate Digital Responsibility. Wie Unternehmen im digitalen Wandel Verantwortung übernehmen. *Schriftenreihe des Institut für Ökologische Wirtschaftsforschung*, *227*.

Lehmkemper, D. (2023). *29. Tätigkeitsbericht 2023* (Techn. Ber.). Der Landesbeauftragte für den Datenschutz Niedersachsen.

Lerch, C., & Maloca, S. (2020). *Digitale Geschäftsmodelle – Modethema oder Wettbewerbsfaktor* (Techn. Ber.). Fraunhofer-Institut für System- und Innovationsforschung.

Leukert, B., Müller, J., & Noga, M. (2019). Das intelligente Unternehmen: Maschinelles Lernen mit SAP zielgerichtet einsetzen. In P. Buxmann & H. Schmidt (Hrsg.), *Künstliche Intelligenz. Mit Algorithmen zum wirtschaftlichen Erfolg* (S. 41–62).

Lingle, J. H., & Schiemann, W. A. (1996). From Balanced Scorecard to Strategic Gauges. Is Measurement Worth It. *Management Review*, *85*(3), 57–60.

Liu, Z., Zhang, W., & Zhao, F. (2022). Impact, Challenges and Prospect of Software-Defined Vehicles. *China Society of Automotive Engineers*, *5*(2), 180–194.

Lobschat, L., Mueller, B., Eggers, F., Brandimarte, L., Diefenbach, S., Kroschke, M., & Wirtz, J. (2021). Corporate Digital Responsibility. *Journal of Business Research*, *122*, 875–888.

Loebbecke, C. (2006). Digitalisierung – Technologien und Unternehmensstrategien. In C. Scholz (Hrsg.), *Handbuch Medienmanagement* (S. 357–373).

Matzler, K., Ballom, F., von den Eichen, S. F., & Anschober, M. (2016). *Digital Disruption: Wie Sie Ihr Unternehmen auf das digitale Zeitalter vorbereiten*.

Mell, P., & Grance, T. (2011). *The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology* (Techn. Ber.). NIST.

Mercedes Benz Group. (2024). Nachhaltigkeitsbericht 2023.

Merchant, K., & Van der Stede, W. (2023). *Management Control Systems* (5. Aufl.).

Murphy, K. P. (2012). *Machine Learning. A Probabilistic Perspective*. MIT Press.

Obermaier, R. (2019). Industrie 4.0 und Digitale Transformation als unternehmerische Gestaltungsaufgabe. In R. Obermaier (Hrsg.), *Handbuch Industrie 4.0 und Digitale Transformation* (S. 3–46).

Onwubiko, C., & Onwubiko, A. (2019). Cyber KPI for Return on Security Investment. *International Conference on Cyber Situational Awareness*, 1–9.

Osterwalder, A., & Pigneuer, Y. (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*.

Pflaum, A., & Klötzer, C. (2019). Von der Pipeline zur Plattform – Strategische Implikationen für das Unternehmen. In W. Becker, B. Eierle, A. Fliaster, B. Ivens, A. Leischnig, A. Pflaum & E. Sucky (Hrsg.), *Geschäftsmodelle in der digitalen Welt* (S. 57–74).

Porter, M. E., & Kramer, M. R. (2015). Shared Value – Die Brücke von Corporate Social Responsibility zu Corporate Strategy. In A. Schneider & R. Schmidpeter (Hrsg.), *Corporate Social Responsibility. Verantwortungsvolle Unternehmensführung in Theorie und Praxis* (2. Aufl., S. 145–160).

Reichmann, T., Kißler, M., & Baumöl, U. (2017). *Controlling mit Kennzahlen. Die systemgestützte Controlling-Konzeption* (9. Aufl.).

Reimer, M., & Schäffer, U. (2024). Gedanken zur Nachhaltigkeit in der Unternehmenssteuerung. *Controlling und Management Review*, *68*, 8–13.

Sailer, U. (2022). *Nachhaltigkeitscontrolling. So werden Unternehmen nachhaltig gesteuert* (4. Aufl.).

SAP. (2023). Integrierter Bericht der SAP 2023.

Schäffer, U., & Brückner, L. (2019). Rollenspezifische Kompetenzprofile für das Controlling der Zukunft. *Controlling und Management Review*, *63*, 14–31.

Schmidpeter, R. (2015). Unternehmerische Verantwortung – Hinführung und Überblick. In A. Schneider & R. Schmidpeter (Hrsg.), *Corporate Social Responsibility. Verantwortungsvolle Unternehmensführung in Theorie und Praxis* (2. Aufl., S. 1–20).

Schmidpeter, R., & Günther, E. (2013). CSR – Innovationen im Rechnungswesen. *Controlling und Management Review*, *57*, 26–35.

Schneider, A. (2015). Reifegradmodell CSR – eine Begriffserklärung und -abgrenzung. In A. Schneider & R. Schmidpeter (Hrsg.), *Corporate Social Responsibility. Verantwortungsvolle Unternehmensführung in Theorie und Praxis* (2. Aufl., S. 21–42).

Schön, D. (2022). *Planung und Reporting im BI-gestützten Controlling* (4. Aufl.).

Schönbohm, A., & Egle, U. (2021). Controlling der digitalen Transformation. In D. R. A. Schallmo, A. Rusnjak, J. Anzengruber, T. Werani & K. Lang (Hrsg.), *Digitale Transformation von Geschäftsmodellen* (2. Aufl., S. 199–224).

Seufert, A., & Oehler, K. (2016). Controlling und Big Data: Anforderungen an die Methodenkompetenz [hrsg. von U. Schäffer/J. Weber, Sonderheft 1]. *Controlling und Management Review*, 74–81.

Spiekermann, S. (2016). *Ethical IT Innovation. A Value-Based System Design Approach*.

Strasser, K., & Niedermayer, B. (2021). Unvoreingenommenheit von Künstliche-Intelligenz-Systemen. Die Rolle von Datenqualität

und Bias für den verantwortungsvollen Einsatz von Künstlicher Intelligenz. In R. Altenburger & R. Schmidpeter (Hrsg.), *CSR und Künstliche Intelligenz* (S. 121–136).

Thorun, C., Kettner, S. E., & Merck, J. (2018). *Ethik in der Digitalisierung: Der Bedarf für eine Corporate Digital Responsibility* (Techn. Ber.). Abteilung Wirtschafts- und Sozialpolitik der Friedrich-Ebert-Stiftung.

Ulrich, H. (1975). Der allgemeine Systembegriff. In J. Baetge (Hrsg.), *Grundlagen der Wirtschafts- und Sozialkybernetik* (S. 33–39).

Van Alstyne, M. W., Parker, G. G., & Choudary, S. P. (2016). Pipelines, Platforms, and the New Rules of Strategy. *Harvard Business Review*, *94*(4), 54–62.

Van der Aalst, W. (2016). *Process Mining. Data Science in Action* (2. Aufl.).

Vial, G. (2019). Understanding Digital Transformation: A Review and a Research Agenda. *The Journal of Strategic Information Systems*, *28*(2), 118–144.

Weber, J., & Schäffer, U. (2022). *Einführung in das Controlling* (17. Aufl.).

Wehrum, K., & Mainoo, M. (2013). Nachhaltigkeit in Management-Systemen – Stakeholder als Erfolgsfaktor. *Controlling und Management Review*, *57*, 58–65.

Weißenberger, B. E. (2021). Künstliche Intelligenz als Zukunftstechnologie im Controlling. *Controlling und Management Review*, *65*, 8–17.

Weißenberger, B. E., & Marrocco, A. (2022). Corporate Digital Responsibility und ihre Integration in die Unternehmensführung. In S. Roth & H. Corsten (Hrsg.), *Handbuch Digitalisierung* (S. 41–58).

Westerman, G., Tannou, M., Bonnet, D., Ferraris, P., & McAfee, A. (2012). *The Digital Advantage: How Digital Leaders Outperform Their Peers in Every Industry* (Techn. Ber.). The MIT Center for Digital Business/Capgemini Consulting.

Wilts, H., & Berg, H. (2017). The Digital Circular Economy: Can the Digital Transformation Pave the Way for Resource-Efficient Materials Cycles. *International Journal of Environmental Sciences and Natural Resources*, *7*(5), 135–139.

Wojcik, P. (2016). How Creating Shared Value Differs from Corporate Social Responsibility. *Central European Management Journal*, *24*(2), 32–55.

World Bank Group. (2024). Digital Progress and Trends Report 2023.

World Commission on Environment and Development. (1987). *Our Common Future*.

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Steuerliche Analyse von Direktinvestitionen in Bulgarien im Vergleich zu einer Geschäftsausweitung in Deutschland für mittelständische Unternehmen

Rozmari Genova

Europa-Universität Viadrina Frankfurt (Oder)

Abstract

Eastern European countries are becoming increasingly important for German medium-sized enterprises, with Bulgaria emerging as a relevant investment destination. Against this background, this thesis examines whether direct investment in Bulgaria provides tax advantages compared to domestic business expansion in Germany, focusing on the quantification of tax burden differences. Three forms of direct investment are analysed: a permanent establishment, a partnership, and a corporation. Based on an analysis of relevant tax frameworks in both countries and the applicable tax treaty, a five-year Monte Carlo simulation covering profit and loss scenarios is conducted to assess industry-independent tax effects of a Bulgarian direct investment compared to a domestic expansion. The results show that investment in Bulgaria is generally more tax-efficient, particularly when structured as a permanent establishment (average tax benefit: 26.83% in profit scenario). In loss periods, however, the foreign investment leads to a higher tax burden than the domestic case. The analysis underlines Bulgaria's tax potential for German medium-sized companies while emphasizing realistic tax planning and company- and location-specific factors in cross-border investment decisions.

Zusammenfassung

Osteuropäische Länder gewinnen für den deutschen Mittelstand zunehmend an Bedeutung, wobei Bulgarien als Investitionsstandort in den Fokus rückt. Vor diesem Hintergrund untersucht die Masterarbeit, ob eine Direktinvestition in Bulgarien steuerliche Vorteile gegenüber einer Geschäftsausweitung in Deutschland bietet, mit Schwerpunkt auf der Quantifizierung von Steuerbelastungsunterschieden. Analysiert werden drei Direktinvestitionsformen: Betriebsstätte, Personengesellschaft und Kapitalgesellschaft. Auf Grundlage der innerstaatlichen Steuerrechtsnormen beider Staaten sowie des Abkommensrechts wird eine Monte-Carlo-Simulation mit Gewinn- und Verlustszenarien durchgeführt, um branchenunabhängige Steuereffekte der Auslandsinvestition im Vergleich zur Inlandsausweitung zu evaluieren. Die Ergebnisse zeigen, dass Direktinvestitionen in Bulgarien grundsätzlich steuerlich vorteilhafter sind, insbesondere bei der Betriebsstätte (Durchschnittssteuervorteil: 26,83 % im Gewinnszenario). In Verlustphasen führt die Auslandsinvestition hingegen zu einer höheren Steuerbelastung als im Inland. Die Analyse verdeutlicht das steuerliche Potenzial Bulgariens sowie die Bedeutung realistischer Steuerplanung und unternehmens- sowie standortspezifischer Faktoren bei grenzüberschreitenden Investitionen.

Keywords: ausländische direktinvestition; internationales steuerrecht; monte-carlo-simulation; abkommensrecht; steuervergleich deutschland–bulgarien

1 Einleitung

Das wissenschaftliche Interesse an den Direktinvestitionen in Osteuropa besteht bereits seit den Wirtschaftswandlungen des späten 20. Jahrhunderts und wurde durch die Ausweitung der Europäischen Union (EU) auf der Ostflanke weiter intensiviert.[1]

Heute, nach mehreren Jahrzehnten der wirtschaftlichen Integration, sind die Staaten Osteuropas von steigender Bedeutung für den deutschen Mittelstand.[2] Zwar bleibt Deutschland nach wie vor der meistbevorzugte Investiti-

Die vorliegende Masterarbeit widme ich meiner zutiefst geliebten Oma Monika Marinowa. Mein Dank gilt ihrer unermüdlichen Unterstützung während der Erstellung dieser Arbeit. Mit ihrer Liebe, ihrem Glauben an mich, ihrer Ermutigung und ihrer Freude an meinen Erfolgen hat sie mich während meiner gesamten akademischen Laufbahn begleitet, mich bestärkt und für immer geprägt. In tiefster Dankbarkeit und liebevollster Erinnerung.

[1] Vgl. bspw. Kaufmann und Menke (1997); Estrin et al. (2000); Marinova und Marinov (2003); Detscher (2012); Elschner et al. (2015), Makhavikova (2018).

[2] Vgl. Abel-Koch (2023), S. 2: Die deutschen Investitionen in Osteuropa sind im Zeitraum zwischen 2019 und 2022 im Vergleich zum Zeitraum 2012–2015 um 11 Prozent angestiegen.

DOI: https://doi.org/10.5282/jums/v11i3pp488-512

© The Author(s) 2026. Published by Junior Management Science.
This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.

onsstandort für inländische mittelständische Unternehmen, jedoch ist in den nächsten fünf Jahren mit einem steigenden Interesse an Auslandsinvestitionen zulasten Deutschlands zu rechnen.[3]

Im Hinblick auf diese Tendenz rückt die osteuropäische Republik Bulgarien als attraktiver Investitionsstandort für deutsche mittelständische Unternehmen in den Vordergrund. Dies lässt sich unter anderem damit begründen, dass sich der bulgarische Staat durch die niedrigste Unternehmensbesteuerung[4] sowie das niedrigste Lohnkostenniveau[5] in der EU auszeichnet.

1.1 Problemstellung und Aufbau

Vor diesem Hintergrund widmet sich die vorliegende Masterarbeit der Fragestellung, ob und in welchem Ausmaß eine Direktinvestition in Bulgarien im Vergleich zu einer Geschäftsausweitung in Deutschland steuerliche Vorteile für mittelständische deutsche Unternehmen bietet.

Im Rahmen der Arbeit wird als Geschäftsausweitung die Gründung einer inländischen Betriebsstätte durch eine deutsche Personengesellschaft definiert. Direktinvestitionen werden als über einen reinen Kapitaltransfer hinausgehend verstanden und beziehen sich auf die aktive Übernahme geschäftlicher Funktionen im Zielstaat.[6] Drei Formen der Direktinvestition werden betrachtet: die Gründung einer Betriebsstätte, einer Tochtergesellschaft sowie einer Kapitalgesellschaft.

Sämtliche Analysen in der wissenschaftlichen Literatur zu der steuerlich vorteilhaftesten Form der Direktinvestition im Ausland besagen, dass die Beurteilung der Vorteilhaftigkeit stets vom konkreten Einzelfall abhängig ist, da eine Berücksichtigung der spezifischen Vorschriften des konkreten Investitionsstandortes erforderlich ist.[7]

Aufbauend auf dieser Erkenntnis stellt sich in der vorliegenden Arbeit insbesondere die Frage, welche steuerlichen Konsequenzen die drei Direktinvestitionsformen in Bulgarien und Deutschland nach sich ziehen und welche steuerlichen Problemfelder dabei entstehen. Von Interesse für die Fragestellung ist die Quantifizierung der entstehenden Steuerlast im Rahmen der drei Gestaltungsformen und im Vergleich zu einer rein inländischen Geschäftsausweitung.

Zu diesem Zweck bedient sich die Untersuchung der stochastischen Methode der Monte-Carlo-Simulation. Es wird eine realitätsnahe und branchenunabhängige steuerliche Simulation in zwei Hauptszenarien (Gewinn- und Verlustszenario) unter Berücksichtigung der einschlägigen steuerlichen Vorschriften durchgeführt. Dabei wird der Barwert der Gesamtsteuerbelastung beim Vorliegen der drei Alternativen einer Direktinvestition in Bulgarien ausgerechnet und mit einer Geschäftsausweitung in Deutschland verglichen. Der Analyse wird eine natürliche Person mit Wohnsitz und gewöhnlichem Aufenthalt in Deutschland (deutscher Investor) zugrunde gelegt, die entweder mittelbar über eine GmbH & Co KG eine Betriebsstätte oder Kapitalgesellschaft oder unmittelbar eine Personengesellschaft in Bulgarien gründet. Gemäß der Zielsetzung der vorliegenden Masterarbeit sind sowohl die Besteuerung der ausländischen Einkünfte als auch die Besteuerung der inländischen Einkünfte von Bedeutung und in die Analyse einzubeziehen.

Für die Zwecke der Untersuchung wird im zweiten Kapitel der steuerrechtliche Rahmen der Besteuerung der drei Fallkonstellationen bzw. Formen der Direktinvestition in Bulgarien separat erläutert. Die Ausführungen erfolgen unter Rückgriff auf die einschlägigen Gesetzesnormen der beiden beteiligten Steuerhoheiten, Deutschland und Bulgarien, sowie anschließend unter Einbeziehung des Abkommensrechts. Zudem wird auf mögliche steuerliche Qualifikationskonflikte hingewiesen. Auf dem steuerrechtlichen Rahmen aufbauend legt das dritte Kapitel das Vorgehen und die Annahmen der Monte-Carlo-Simulation dar. In diesem Zusammenhang werden die Simulationsszenarien konstruiert und die Ergebnisse dargestellt und ausgewertet. Anschließend werden die Grenzen der Untersuchung kurz erörtert. Schließlich erfolgt im vierten Kapitel (Fazit) eine Zusammenfassung der Ergebnisse der Arbeit.

1.2 Abgrenzung des Untersuchungsgegenstands

Aufgrund des begrenzten Umfangs der Arbeit bezieht sich die Darstellung nur auf die für das Verständnis und die Durchführung der vorgesehenen methodischen Untersuchung relevanten Aspekte und steuerlichen Vorschriften und Normen des Abkommensrechts. Aus Gründen der Realitätsnähe bzw. Haftungsbeschränkung fällt die Wahl auf die Rechtsform der GmbH & Co KG. Auf Aspekte dieser hybriden Rechtsform wird nicht eingegangen. Die Besteuerung auf Ebene der Gesellschaft mit beschränkter Haftung (GmbH) als Gesellschafterin der Personengesellschaft wird durchgehend vernachlässigt. Ein gesonderter Hinweis hierzu erfolgt im weiteren Verlauf der Arbeit nicht. Diese Masterarbeit geht von einer schon getroffenen Standortentscheidung aus, bei der Bulgarien als optimaler Alternativstandort zu einer Geschäftsausweitung in Deutschland gewählt wurde.

Im Rahmen der Arbeit wird ein mittelständisches Unternehmen in der Rechtsform einer Personengesellschaft betrachtet. Das Unternehmen erzielt unter Berücksichtigung der Geschäftsausweitung im In- und Ausland in der ersten Periode Jahresüberschüsse in Höhe von bis zu 1.000.000 Euro. Der Verlust in der ersten Periode ist auf den jeweiligen Negativbetrag begrenzt. In den Folgeperioden erhöhen sich die Bandbreiten aufgrund inflationsbedingter Anpassungen.

Es werden ausschließlich Ertragsteuern betrachtet. Aperiodische Sachverhalte werden nicht berücksichtigt. Sondervergütungen und Sonderbetriebsvermögen sowie

[3] Vgl. Deloitte (2024), S. 9: Circa 80 Prozent der befragten Geschäftsführer aus dem Mittelstand benennen Deutschland als ihren vorrangigen Investitionsstandort. Im Hinblick auf die kommenden fünf Jahre sinkt der Prozentsatz jedoch um 12 Prozentpunkte.

[4] Vgl. BMF (2024), S. 12.

[5] Vgl. Eurostat (2025), Spalte: 2024, Zeile: Bulgarien.

[6] So auch Hundsdoerfer / Jamrozy in: Kudert (2007), S. 535.

[7] So auch Kaminski (1996), S. 110; vgl. auch Rademacher-Gottwald (2016), HI2320166.

ihre steuerliche Behandlung bleiben außen vor. Treaty-Override-Regelungen sind nicht thematisiert. Es wird im Rahmen der Masterarbeit davon ausgegangen, dass der deutsche Investor eine aktive gewerbliche Geschäftstätigkeit beabsichtigt, die nicht der Steuervermeidung dient. Deshalb sind sämtliche Rechtsfolgen und Besonderheiten beim Vorliegen passiver Tätigkeiten im Ausland nicht dargelegt. Auf die Tatbestandsmerkmale des Betriebsstättenbegriffs in den verschiedenen Rechtskreisen wird ebenso wenig eingegangen wie auf die Besonderheiten und Ausnahmen bei dem Bestehen oder Beenden einer ausländischen Betriebsstätte oder Gesellschaft. Auch die Fragen der Einkunftsermittlung und -abgrenzung insbesondere des Betriebsstättengewinns sowie der Finanzierung des Auslandsengagements sind nicht Gegenstand dieser Masterarbeit. Zivilrechtliche, buchhalterische, verfahrens- und haftungsrechtliche Aspekte sowie Gründungsvoraussetzungen werden ebenfalls nicht thematisiert.

2 Besteuerung von Direktinvestitionen – steuerrechtlicher Rahmen

Die Auswahl der optimalen Ausgestaltung der ausländischen Direktinvestition ist für einen deutschen Investor von erheblicher Relevanz.[8] Im Rahmen dieser Entscheidungsfindung sind drei als notwendige Bedingungen fungierende Ausschlusskriterien zwingend zu berücksichtigen. Darunter fällt unter anderem die Steuerbelastung.[9] Auf dieser Grundlage gewinnt die nachfolgend dargelegte steuerliche Analyse der drei Direktinvestitionsformen am Standort Bulgariens zusätzlich an Relevanz.

Im Rahmen der Analyse ist zu beachten, dass das innerstaatliche Steuerrecht der beiden beteiligten Staaten und das zwischenstaatliche Recht der Doppelbesteuerungsabkommen (im weiteren Verlauf DBA) jeweils eigenständige rechtliche Kreise darstellen, die strikt voneinander zu unterscheiden sind.[10] Diese Trennung wird im Rahmen der steuerlichen Analyse konsequent eingehalten. Nachfolgend wird daher für jede einzelne Form des Auslandsengagements zunächst die konkrete Ausgestaltung der Besteuerungsrechte gemäß den innerstaatlichen Steuerrechtsordnungen der beteiligten Staaten dargelegt. Anschließend ist zu prüfen, ob und inwieweit das Abkommensrecht die jeweiligen nationalen Steuerhoheiten einschränkt. In allen Fällen wird davon ausgegangen, dass der Steuerpflichtige Einkünfte aus seiner Tätigkeit im Ausland und weiterhin Inlandseinkünfte erzielt.[11]

2.1 Gründung einer Betriebsstätte

Die Gründung einer Betriebsstätte stellt eine mögliche Gestaltungsart des unternehmerischen Auslandsengagements dar. Sie gilt weder für rechtliche noch für steuerliche Zwecke als eigenständiges Subjekt.[12] Dabei wird eine feste Geschäftseinrichtung im Ausland etabliert, die einen wirtschaftlich in unterschiedlichem Maße eigenständig operierenden Teil des Unternehmens darstellt.[13]

Die als Basis für die folgenden Ausführungen dienende konkrete Konstellation baut auf einem deutschen Investor auf, der in Deutschland über eine GmbH & Co KG gewerblich tätig ist und die Option eines Auslandsengagements über eine Betriebsstätte in Bulgarien in Betracht zieht. An der deutschen Kommanditgesellschaft (KG) ist der deutsche Investor als Kommanditist zu 100 % beteiligt. Der Beteiligung der GmbH als Komplementärin ist auf null Prozent festgelegt. Die Struktur ist schematisch in Abbildung 1 veranschaulicht.

Die bulgarische Betriebsstätte erzielt aktive Einkünfte im Sinne des § 8 Abs. 1 AStG und § 2a Abs. 1 EStG und erfüllt annahmegemäß die Tatbestandsmerkmale einer Betriebsstätte in allen Rechtskreisen, nämlich nach dem nationalen Steuerrecht Bulgariens, dem deutschen Steuerrecht sowie dem Abkommensrecht.

2.1.1 Nationales Recht

Der vorliegende Sachverhalt wird aus deutscher und bulgarischer Sicht steuerlich gewürdigt.

Deutschland

Unter Zugrundelegung der getroffenen Annahmen liegen die Voraussetzungen des deutschen Begriffs der Betriebsstätte vor. Davon hängt in Deutschland vor allem die Anwendung der unilateralen Maßnahmen zur Vermeidung der Doppelbesteuerung ab.[14] Aufgrund des bestehenden DBA zwischen Deutschland und Bulgarien ist diese in dem vorliegenden Fall irrelevant (§ 34c Abs. 6 Satz 1 EStG). Aus den oben genannten Gründen wird im Folgenden auf eine nähere Betrachtung des deutschen Betriebsstättenbegriffs verzichtet.

Die Ausführungen befassen sich im weiteren Verlauf mit der Prüfung der persönlichen und sachlichen Steuerpflicht des deutschen Stammhauses. Da die GmbH & Co KG nach deutschem Steuerrecht eine Personengesellschaft darstellt, gilt das Transparenzprinzip. Der Begriff der *Transparenz* bezieht sich dabei auf die fehlende Steuerpflicht der Gesellschaft selbst. Die maßgeblichen Steuersubjekte sind ihre Gesellschafter.[15] Sind diese im Inland persönlich steuerpflichtige natürliche Personen (§ 1 EStG), unterliegen sie mit ihrem Gewinnanteil[16] als Einkünfte aus Gewerbebetrieb der Einkommensteuer (§ 15 Abs. 1 S. 1 Nr. 2 S. 1 EStG). Die Zurechnung erfolgt im Jahr der Gewinnentstehung, wobei Fragen der Ausschüttung oder Thesaurierung irrelevant sind.[17] Die Gewinnermittlung der Personengesellschaften richtet sich nach den allgemeinen Vorschriften des § 4 EStG.[18] Nach dem Maßgeblichkeitsprinzip (§ 5 Abs. 1

[8] Vgl. Kudert in: Elschner et al. (2015), S. 29.
[9] Vgl. Kudert in: Elschner et al. (2015), S. 30f.
[10] Vgl. Puls in: Wassermeyer et al. (2018), Rn. 2.1.
[11] So auch Kudert et al. (2012), S. 2f.
[12] Vgl. Wassermeyer in: Wassermeyer et al. (2018), Rn. 1.4.
[13] Vgl. Spengel in: Jacobs (2023), S. 384.
[14] Vgl. Spengel in: Jacobs (2023), S. 385.
[15] Vgl. Wassermeyer in: Wassermeyer et al. (2015), Rn. 2.5.
[16] Die Sondervergütungen (§ 15 Abs. 1 S. 1 Nr. 2 S. 1 Halbsatz 2 EStG) werden im Rahmen der Masterarbeit durchgängig ausgeklammert.
[17] Vgl. Hennrichs in: Tipke und Lang (2024), Rn. 13.6.
[18] Vgl. Hennrichs in: Tipke und Lang (2024), Rn. 10.100.

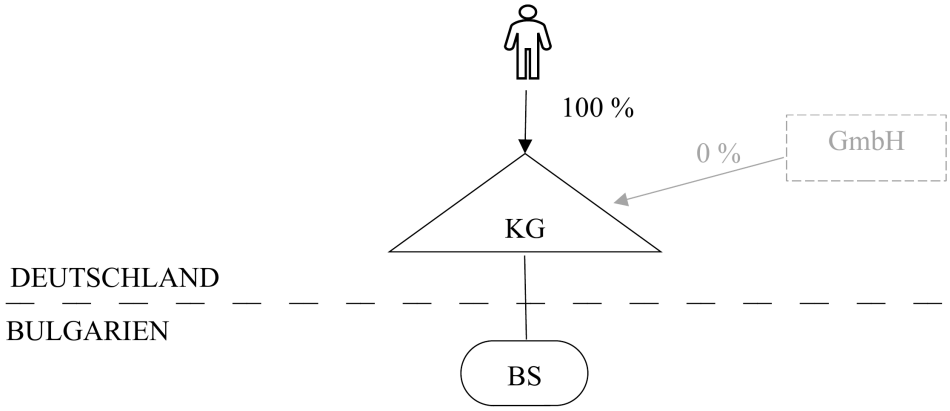


Abbildung 1: Gründung einer Betriebsstätte in Bulgarien (Schematische Darstellung)

EStG) ist bei Gewerbetreibenden die Handelsbilanz[19] für die Steuerbilanz maßgeblich, soweit das Steuergesetz keine abweichenden Normen vorsieht.[20] Somit werden die handelsrechtlich gebildeten Rückstellungen mit gewissen Ausnahmen grundsätzlich für steuerliche Zwecke anerkannt.[21]

Bei der Besteuerung einer von der Personengesellschaft unterhaltenen Betriebsstätte ist analog die Ebene der Gesellschafter maßgeblich.[22] Sie wird als anteilige Betriebsstätte der jeweiligen Gesellschafter angesehen.[23] Dabei werden die Erfolge der ausländischen Betriebsstätte bereits bei erfolgter Feststellung dem inländischen Stammhaus bzw. den Gesellschaftern vollständig zugerechnet (Feststellungsprinzip).[24] Die Ermittlung des Gewinnanteils aus der ausländischen Betriebsstätte erfolgt nach Maßgabe des § 4 Abs. 1 EStG und daher nach den deutschen Vorschriften.[25]

Übertragen auf den konkreten Sachverhalt (s. Abbildung 1) bedeutet dies Folgendes: Der deutsche Investor als Gesellschafter der GmbH & Co KG und natürliche Person (§ 1 BGB) ist aufgrund seines Wohnsitzes sowie seines gewöhnlichen Aufenthalts in Deutschland unbeschränkt steuerpflichtig (§ 1 Abs. 1 Satz 1 EStG i. V. m. §§ 8, 9 AO). Folglich werden die Einkünfte aus der Beteiligung an der inländischen GmbH & Co KG wie auch der bulgarischen Betriebsstätte anteilig (zu 100 Prozent) dem deutschen Investor mit dem Ziel der Besteuerung zugeordnet[26] und zählen unmittelbar zu seinem Welteinkommen.[27] Vorbehaltlich der Anwendung bilateraler Maßnahmen (s. Kap. 2.1.2) unterliegen sie im Entstehungsjahr der Einkommensteuer in Deutschland (§ 1 Abs. 1 i. V. m. § 2 EStG).[28]

Im weiteren Verlauf stellt sich die Frage, wie Verluste, die im Zusammenhang mit einer ausländischen Betriebsstätte entstehen, steuerlich behandelt werden. Die Berücksichtigung von Verlusten ist im Steuersystem verankert. Somit finden auch ausländische Verluste (mit Ausnahme von § 2a EStG), die in das Welteinkommen einer unbeschränkt steuerpflichtigen Person einbezogen werden, im Inland Berücksichtigung,[29] vorbehaltlich der einschlägigen Regelungen des anwendbaren DBA (s. Kap. 2.1.2). Abgesehen von einer möglichen Freistellung der negativen Einkünfte im Rahmen des DBA ist kurz auf die allgemeinen Vorschriften zur Verlustverrechnung nach deutschem Recht einzugehen. Dies ist vor allem deshalb relevant, weil der Fall einer Geschäftsausweitung in Deutschland (s. Abbildung 10, S. XI) als Vergleichsszenario in der Untersuchungsmethode einbezogen wird. Nach dem Verlustausgleich (§ 2 Abs. 1, 3 EStG)[30] verbleibende negative Einkünfte können im Rahmen des Verlustrücktrags bis zu einem maximalen Betrag von einer Million Euro[31] mit Gewinnen der beiden zuvor liegenden Veranlagungszeiträume verrechnet werden (§ 10d Abs. 1 Satz 1, Satz 2 EStG). Darauf kann auf Antrag verzichtet werden (§ 10d Abs. 1 Satz 6 EStG). Im Rahmen des Verlustrücktrags nicht berücksichtigte Verluste sind mit den in nächsten Veranlagungszeiträumen erzielten positiven Einkünften zu verrechnen (§ 10d Abs. 2 Satz 1 EStG).[32]

[19] Für gewerbliche Mitunternehmerschaften besteht die Pflicht zur Buchführung (§ 140 AO i. V. m. § 238ff. HGB).
[20] Vgl. Querbach (2024), HI2282706.
[21] Die Voraussetzungen für die steuerliche Anerkennung sowie die Ansatzverbote für bestimmte Rückstellungen sind im Wesentlichen in § 5 EStG geregelt.
[22] Vgl. Wassermeyer in: Wassermeyer et al. (2015), Rz. 2.12.
[23] Vgl. BMF-Schreiben v. 26.9.2014 - IV B 5 - S 1300/09/10003, BStBl I S. 125, Tz. 2.2.3; BFH-Urteil v. 16.10.2002 - I R 17/01, BStBl 2003 II S. 631 III, Nr. 3, sublit. aaa.
[24] Vgl. Spengel in: Jacobs (2023), S. 391.
[25] Vgl. BFH-Urteil v. 13.9.1989 - I R 117/87, BStBl 1990 II S. 57, II., Nr. 5, lit. b.
[26] Vgl. Kempermann in: Wassermeyer et al. (2015), Rn. 2.63.
[27] Das Welteinkommensprinzip ist nicht explizit gesetzlich kodifiziert, sondern ergibt sich im Wege eines Umkehrschlusses aus § 2 Abs. 1 Satz 1 Halbsatz 2 EStG.
[28] Vgl. Ditz / Quilitzsch in: Wassermeyer et al. (2018), Rn. 3.32.
[29] Vgl. Cloer in: Elschner et al. (2015), S. 68; Haase (2023), Rn. 194.
[30] Damit ist der Ausgleich der negativen Einkünfte innerhalb derselben Einkunftsart und bei der Aufsummierung der Einkünfte zwischen den Einkunftsarten im Kalenderjahr gemeint.
[31] In dem vorliegenden Fall ist stets von einer Einzelveranlagung auszugehen.
[32] Der Verlustvortrag ist bis zu einem Betrag der Gesamteinkünfte in Höhe von einer Million Euro uneingeschränkt möglich, darüber hinaus zu maximal 70 Prozent (ab Veranlagungszeitraum 2028: 60 Prozent) der nach Abzug der einen Million Euro verbleibenden Gesamteinkünfte (§ 10d Abs. 2 Satz 1 EStG).

Der dargelegte Verlustabzug ist unter anderem bei der Ermittlung des zu versteuernden Einkommens, das die Bemessungsgrundlage der Einkommensteuer bildet (§ 2 Abs. 5 EStG), zu berücksichtigen. Für eine vereinfachte Ermittlung des zu versteuernden Einkommens wird auf Tabelle 9 (s. S. XI) verwiesen. Zur Berechnung der tariflichen Einkommensteuer wird auf Basis des auf volle Euro abgerundeten zu versteuernden Einkommens der progressive Tarif des § 32a Abs. 1 Satz 2 EStG angewendet (für den im Jahr 2025 geltenden Tarif s. Tabelle 11, S. XII). Für die Ermittlung der festzusetzenden Einkommensteuer (s. Tabelle 10, S. XII) ist zudem zu beachten, dass unbeschränkt steuerpflichtigen natürlichen Personen als Mitunternehmern mit Einkünften aus Gewerbebetrieb (§ 15 Abs. 1 Satz 1 Nr. 2 EStG) eine Steuerermäßigung in Form der Anrechnung der Gewerbesteuer auf die tarifliche Einkommensteuer gewährt wird.[33] Diese Steuerermäßigung beläuft sich auf das Vierfache des nach § 14 GewStG festgesetzten anteiligen Gewerbe-Messbetrags (§ 35 Abs. 1 Satz 1 Nr. 2 EStG) und ist höchstens auf die tatsächlich entrichtete anteilige Gewerbesteuer begrenzt (§ 35 Abs. 1 Satz 5 EStG).[34] Auf die festzusetzende Einkommensteuer wird ein Solidaritätszuschlag in Höhe von 5,5 Prozent (§ 4 Abs. 1 Satz 1 SolZG) erhoben (§ 1 Abs. 1, § 3 Abs. 1 Nr. 1 SolZG).[35]

Ferner ist die Personengesellschaft anders als im Rahmen der Einkommensteuer Schuldnerin der Gewerbesteuer (§ 5 Abs. 1 Satz 3 GewStG). Die der ausländischen Betriebsstätte zuzuordnenden Gewinne unterliegen nicht der Gewerbesteuer.[36] Vor diesem Hintergrund und unter Berücksichtigung der dargestellten Ermäßigungsvorschrift spielt die Ermittlung der Gewerbesteuer in der Untersuchung betragsmäßig eine untergeordnete Rolle und wird im Folgenden nicht weiter dargelegt. Für die schematische Ermittlung der Gewerbesteuer wird auf Tabelle 12 (s. S. XIII) verwiesen.

Bulgarien

Entscheidend für die Begründung der beschränkten Steuerpflicht im Quellenstaat (Bulgarien) ist die Beantwortung der Frage, ob die Betätigung ausschließlich nach der Definition des bulgarischen Steuerrechts als Betriebsstätte einzustufen ist.[37] In diesem Zusammenhang erkennt das bulgarische Steuerrecht den Vorrang von internationalen Verträgen, insbesondere von DBA, an (Art. 13 KStG-BG). Maßgeblich für die Beurteilung der Frage nach dem Betriebsstättenstatus sind somit die Bestimmungen des DBA mit dem jeweiligen Staat.[38] Im vorliegenden Fall ist annahmegemäß vom Vorliegen einer Betriebsstätte auszugehen, weshalb im nächsten Schritt auf ihre steuerliche Behandlung nach bulgarischem Steuerrecht einzugehen ist.

Die Frage, welche Steuern und in welcher Höhe im Ausland anfallen, richtet sich lediglich nach der Rechtsform der deutschen Stammeinheit, da die Betriebsstätte kein eigenständiges Steuersubjekt darstellt.[39] In diesem Zusammenhang stellt sich insbesondere die Frage, wie die deutsche GmbH & Co KG nach bulgarischem Steuerrecht zu qualifizieren ist. Im Gegensatz zur deutschen Besteuerung der Personengesellschaften nach dem Mitunternehmerkonzept richtet sich die Besteuerung in Bulgarien nach dem Kapitalgesellschaftskonzept.[40] Der Transparenzgrundsatz in Deutschland ändert die Behandlung der Gesellschaft nach bulgarischem Recht nicht.[41] Somit gilt die deutsche Personengesellschaft als eine nicht in Bulgarien ansässige, juristische und somit steuerpflichtige Person, die Einkünfte aus einer Quelle in Bulgarien erzielt (Art. 4 Abs. 2 KStG-BG). Ihre steuerrechtliche Persönlichkeit wird für bulgarische Steuerzwecke fingiert.[42] Daher wird die deutsche Gesellschaft der Körperschaftsteuer gemäß Art. 5 Abs. 1 KStG-BG und nicht der bulgarischen Einkommensteuer der natürlichen Personen bzw. Einzelunternehmer unterworfen.[43] Die GmbH & Co KG unterliegt mit den von ihr erzielten Einkünften im Rahmen ihrer wirtschaftlichen Tätigkeit durch die bulgarische Betriebsstätte mit den ihr zuzuordnenden Gewinnen der beschränkten Steuerpflicht nach dem bulgarischen Körperschaftsteuergesetz (Art. 2 Abs. 1 Nr. 2 und Art. 12 Abs. 1 KStG-BG).

Es existieren keine besonderen Vorschriften für die Besteuerung von ausländischen Personen, die in Bulgarien über eine Betriebsstätte wirtschaftlich tätig sind. Für sie gelten daher die allgemeinen gesetzlichen Bestimmungen, die in der weiteren Darstellung kurz erörtert werden.[44] Für ausländische juristische Personen besteht eine Buchführungspflicht über die Tätigkeit ihrer Betriebsstätte in Bulgarien. Diese ergibt aus ihrer Einordnung als *Unternehmen* im Sinne des bulgarischen Gesetzes über die Rechnungslegung (RLG-BG, Art. 2 Nr. 7 RLG-BG). Die Ermittlung des handelsrechtlichen Ergebnisses erfolgt unter Einhaltung der Bestimmungen des RLG-BG. Das handelsrechtliche Ergebnis wird in der Landeswährung Bulgariens, Lew (BGN),[45] ermittelt, unabhängig davon, in welcher Währung die Geschäftsvorfälle

[33] Vgl. BMF-Schreiben v. 3.11.2016 - IV C 6 - S 2296-a/08/10002 :003, BStBl 2016 I S. 1187, Rn. 1.

[34] Der Ermäßigungshöchstbetrag wird nicht beachtet, da im vorliegenden Fall nur Einkünfte aus Gewerbebetrieb vorliegen.

[35] Die Freigrenzen des Solidaritätszuschlags werden im Folgenden nicht thematisiert.

[36] Dies ergibt sich aus dem deklarierten Inlandsbezug der Gewerbesteuer unter § 2 Abs. 1 Satz 3 GewStG und wird nochmal durch die unter § 9 Nr. 3 GewStG kodifizierte Kürzungsvorschrift verdeutlicht, vgl. dazu Ditz / Quilitzsch in: Wassermeyer et al. (2018), Rn. 3.60.

[37] Vgl. Puls in Wassermeyer et al. (2018), Rn. 2.2. Vgl. auch Spengel in: Jacobs (2023), S. 384.

[38] Vgl. Bulgarische Nationale Steuerverwaltung (2012), Erläuterung Nr. 24-96, S. 2.

[39] Vgl. dazu Spengel in: Jacobs (2023), S. 386.

[40] Die bulgarische Personengesellschaft ist nach bulgarischer Rechtsordnung für Steuerzwecke eine juristische Person (Art. 63 Abs. 3 HGB-BG i. V. m. Art. 64 Abs. 1 Nr. 1, Nr. 2 HGB-BG und Art. 2 Abs. 1 Nr. 1 KStG-BG).

[41] Vgl. Bulgarische Nationale Steuerverwaltung (2023), Erläuterung Nr. 07-00-502, S. 3.

[42] Vgl. Bulgarische Nationale Steuerverwaltung (2021a), S. 2.

[43] Vgl. Bulgarische Nationale Steuerverwaltung (2014), Erläuterung Nr. 5304654, S. 2.

[44] Vgl. Bulgarische Nationale Steuerverwaltung (2023), Erläuterung Nr. 07-00-502, S. 3.

[45] Der Wechselkurs des Lew zum Euro ist fest: 1 BGN = 0,51129 EUR;

abgewickelt wurden.[46] Das handelsrechtliche Ergebnis ist für die Ermittlung des steuerlichen Ergebnisses maßgeblich. Es wird gemäß den Bestimmungen des KStG-BG angepasst (Art. 18 Abs. 1 KStG-BG). Dabei ist zu beachten, dass alle handelsrechtlich gebildeten Rückstellungen nach bulgarischem Steuerrecht nicht ansetzbar sind (Art. 38 Abs. 1 KStG-BG). Bemessungsgrundlage für die Körperschaftsteuerfeststellung stellt das positive steuerliche Ergebnis bzw. der steuerliche Gewinn dar (Art. 18 Abs. 2, Art. 19 KStG-BG). Der Körperschaftsteuersatz beträgt 10 Prozent (Art. 20 KStG-BG). Der Veranlagungszeitraum beläuft sich auf ein Kalenderjahr (Art. 21 Abs. 1 KStG-BG). Die Entnahmen aus der bulgarischen Betriebsstätte werden nicht mit Quellensteuer in Bulgarien belastet (Art. 194 Abs. 1 Nr. 1 KStG-BG).

Ferner wird im Rahmen der Veranlagung der deutschen Personengesellschaft als Spitzeneinheit in Bulgarien das Recht gewährt, die steuerlichen Verluste der Betriebsstätte innerhalb eines Zeitraums von fünf Jahren der Reihe ihres Entstehens nach bis zur vollständigen Ausschöpfung vorzutragen (Art. 70 Abs. 1 KStG-BG). Es ist zu beachten, dass der Verlustvortrag nicht automatisch gewährt wird. Vielmehr stellt er ein Wahlrecht dar, das innerhalb des vom Gesetz festgelegten Zeitraums auszuüben ist, nämlich in dem ersten Jahr nach der Realisierung des steuerlichen Verlusts, in dem ein positives steuerliches Ergebnis vor dem Verlustabzug erzielt wird (Art. 70 Abs. 2 KStG-BG). Andernfalls geht das Recht auf Verlustvortrag unwiderruflich verloren.[47] Ist der vorgetragene Verlust kleiner als das positive steuerliche Ergebnis vor dem Verlustabzug, wird der Verlustvortrag vollständig ausgeschöpft (Art. 71 Abs. 1 Satz 2 KStG-BG). Es ist darauf hinzuweisen, dass für jeden neu entstandenen Verlust die fünfjährige Frist mit dem Jahr beginnt, das auf das Jahr der Entstehung folgt (Art. 72 Abs. 1 Satz 2 KStG-BG).

2.1.2 Abkommensrecht

Wie die vorangegangene Analyse zeigt, führen die unbeschränkte Steuerpflicht in Deutschland und die beschränkte Steuerpflicht in Bulgarien zu einer Überschneidung der ertragsteuerlichen Ansprüche der beiden Staaten. An dieser Stelle ist die Frage aufzuwerfen, ob und wie diese drohende Doppelbesteuerung durch das bestehende DBA zwischen Deutschland und Bulgarien aufzulösen ist.

Der Abkommensschutz kann nur dann in Anspruch genommen werden, wenn der Anwendungsbereich des DBA D/BG eröffnet ist.[48] Im Folgenden wird daher in einem ersten Schritt geprüft, ob das DBA D/BG aus räumlicher, zeitlicher, sachlicher und persönlicher Sicht angewendet werden kann. In dem Fall sind der räumliche (Art 3 Abs. 1 lit. a und lit. b DBA D/BG)[49] und der zeitliche Aspekt (Art. 30 DBA D/BG) unproblematisch. Von dem Abkommen sind unter anderem die Einkommensteuer, Gewerbesteuer und die darauf erhobenen Zuschläge in Deutschland sowie die Körperschaftsteuer in Bulgarien erfasst (Art. 2 Abs. 3 lit. a, lit. b DBA D/BG). Die in diesem Fall betroffenen Steuern sind folglich vom DBA D/BG sachlich umfasst. Ferner ist der persönliche Anwendungsbereich einer Prüfung zu unterziehen. Bei der Anwendung des DBA muss jeder Staat die begriffliche Auslegung einer „steuerpflichtigen Person" nach seinem innerstaatlichen Recht anwenden.[50] Für den Erhalt der Abkommensberechtigung muss die GmbH & Co KG „eine in einem Vertragsstaat ansässige Person"[51] darstellen, d. h., sie muss in dem Ansässigkeitsstaat, in diesem Fall Deutschland, der unbeschränkten Steuerpflicht unterliegen.[52] Wie bereits erläutert, ist dies nicht der Fall. Die GmbH & Co KG mit Sitz und Geschäftsleitung in Deutschland kann deshalb nach dominierender Auffassung in der Fachliteratur keinen Abkommensschutz von sich aus beanspruchen.[53]Als abkommenberechtigt ist vielmehr der Gesellschafter anzusehen, soweit er in Bulgarien oder Deutschland ansässig ist.[54] Folglich ist im vorliegenden Fall der Anwendungsbereich des DBA D/BG eröffnet.

Des Weiteren ist zu beurteilen, welcher Einkunftsartikel des Abkommens zur Anwendung kommt. Für diese Beurteilung ist ausschließlich die in dem Abkommen kodifizierte begriffliche Auslegung der Betriebsstätte entscheidend.[55] Im Gegensatz zu der Begriffsauslegung nach dem innerstaatlichen Steuerrecht konzentriert sich der DBA-Begriff der Betriebsstätte auf die klare Zuweisung und Abgrenzung der Besteuerungsrechte auf die Gewinne des jeweiligen Unternehmens zwischen dem Quellen- und dem Wohnsitzstaat.[56] An dieser Stelle wird auf eine detaillierte Auslegung des DBA-Betriebsstättenbegriffs verzichtet, da davon ausgegangen wird, dass die Tätigkeit in Bulgarien als Betriebsstätte im Abkommensrecht zu qualifizieren ist. Es sei jedoch darauf hingewiesen, dass der DBA-Begriff der Betriebsstätte in der Regel deutlich restriktiver gefasst ist als jener nach den nationalen Steuerrechtsordnungen der Vertragsstaaten.[57] Folglich kann es in bestimmten Fällen vorkommen, dass der Betriebsstättenbegriff nach dem bulgarischen Steuerrecht

1 EUR = 1,95583 BGN.

[46] Vgl. Dimitrova (2024), S. 376.

[47] Vgl. Oberster Verwaltungsgericht Bulgariens, Beschluss Nr. 2075 v. 4.3.2025 in der verwaltungsrechtlichen Angelegenheit Nr. 8148/2024, I, S. 2.

[48] Vgl. Haase (2023), Rn. 592.

[49] Vgl. Becker in: Wassermeyer (2025), Art. 3 DBA D/BG 2010, Rn. 15.

[50] Vgl. Becker in: Wassermeyer (2025), Art. 3 DBA D/BG 2010, Rn. 10. Vgl. auch Spengel in: Jacobs (2023), S. 412: Dazu gibt es in der Literatur konträre Ansichten, die dem innerstaatlichen Recht bei der Auslegung der Begriffe nur eine sekundäre Bedeutung einräumen.

[51] Art. 4 Abs. 1 DBA D/BG

[52] Vgl. Becker in: Wassermeyer (2025), Art. 1 DBA D/BG 2010, Rn. 11. Die Personengesellschaft mag unter dem Begriff der Person als „sonstige Personenvereinigung" gemäß Art. 3 Abs. 1 lit. d DBA D/BG fallen, ist aber als solche nicht ansässig.

[53] Vgl. BMF-Schreiben v. 26.9.2014 - IV B 5 - S 1300/09/10003, BStBl I S. 1258, Tz. 2.1.1. Vgl. auch Weggenmann / Nehls in: Vogel und Lehner (2021), Art. 1 OECD-Musterabkommen (MA) 2017, Rn. 36.

[54] Vgl. Becker in: Wassermeyer (2025), Art. 3 DBA D/BG 2010, Rn. 7.

[55] Vgl. Looks und Heinsen (2017), Rn. 181.

[56] Vgl. Spengel in: Jacobs (2023), S. 385; weiterführend zum Betriebsstättenbegriff bspw. Looks und Heinsen (2017), S. 45–101.

[57] Vgl. Spengel in: Jacobs (2023), S. 385.

durch das Abkommen eingeschränkt wird.

Die Unternehmensgewinne der GmbH & Co KG können in Deutschland besteuert werden, es sei denn, das Unternehmen übt seine Tätigkeit in Bulgarien durch eine Betriebsstätte aus (Art. 7 Abs. 1 Satz 1 Halbsatz 2 DBA D/BG). In einem solchen Fall ist die GmbH & Co KG ausschließlich in diesem ausländischen Quellenstaat (Bulgarien) mit den dieser Betriebsstätte zuzurechnenden Gewinnen der Besteuerung zu unterwerfen (Art. 7 Abs. 1 Satz 2 DBA D/BG), sog. Betriebsstättenprinzip.[58] Demnach werden die der Betriebsstätte zugehörigen Gewinne in Bulgarien der Besteuerung unterworfen (s. zu der Besteuerung in Bulgarien Kap. 2.1.1.2). Die Gewinne, die der bulgarischen Betriebsstätte zuzurechnen sind, werden durch die Anwendung der Freistellungsmethode von der deutschen Besteuerung befreit (Art. 7 Abs. 1 Satz 2 i. V. m. Satz 1 Halbsatz 2 und Art. 22 Abs. 1 lit. a DBA D/BG). Das DBA schränkt den bereits beschriebenen Steueranspruch Deutschlands ein. Zu achten ist hierbei insbesondere darauf, dass das DBA die Freistellung nur beim Vorliegen einer aktiven Tätigkeit gewährt (Art. 22 Abs. 1 lit. c DBA D/BG).

Das DBA gestattet jedoch, dass die in Deutschland freigestellten ausländischen Einkünfte im Rahmen der Steuersatzfestsetzung berücksichtigt werden (Art. 22 Abs. 1 lit. d DBA D/BG).[59] Diese werden dementsprechend unter Anwendung des positiven Progressionsvorbehalts (im weiteren Verlauf PVB) für die Zwecke der Berechnung des im Inland anzuwendenden Einkommensteuersatzes in der Summe der Gesamteinkünfte einbezogen (§ 32b Abs. 1 Satz 1 Nr. 3 EStG).[60] Hierbei sind ausschließlich die nationalen Gewinnermittlungsvorschriften heranzuziehen.[61] Bei der Ermittlung des durchschnittlichen Steuersatzes findet der Verlustabzug nach § 10d EStG keine Berücksichtigung (H 32b „Allgemeines" EStH).

Im Falle der abkommensrechtlichen Freistellung des ausländischen Betriebsstättengewinns im Ansässigkeitsstaat des Steuerpflichtigen obliegt die Entscheidung, wie mit einem Verlust aus dieser Betriebsstätte zu verfahren ist, grundsätzlich den Vertragsstaaten.[62] In Bezug auf die steuerrechtliche Behandlung von Verlusten aus Betriebsstätten, deren Einkünfte im Rahmen der DBA freigestellt sind, stützt sich der Bundesfinanzhof (BFH) gemäß seiner konstanten Rechtsprechung auf die Symmetriethese.[63] Demzufolge werden die Verluste der ausländischen Betriebsstätte als Teil der freigestellten Einkünfte im Inland nicht berücksichtigt.[64] In diesem Kontext war die Frage der Nichtabziehbarkeit finaler Betriebsstättenverluste über mehrere Jahre hinweg umstritten und durch eine inkonsistente Rechtsprechung[65] geprägt.[66] In aktueller Rechtssache lehnt der Europäische Gerichtshof eine für die Vertragsstaaten verpflichtende Berücksichtigung von laufenden oder abschließenden Verlusten einer ausländischen Betriebsstätte ab.[67] Diesbezüglich ist darauf hinzuweisen, dass die Nichtberücksichtigung der (finalen) Verluste von der ausländischen Betriebsstätte im Rahmen der Veranlagung in Deutschland keine unzulässige Ungleichbehandlung zwischen einer inländischen und einer ausländischen Betriebsstätte darstellt. Ein solches steuerliches Vorgehen verstößt daher nicht gegen die Niederlassungsfreiheit (Art. 49 und 54 AEUV).[68] Dies lässt sich damit begründen, dass der Fall einer Betriebsstätte im Ausland nicht mit dem Bestehen einer inländischen Betriebsstätte zu vergleichen ist, da nur im erstgenannten Fall eine Freistellung der Gewinne im Inland stattfindet.[69]

Bei der Besteuerung in Deutschland können die Betriebsstättenverluste analog zu den Gewinnen innerhalb des negativen PVB eine Berücksichtigung finden (§ 32b Abs. 1 Satz 1 Nr. 3 EStG).[70] Im Rahmen des negativen PVB wird das Gesamteinkommen bei der Anwendung des progressiven Einkommensteuertarifs um die ausländischen Betriebsstättenverluste reduziert. Dadurch kann der anzuwendende Einkommensteuersatz ermäßigt werden.[71] Der negative PVB kann auch einen Steuersatz von null Prozent als Folge haben.[72]

2.1.3 Qualifikationskonflikte

Bei der Anwendung des DBA seitens der beiden Vertragsstaaten kann es zu Qualifikationskonflikten kommen, die aus der unterschiedlichen Behandlung von Personengesellschaften in beiden Vertragsstaaten – als intransparent im Quellenstaat und als transparent im Ansässigkeitsstaat – resultiert.[73]

Bei der Anwendung des DBA aus Sicht Bulgariens ist in diesem Zusammenhang der Umstand zu beachten, dass die deutsche Personengesellschaft für die Zwecke der Abkommensberechtigung gegenüber den bulgarischen Steuerbehörden ihren Status als steuerpflichtige Person in Deutschland durch ein offizielles Dokument der deutschen Steuerverwaltung nachweisen soll. Die fehlende Steuerpersönlichkeit der Gesellschaft in Deutschland steht der

[58] Vgl. Becker in: Wassermeyer (2025), Art. 7 DBA D/BG 2010, Rn. 14.
[59] Nach der Rechtsprechung ergibt sich der PVB bereits durch die Freistellung in dem Verteilungsartikel; s. BFH-Urteil v. 14.7.2010 - X R 37/08, BStBl 2011 II S. 628, Rn. 20.
[60] Siehe dazu Becker in: Wassermeyer (2025), Art. 22 DBA D/BG 2010, Rn. 3. Bei passiver Tätigkeit ist der PVB gemäß § 32b Abs. 1 S. 2 EStG verboten.
[61] Vgl. Spengel in: Jacobs (2023), S. 396.
[62] Vgl. OECD (2019), C(23)-23, Art. 23, Rn. 44.
[63] Vgl. dazu bspw. BFH-Urteil v. 5.2.2014 - I R 48/11, BB 2014 S. 1046 Nr. 18, Leitsatz, Nr. 1.
[64] Vgl. Ditz/Quilitzsch in Wassermeyer et al. (2018), S. 145, Rn. 3.66.
[65] Vgl. dazu bspw. EuGH-Urteil v. 13.12.2005 - C.-446/03, NWB ZAAAB-79456, Rn. 59; EuGH-Urteil v. 15.5.2008 - C.-414/06, BStBl 2009 II S. 692, Rn. 53; EuGH-Urteil v. 17.12.2015 - C-388/14, BStBl 2016 II S. 362, Rn. 53; BFH-Urteil v. 22.2.2017 - I R 2/15, BStBl 2017 II S. 709; Rn. 38; EuGH-Urteil v. 12.6.2018 - C-650/16, NWB PAAAG-87352, Rn. 66.
[66] Vgl. Schulz-Trieglaff (2022), S. 865.
[67] Vgl. EuGH-Urteil v. 22.9.2022 - C-538/20, NWB QAAAJ-22773, Rn. 29.
[68] Vgl. EuGH-Urteil v. 22.9.2022 - C-538/20, NWB QAAAJ-22773, Rn. 27f.
[69] Vgl. EuGH-Urteil v. 17.12.2015 - C-388/14, BStBl 2016 II S. 362, Rn. 64f.
[70] Dies gilt entsprechend nur im Falle einer aktiven Tätigkeit, andernfalls greift das Verbot nach § 32b Abs. 1 S. 2 EStG.
[71] Vgl. Spengel in: Jacobs (2023), S. 398.
[72] Vgl. BFH-Urteil v. 25.5.1970 - I R 109/68, BStBl 1970 II S. 660, Abschnitt 7 unter „Gründe".
[73] Vgl. Trencsik in: Elschner et al. (2015), S. 245.

Anwendung des Abkommens jedoch nicht entgegen.[74] Im Rahmen des DBA D/BG wird diese Problematik seitens des Quellenstaats Bulgarien mithilfe der Regelung in Art. 4 Abs. 3 DBA D/BG gelöst.[75] Die Regelung ist anwendbar, da Deutschland die Einkünfte den Gesellschaftern zuweist und Bulgarien dagegen auf die Gesellschaft abstellt, die in Deutschland als solche nicht steuerpflichtig ist.[76] Werden dementsprechend Einkünfte über die transparente GmbH & Co KG erzielt, wird ihre Ansässigkeit für die Zwecke des Abkommens fingiert.[77] Dies greift nur, wenn die Gesellschafter der Personengesellschaft, denen die Einkünfte zugeordnet werden, in Deutschland ansässig sind,[78] was in dem vorliegenden Fall zutrifft (Art. 4 Abs. 3 DBA D/BG).

Im Rahmen der vorliegenden Konstellation wird von dem Vorliegen einer Betriebsstätte ausgegangen. Es sollte jedoch darauf hingewiesen werden, dass sich in der Praxis in bestimmten Fällen durchaus Qualifikationskonflikte bei der Auslegung ergeben können.[79] Die potenziellen Qualifikationskonflikte werden im nachfolgenden Fall des Auslandsengagements in Form einer bulgarischen Personengesellschaft evidenter.

2.2 Gründung einer Personengesellschaft

Als Ausgangspunkt der zweiten Gestaltungsmöglichkeit des Auslandsengagements dient ein deutscher Investor, der eine gewerblich tätige Personengesellschaft nach bulgarischem Recht in Bulgarien gründet. In diesem Zusammenhang werden im Folgenden die steuerlichen Besonderheiten erörtert, die sich aus dieser konkreten Konstellation (s. Abbildung 2) ergeben.

Der deutsche Investor erwirtschaftet annahmegemäß ausländische aktive und inländische gewerbliche Einkünfte. Hinsichtlich der steuerlichen Behandlung der Inlandseinkünfte aus der inländischen Tätigkeit des deutschen Investors wird auf die Ausführungen zu den allgemeinen Regelungen nach dem deutschen Steuerrecht im Kapitel 2.1.1.1 verwiesen.

2.2.1 Nationales Recht

Der Sachverhalt ist, analog zu dem vorangegangenen Fall, systematisch aus der Perspektive der beiden nationalen Steuerrechte Deutschlands und Bulgariens zu betrachten und wird anschließend im Rahmen des Abkommenrechts gewürdigt.

Deutschland

Die Beantwortung der Frage nach der steuerlichen Behandlung des Sachverhalts seitens Deutschlands setzt zunächst die Einordung des bulgarischen Rechtsgebildes als Mitunternehmerschaft oder Körperschaft nach deutschem Verständnis voraus. Dafür ist ausschließlich das deutsche Steuerrecht maßgeblich. Es ist ein Rechtstypenvergleich durchzuführen. Die zivil- oder steuerrechtliche Behandlung in dem ausländischen Staat ist irrelevant.[80] Als Orientierungshilfen zur Durchführung des Rechtstypenvergleichs können die vorhandenen Qualifizierungen von bestimmten ausländischen Rechtsgebilden der Finanzverwaltung, insbesondere die Betriebsstätten-Verwaltungsgrundsätze herangezogen werden.[81] Infolgedessen werden die bulgarischen Personengesellschaften ungeachtet ihrer Stellung als juristische Personen nach bulgarischem Recht, in Deutschland als deutsche Personengesellschaften behandelt.[82] Eine zusätzliche Durchführung der zweistufigen Methodik des Rechtstypenvergleichs[83] ist in der Praxis nicht weiter notwendig.[84] Auf dieser Grundlage wird im Folgenden auf ihre Anwendung verzichtet.

Darüber hinaus muss überprüft werden, ob die Gesellschaft als eine gewerbliche (§ 15 Abs. 1 S. 1 Nr. 2 EStG) oder nicht gewerbliche (§ 13 Abs. 7 EStG, § 18 Abs. 4 S. 2 EStG und § 21 EStG) Personengesellschaft zu werten ist. Dafür müssen die Tatbestandsmerkmale eines Gewerbebetriebs[85] erfüllt sein,[86] was im vorliegenden Sachverhalt annahmegemäß der Fall ist.

Nachfolgend ist die Besteuerung des ausländischen Rechtsgebildes in Deutschland zu untersuchen. Dessen Gesellschafter, der deutsche Investor, ist gemäß den Ausführungen im Kapitel 2.1.1.1 mit seinen ausländischen und inländischen Einkünften in Deutschland unbeschränkt einkommensteuerpflichtig (§ 1 Abs. 1 Satz 1 EStG), sofern das anwendbare DBA keine Befreiung von der inländischen Besteuerung vorsieht (zum Abkommensrecht s. Kap. 2.2.3).

Der BFH geht in ständiger Rechtsprechung davon aus, dass eine Beteiligung an einem gewerblich tätigen Rechtsgebilde mit ausländischer Ansässigkeit, die einer Personengesellschaft nach deutschem Recht entspricht, bei den inländischen Gesellschaftern zu Einkünften aus Gewerbebetrieb gemäß § 15 Abs. 1 Satz 1 Nr. 2 EStG führt. Diese steuerliche Vorgehensweise gilt unabhängig davon, ob der ausländische

[74] Vgl. Bulgarische Nationale Steuerverwaltung (2021a), S. 2.

[75] So auch Trencsik in: Elschner et al. (2015), S. 245; Weggenmann / Nehls in: Vogel und Lehner (2021), Art. 1 OECD-MA 2017, Rn. 79b; Die Regelung entspricht Art. 1 Abs. 2 OECD-MA 2017.

[76] So auch Weggenmann / Nehls in: Vogel und Lehner (2021), Art. 1 OECD-MA 2017, Rn. 66a.

[77] Vgl. Weggenmann / Nehls in: Vogel und Lehner (2021), Art. 1 OECD-MA 2017, Rn. 65.

[78] Basierend auf den Ausführungen in: Becker in: Wassermeyer (2025), Art. 3 DBA D/BG 2010, Rn. 7; vgl. dazu auch BMF-Schreiben v. 26.9.2014 - IV B 5 - S 1300/09/10003, BStBl I S. 1258, Tz. 2.1.1.

[79] Vgl. Spengel in: Jacobs (2023), S. 384.

[80] Vgl. BMF-Schreiben v. 26.9.2014 - IV B 5 - S 1300/09/10003, BStBl I S. 1258, Tz. 1.2.

[81] Vgl. Rupp (2025), HI2596048; vgl. auch Spengel in: Jacobs (2023), S. 332.

[82] Siehe dazu BMF-Schreiben v. 24.12.1999, IV B 4 - S 1300 - 111/99, BStBl 1999 I S. 1076, Tab. 2, Bulgarien, Nr. 1–2, BMF-Schreiben v. 26.9.2014, IV B 5 - S 1300/09/10003, BStBl I S. 1258, Anlage. So auch Weggenmann / Nehls in: Vogel und Lehner (2021), Art. 1 OECD-MA 2017, Rn. 79d.

[83] Weiterführend zum Rechtstypenvergleich siehe bspw. Schnittker in: Wassermeyer et al. (2015), S. 155–193.

[84] Vgl. Schnittker in: Wassermeyer et al. (2015), Rn. 3.6.

[85] Ein Gewerbebetrieb zeichnet sich durch Selbständigkeit, Nachhaltigkeit, Gewinnerzielungsabsicht und Beteiligung am allgemeinen wirtschaftlichen Verkehr aus – mit Ausnahme von Tätigkeiten in den Bereichen der Land- und Forstwirtschaft, selbstständiger Arbeit oder Vermögensverwaltung (§ 15 Abs. 2 EStG).

[86] Vgl. Spengel in: Jacobs (2023), S. 407.

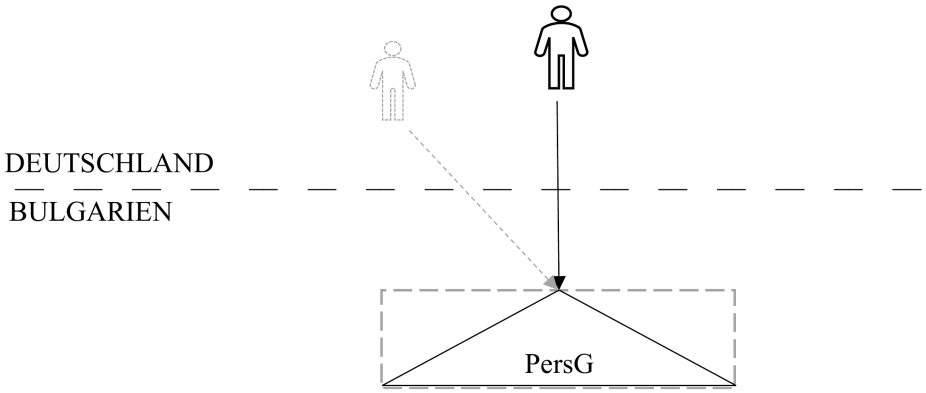


Abbildung 2: Gründung einer Personengesellschaft in Bulgarien (Schematische Darstellung)

Zur Wahrung der Vergleichbarkeit zwischen den Fällen insbesondere im Rahmen der Untersuchung wird unterstellt, dass der Gewinn bzw. Verlust gemäß Gesellschaftsvertrag zu 100 % dem deutschen Investor zuzuordnen ist. Weitere gesonderte Hinweise dazu erfolgen nicht.

nach dem Kapitalgesellschaftskonzept besteuerte Rechtsträger eine Gewinnausschüttung tatsächlich beschlossen, ausgeführt oder ausgezahlt hat.[87] Irrelevant ist auch, ob die jeweilige Gesellschaft in ihrem Sitzstaat als juristische Person behandelt wird und einer Körperschaftsteuer unterliegt.[88]

Der anteilige Gewinn des inländischen Gesellschafters wird auf Basis jenes Betrags berechnet, der sich aus dem im Gesellschaftsvertrag kodifizierten Beteiligungsprozentsatz ergibt. Dabei ist die Handelsbilanz der ausländischen Gesellschaft in die Steuerbilanz unter Beachtung der deutschen Normen (§ 4 ff. EStG) zu überführen. Der sich am Ende daraus ergebende Betrag ist der „Gewinnanteil des Gesellschafters" im Sinne des § 15 Abs. 1 S. 1 Nr. 2 EStG.[89]

Im Verlustfall sind die allgemeinen steuerlichen Regelungen zur Behandlung von Verlusten, die im Kapitel 2.1.1.1 erläutert wurden, grundsätzlich anzuwenden, wenn das Besteuerungsrecht Deutschlands in Bezug auf diese ausländischen Einkünfte durch das anwendbare DBA nicht eingeschränkt ist. Eine eingehende Prüfung dieser Frage hat separat zu erfolgen (s. Kap. 2.2.3).

Gewebebetriebe unterliegen der Gewerbesteuerpflicht, nur wenn sie im Inland betrieben werden (§ 2 Abs. 1 Satz 1 GewStG). Die Gewinne einer ausländischen Personengesellschaft unterliegen daher nicht der Gewerbesteuer, es sei denn, es besteht eine inländische Betriebsstätte,[90] was hier jedoch nicht der Fall ist.

Aus deutscher steuerlicher Sicht wird die bulgarische Personengesellschaft im Ergebnis wie eine ausländische Betriebsstätte besteuert.[91] In Bezug auf die weiteren Steueraspekte im Rahmen der Besteuerung des deutschen Investors wird auf die Ausführungen im Kapitel 2.1.1.1 verwiesen.

Bulgarien

Die Behandlung der bulgarischen Personengesellschaften nach dem Kapitalgesellschaftskonzept macht es notwendig, im Rahmen der Besteuerung in Bulgarien die Ebene der Gesellschaft und der Gesellschafter separat zu betrachten.[92]

Gesellschaftsebene

Die bulgarische Personengesellschaft mit Sitz und Geschäftsleitung in Bulgarien unterliegt der unbeschränkten Steuerpflicht mit ihren weltweiten Einkünften in Bulgarien (Art. 3 Abs. 2 KStG-BG). Erfolgt keine Gewinnausschüttung, handelt es sich folglich aus der Sicht des bulgarischen Steuerrechts um einen rein inländischen Sachverhalt.[93] In diesem Fall wird die Personengesellschaft in Bulgarien laufend nach den Bestimmungen des KStG-BG besteuert. Eine Anwendung des DBA in Bezug auf die Besteuerung der Gesellschaft in Bulgarien ist nicht erforderlich.[94]

Zur Vermeidung von Wiederholungen wird hinsichtlich der Besteuerung einer Personengesellschaft nach bulgarischem Steuerrecht im Gewinn- und Verlustfall auf Kapitel 2.1.1.2 verwiesen.

Gesellschafterebene

Im Gegensatz zur Wertung nach dem deutschen Steuersystem wird der Gewinn aus der Beteiligung an der bulgarischen Personengesellschaft als Dividende im Gesetz aufgefasst (§ 1 Nr. 4 lit. b der Zusatzbestimmungen des KStG-BG). Erfolgt eine Gewinnausschüttung von der bulgarischen Personengesellschaft an den deutschen Investor, wird eine beschränkte Steuerpflicht des deutschen Investors in Bulgarien begründet (Art. 7 EStG-BG, Art. 8 Abs. 3 EStG-BG). Gemäß

[87] Vgl. Möller-Gosoge / Keilen in: Gosch et al. (2024), Art. 7 DBA OECD MA, Rn. 813.
[88] Vgl. BFH-Beschluss v. 4.4.2007 - I R 110/05, BStBl 2007 II S. 521, III, 2, lit. a.
[89] Vgl. Spengel in: Jacobs (2023), S. 416.
[90] Vgl. Weggenmann in: Wassermeyer et al. (2015), S. 295, Rn. 6.42.
[91] Vgl. dazu Wellmann in: Elschner et al. (2015), S. 92.
[92] So auch Spengel in: Jacobs (2023), S. 332.
[93] Vgl. dazu Weggenmann / Nehls in: Vogel und Lehner (2021), Art. 1 OECD-MA 2017, Rn. 68a.
[94] So auch Wellmann in: Elschner et al. (2015), S. 92.

Art. 38 Abs. 1 Nr. 2 lit. a EStG-BG unterliegen Dividenden zugunsten ausländischer natürlicher Personen aus einer Quelle in Bulgarien einer endgültigen Besteuerung in Höhe von fünf Prozent (Art. 46 Abs. 3 EStG-BG). Die Steuer wird auf Ebene der Gesellschaft einbehalten (Art. 65 Abs. 2 EStG-BG).

2.2.2 Abkommensrecht

Im Folgenden ist zu prüfen, wie der Sachverhalt im Rahmen des Abkommensrechts zu würdigen ist. Der Anwendungsbereich des DBA D/BG ist in zeitlicher, räumlicher und sachlicher Hinsicht erfüllt (Prüfung analog zum Kap. 2.1.2). Der deutsche Investor stellt eine in einem der Vertragsstaaten des Abkommens ansässige Person dar (Art. 4 Abs. 1 DBA D/BG). Somit ist der Anwendungsbereich des DBA D/BG in persönlicher Hinsicht ebenfalls erfüllt.

Die zuvor erläuterte Einordnung der bulgarischen intransparenten Personengesellschaft nach deutschem Verständnis als transparent macht sie zu einer steuerlich hybriden Gesellschaft im Kontext des internationalen Steuerrechts. Im vorliegenden Kontext bedeutet „hybrid", dass die bulgarische Personengesellschaft von den beiden Vertragsstaaten unterschiedlich qualifiziert wird.[95]

Diese Problematik erschwert die Anwendung des DBA.[96] Im Rahmen des DBA D/BG[97] wird durch Art. 4 Abs. 3 DBA D/BG der Versuch unternommen, das Problem mit hybriden Rechtsgebilden zu lösen.[98] Aufgrund des Vorliegens einer hybriden Gesellschaft (siehe oben) ist der Anwendungsbereich der Regelung grundsätzlich eröffnet.[99] Es stellt sich die Frage, ob die Vorschrift bei der weiteren Anwendung des DBA in dem konkreten Fall einen Beitrag zur Klärung leistet. Jedoch verhilft die Regelung in Art. 4 Abs. 3 DBA D/BG in diesem konkreten Fall zu keiner Lösung.[100] Insbesondere die Qualifizierung der Einkünfte, die der Gesellschafter über die Personengesellschaft erwirtschaftet, ist mit Zweifelsfragen behaftet.

Die genaue Auslegung des DBA im deutschen Steuerrecht ist durch Uneinigkeit geprägt, die nicht nur innerhalb der Literatur, sondern zudem auch zwischen Teilen der Literatur und der Finanzverwaltung besteht.[101] Nach der in der Literatur vorherrschenden Sichtweise sollte sich Deutschland bei der Behandlung der Einkünfte nach dem Abkommensrecht an der Einordnung der Personengesellschaft im Sitzstaat orientieren. Nach diesem Vorgehen sollten die Gewinne der Personengesellschaft auch in Deutschland als Dividenden behandelt und erst nach erfolgter Ausschüttung der Besteuerung unterworfen werden.[102]

Die Finanzverwaltung nimmt jedoch ausdrücklich eine davon abweichende, an dem Anwenderstaat des DBA orientierte Vorgehensweise an. Somit ist für die deutsche Besteuerung allein die Sichtweise des deutschen Steuerrechts maßgeblich (s. Kap. 2.2.2.1).[103] Die Tatsache, dass die bulgarische Personengesellschaft als juristische Person in Bulgarien grundsätzlich eine nach dem DBA ansässige Person und ein Steuersubjekt darstellt, steht der Zurechnung der Einkünfte auf den Gesellschafter nicht entgegen.[104] Infolge der Anwendung des steuerlichen Verständnisses des Ansässigkeitsstaates des Gesellschafters, Deutschlands, handelt es sich bei der intransparenten gewerblich tätigen bulgarischen Personengesellschaft um eine ausländische Betriebsstätte des unbeschränkt Steuerpflichtigen.[105] Es ist folglich das Betriebsstättenprinzip anzuwenden.[106] Der Anteil an der Personengesellschaft wird somit seitens Deutschlands von der Besteuerung befreit (Art. 7 Abs. 1 Satz 2 i. V. m. Satz 1 Halbsatz 2 und Art. 22 Abs. 1 lit. a DBA D/BG).[107] Die freigestellten Einkünfte sind analog zum Fall der Betriebsstätte (s. dazu Kap. 2.1.2) bei der Veranlagung in Deutschland ausschließlich im Rahmen des positiven bzw. negativen PVB zu berücksichtigen. Der Umstand, dass die Gesellschaft ausländischen Rechts als juristische Person im Sitzstaat gilt, ändert daran nichts.[108]

Im Falle einer tatsächlichen Gewinnausschüttung von der intransparenten bulgarischen Personengesellschaft kommt es zu keiner weiteren Besteuerung in Deutschland, da diese wie eine steuerlich irrelevante Privatentnahme behandelt wird.[109] Da die Gewinnrepatriierung an den deutschen Investor aus deutscher Sicht nicht steuerbar ist, ist auch eine Anrechnung der in Bulgarien auf der Dividende einbehaltenen Kapitalertragsteuer auf die deutsche Einkommensteuer ausgeschlossen.[110] Diese Auffassung wird sowohl von der Finanzverwaltung als auch mehrheitlich im Schrifttum befürwortet.[111]

[95] Vgl. Weggenmann / Nehls in: Vogel und Lehner (2021), Art. 1 OECD-MA 2017, Rn. 17.

[96] Vgl. Weggenmann / Nehls in: Vogel und Lehner (2021), Art. 1 OECD-MA 2017, Rn. 30.

[97] Vgl. Weggenmann / Nehls in: Vogel und Lehner (2021), Art. 1 OECD-MA 2017, Rn. 79b.

[98] Vgl. Trencsik in: Elschner et al. (2015), S. 245.

[99] So auch Jacob und Klein (2014), S. 123.

[100] Vgl. Weggenmann / Nehls in: Vogel und Lehner (2021), Art. 1 OECD-MA 2017, Rn. 68a; Grund dafür ist, dass sich der Artikel lediglich an dem Quellenstaat richtet. Im konkreten Fall besteuert Bulgarien nicht in seiner Rolle als Quellenstaat, sondern betrachtet sich vielmehr als der Staat, in dem die Gesellschaft ansässig ist.

[101] Vgl. Spengel in: Jacobs (2023), S. 419f. Die Diskussion über die abkommensorientierte oder anwenderorientierte Auslegung ist sehr umfangreich und wird daher nur kurz angerissen.

[102] Vgl. Spengel in: Jacobs (2023), S. 419.

[103] Vgl. BFH-Urteil v. 25.5.2011 - I R 95/10, BStBl 2014 II S. 760, Rn. 16; so auch Weggenmann in: Wassermeyer et al. (2015), Rn. 6.52.

[104] Vgl. BFH-Beschluss v. 4.4.2007 - I R 110/05, BStBl 2007 II S. 521, III, Nr. 2, lit. b.

[105] Vgl. Weggenmann / Nehls in: Vogel und Lehner (2021), Art. 1 OECD-MA 2017, Rn. 68f.

[106] So auch Weggenmann in Wassermeyer et al. (2015), Rn. 6.45.

[107] Vgl. Spengel in: Jacobs (2023), S. 445.

[108] Vgl. BFH-Beschluss v. 4.4.2007 - I R 110/05, BStBl 2007 II S. 521, III, Nr. 2, lit. d.

[109] Vgl. dazu Wellmann in: Elschner et al. (2015), S. 93; so auch Weggenmann / Nehls in: Vogel und Lehner (2021), Art. 1 OECD-MA 2017, Rn. 68f.

[110] Vgl. BMF-Schreiben v. 26.9.2014 - IV B 5 - S 1300/09/10003, BStBl I S. 1258, Tz. 4.1.4. Ein anderes Argument in Bezug auf die ausgeschlossene Anrechnung ist die fehlende Periodenidentität.

[111] So auch Weggenmann in: Wassermeyer et al. (2015), Rn. 6.39.

2.2.3 Qualifikationskonflikte

Im vorliegenden Fall kommen bei der Abkommensanwendung subjektive Qualifikationskonflikte[112] zur Geltung. Bei der Anwendung des DBA sind jeweils unterschiedliche Artikel je nach Sichtweise des jeweiligen Vertragsstaates einschlägig. Im Folgenden ist dies kurz aus der steuerlichen Perspektive Bulgariens als der Sitzstaat der nach dem Kapitalgesellschaftskonzept behandelten Personengesellschaft aufzuzeigen. Wie bereits erläutert (s. Kap. 2.2.2.2.1), richtet sich im Thesaurierungfall die Besteuerung ausschließlich nach dem innerstaatlichen Steuerrecht Bulgariens. Im Falle einer Gewinnrepatriierung wird diese unter dem Dividendenartikel gemäß Art. 10 Abs. 3 DBA D/BG subsumiert.[113] Die Dividenden, die die bulgarische Personengesellschaft an dem deutschen Investor auszahlt, können im Quellenstaat (in Bulgarien) besteuert werden (Art. 10 Abs. 2 Satz 1 Halbsatz 1 DBA D/BG). Das Besteuerungsrecht von Bulgarien kann jedoch von dem DBA der Höhe nach eingeschränkt werden. Laut Art. 10 Abs. 2 Satz 1 Halbsatz 2 lit. b DBA D/BG beträgt der maximale anwendbare Steuersatz 15 Prozent. Als Folge des unterhalb dieses Höchststeuersatzes liegenden fünfprozentigen Kapitalertragsteuersatzes erfolgt keine Entlastung aus dem DBA. Die Dividendeneinkünfte werden nach dem EStG-BG besteuert.[114] Die Kapitalertragsteuer stellt eine endgültige Steuerbelastung dar, da die Steuer in Deutschland, wie bereits im Kapitel 2.2.3 erläutert, nicht anrechenbar ist. Die Gewinne der die Dividenden ausschüttenden bulgarischen Personengesellschaft unterliegen in Bulgarien weiterhin der Besteuerung (Art. 10 Abs. 2 Satz 3 DBA D/BG).

2.3 Gründung einer Kapitalgesellschaft

Die Gründung einer ausländischen Kapitalgesellschaft stellt grundsätzlich die bevorzugte Form des Auslandsengagements dar.[115] In der vorliegenden Konstellation (s. Abbildung 3) wird davon ausgegangen, dass der deutsche Investor[116] über eine deutsche Spitzeneinheit in der Form einer Personengesellschaft eine Kapitalgesellschaft in Bulgarien gründet, die aktive Einkünfte erzielt.

Die steuerrechtliche Behandlung der GmbH & Co KG in Deutschland wurde bereits in den vorangegangenen Ausführungen erläutert (s. Kap. 2.1.1.1). Nachfolgend werden lediglich die sich unmittelbar aus dieser Konstellation ergebenden Steuerwirkungen erörtert.

2.3.1 Nationales Recht

Im weiteren Verlauf erfolgt eine steuerrechtliche Würdigung des geschilderten Auslandssachverhalts unter Berücksichtigung der nationalen Regelungen der beiden beteiligten Staaten.

Deutschland

Grundlegend für die weitergehende steuerliche Analyse ist die Beantwortung der Frage nach der steuerlichen Qualifikation der bulgarischen Kapitalgesellschaft nach deutschem Steuerrecht. Wie in dem Fall einer bulgarischen Personengesellschaft (s. dazu Kap. 2.2.2.1), wird für diesen Zweck die Rechtshilfe der Finanzverwaltung herangezogen. Demnach werden die bulgarischen Kapitalgesellschaften dem Rechtstyp einer deutschen Kapitalgesellschaft gleichgestellt.[117] Im Gegensatz zum Fall der bulgarischen Personengesellschaft wird die bulgarische Kapitalgesellschaft auch in Deutschland als selbstständiger Rechtsträger auf Basis des Trennungsprinzips steuerlich behandelt. Somit ist im Folgenden eine klare Trennung der Vermögenssphären zwischen der Gesellschaft und ihren Gesellschaftern erforderlich.

Gesellschaftsebene

Infolgedessen bleibt die Leistungsfähigkeit der Anteilseigner bei der Besteuerung der Kapitalgesellschaft unberücksichtigt.[118] Damit diese Abschirmwirkung der Kapitalgesellschaft uneingeschränkt zum Tragen kommt, sollten grundsätzlich weitere nationale steuerliche Sondernormen beachtet werden. Es ist der Frage nachzugehen, ob die Tatbestandsmerkmale von §§ 39, 41, 42 AO sowie §§ 7–13 AStG erfüllt sind. Da im vorliegenden Sachverhalt die ausländische Kapitalgesellschaft annahmegemäß keine passiven Einkünfte erwirtschaftet, wird auf eine Prüfung der genannten Vorschriften verzichtet.

Folglich kommt keine Einschränkung zum Tragen und die bulgarische Kapitalgesellschaft wird aufgrund der Abschirmung ausschließlich in Bulgarien der bulgarischen Körperschaftsteuer unterworfen.[119] Im Falle einer Thesaurierung des Gewinns bleibt somit jegliche Besteuerung in Deutschland aus. Analog sind die Verluste der Kapitalgesellschaft ausschließlich in Bulgarien steuerlich zu berücksichtigen.[120] Für die steuerrechtliche Behandlung der Kapitalgesellschaft in Bulgarien wird auf die Ausführungen zur Besteuerung juristischer Personen im Kapitel 2.1.1.2 Bezug genommen.

Gesellschafterebene

Im Rahmen der betrachteten Form des Auslandsengagements setzt die deutsche Steuerhoheit erst im Rahmen einer Gewinnrepatriierung aus der bulgarischen Kapitalgesellschaft ein.[121] Diese erhöht das Welteinkommen des unbeschränkt steuerpflichtigen deutschen Investors (§ 15 Abs. 1

[112] Vgl. Spengel in: Jacobs (2023), S. 444.
[113] Vgl. BMF-Schreiben v. 26.9.2014 - IV B 5 - S 1300/09/10003, BStBl I S. 125, Tz. 4.1.4; vgl. auch Spengel in: Jacobs (2023), S. 419.
[114] Vgl. Bulgarische Nationale Steuerverwaltung (2021b), S. 3.
[115] Vgl. Spengel in: Jacobs (2023), S. 332.
[116] Zur Vereinfachung wird im Rahmen der Gewinnausschüttung von einem deutschen Investor ausgegangen. Rechtliche Aspekte der Gewinnausschüttung werden nicht berücksichtigt.
[117] Siehe dazu die sog. Betriebsstätten-Verwaltungsgrundsätze: BMF-Schreiben v. 24.12.1999, IV B 4 – S 1300 – 111/99, BStBl 1999 I S 1076, Tabelle 2, Bulgarien, Nr. 3–5.
[118] Vgl. Bundesverfassungsgerichts-Beschluss v. 29.3.2017 - 2 BvL 6/11, lexetius.com/2017, 1008, Rn. 42.
[119] Vgl. Spengel in: Jacobs (2023), S. 354.
[120] Vgl. Spengel in: Jacobs (2023), S. 376.
[121] Vgl. Hey in Tipke und Lang (2024), S. 971, Rn. 13.135.

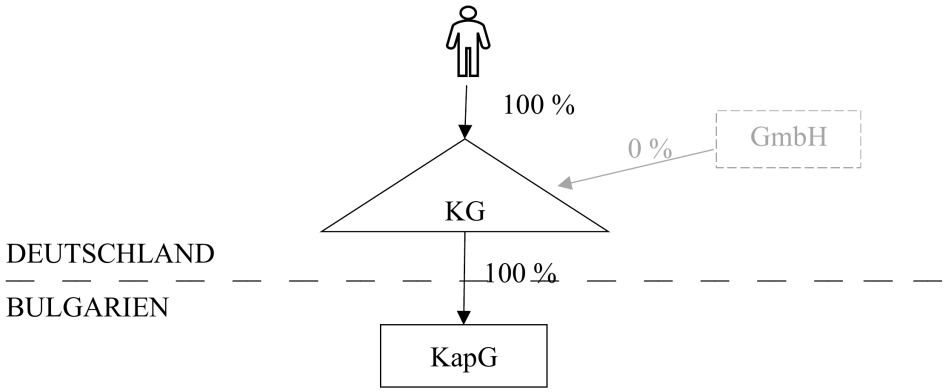


Abbildung 3: Gründung einer Kapitalgesellschaft in Bulgarien (Schematische Darstellung)

S. 1 Nr. 2 EStG i. V. m. § 1 Abs. 1 Satz 1 EStG). Die Gewinnausschüttung der bulgarischen Kapitalgesellschaft fällt grundsätzlich unter den Einkünften aus Kapitalvermögen gemäß § 20 Abs. 1 Nr. 1 S. 1 EStG. Nach § 20 Abs 8 Satz 1 EStG erfolgt jedoch eine Umqualifizierung in Einkünfte aus Gewerbebetrieb.[122] Daher ist der 25-prozentige Tarif für Einkünfte aus Kapitaleinkünfte nicht anzuwenden (§ 32d Abs. 1 Satz 1 EStG). Da der deutsche Investor der Gesellschafter einer Personengesellschaft ist, findet das Teileinkünfteverfahren (im folgenden Verlauf TEV) auf die ausländische Dividende Anwendung.[123] Demnach werden lediglich 60 Prozent der Dividende der Besteuerung unterworfen (§ 3 Nr. 40 Satz 1 lit. d Satz 1 EStG i. V. m. § 3 Nr. 40 Satz 2 EStG). Dabei ist die ausländische einbehaltene Kapitalertragsteuer nicht abzugsfähig (§ 12 Nr. 3 EStG). Es wird somit die Bruttodividende herangezogen. Korrespondierend können die damit wirtschaftlich zusammenhängenden Betriebsausgaben nur zu 60 Prozent steuerlich berücksichtigt werden (§ 3c Abs. 2 Satz 1 EStG).

Des Weiteren ist der Gewerbeertrag um die Gewinne aus Beteiligungen an ausländischen Kapitalgesellschaften zu kürzen, sofern eine Beteiligung in Höhe von 15 Prozent oder mehr besteht und die Anteile dem Betriebsvermögen zuzurechnen sind (§ 9 Nr. 7 GewStG). Folglich unterliegen die im vorliegenden Fall erzielten Dividenden nicht der Gewerbesteuer.

Die laufende Besteuerung der inländischen Einkünfte der deutschen GmbH & Co KG bzw. des deutschen Investors in Deutschland vollzieht sich nach den im Kapitel 2.1.1.1 dargestellten allgemeinen Grundsätzen zu der Besteuerung von Personengesellschaften.

Bulgarien

Aufgrund des geltenden Trennungsprinzips sind die Ebenen der Gesellschaft und der Gesellschafter voneinander separat zu würdigen.

Gesellschaftsebene

Die bulgarische Kapitalgesellschaft mit Sitz und Geschäftsleitung in Bulgarien stellt gemäß Art. 3 Abs. 1 Nr. 1 KStG-BG eine inländische (bulgarische) juristische Person dar. Daher unterliegt sie der unbeschränkten Steuerpflicht und besteuert somit ihre weltweiten Einkünfte in Bulgarien (Art. 3 Abs. 2 KStG-BG). Die Besteuerung der Kapitalgesellschaft im Gewinn- und Verlustfall erfolgt nach den Bestimmungen des KStG-BG und entspricht damit der zuvor erläuterten Vorgehensweise bei bulgarischen Personengesellschaften bzw. bulgarischen Betriebsstätten einer ausländischen Personengesellschaft (s. Kap. 2.1.1.2).

Gesellschafterebene

Im Falle von Gewinnausschüttungen der bulgarischen Kapitalgesellschaft an die deutsche Personengesellschaft[124] wird in Bulgarien eine beschränkte Steuerpflicht der deutschen Personengesellschaft begründet (Art. 2 Abs. 1 Nr. 2 KStG-BG). Eine Auflistung der Einkünfte von ausländischen Personen, die als Einkünfte aus bulgarischer Quelle gelten, erfolgt in Art. 12 KStG-BG. Darunter fallen unter anderem Dividenden aus Beteiligungen an inländischen juristischen Personen, im vorliegenden Fall an der bulgarischen Kapitalgesellschaft (Art. 12 Abs. 4 KStG-BG). Nach Art. 194 Abs. 1 Nr. 1 i. V. m. Art. 200 Abs. 1 KStG-BG unterliegt die Dividende einer Besteuerung in Höhe von fünf Prozent.

Das Gesetz sieht Ausnahmen von der Besteuerung vor, unter anderem im Falle ausländischer in einem EU-Mitgliedstaat ansässiger juristischer Personen (Art. 194 Abs. 3 Nr. 3 KStG-BG). Dafür hat die ausländische Gesellschaft eine von der deutschen Steuerverwaltung ausgestellte Bescheinigung vorzulegen, die belegt, dass sie in Deutschland steuerlich ansässig ist. Andernfalls besteht die Möglichkeit, dass die Befreiung seitens der bulgarischen nationalen

[122] So auch Weiss (2021), S. 100.
[123] Vgl. Spengel in: Jacobs (2023), S. 361.
[124] Die deutsche Personengesellschaft wird aus bulgarischer Sicht steuerlich als ausländische juristische Person qualifiziert, vgl. dazu Bulgarische Nationale Steuerverwaltung (2021a), S. 2.

Steuerbehörde nicht anerkannt wird.[125] Da die Ausstellung einer solchen Ansässigkeitsbescheinigung für die deutsche Personengesellschaft von der deutschen Finanzverwaltung ausgeschlossen ist,[126] wird die fünfprozentige Quellensteuer bei der Kapitalgesellschaft einbehalten (Art. 194 Abs. 2 KStG-BG).

2.3.2 Abkommensrecht

Der Anwendungsbereich des DBA D/BG ist im vorliegenden Fall eröffnet. Für eine detaillierte Prüfung wird auf Kapitel 2.1.2 verwiesen. Da die bulgarische Gesellschaft eine Dividende an den in Deutschland ansässigen Investor ausschüttet, liegt ein grenzüberschreitender Sachverhalt vor, sodass Art. 10 Abs. 1 DBA D/BG anwendbar ist.

In Bezug auf die Dividenden der im Quellenstaat (Bulgarien) ansässigen Kapitalgesellschaft wird Deutschland ein Besteuerungsrecht zugewiesen (Art. 10 Abs. 1 DBA D/BG). Die Dividende wird im Rahmen der unbeschränkten Steuerpflicht des deutschen Investors unter Anwendung des TEV besteuert (s. detailliert zur steuerlichen Behandlung in Deutschland Kap. 2.3.1.1.2). Deutschland ist in diesem Zusammenhang verpflichtet, die entstehende Doppelbesteuerung durch Anrechnung der in Bulgarien auf die Dividende erhobenen Quellensteuer zu vermeiden (Art. 22 Abs. 1 lit. b sublit. aa DBA D/BG). Die Anrechnung erfolgt nach dem DBA D/BG ausdrücklich nach Maßgabe der einschlägigen Normen des deutschen Steuerrechts über die Anrechnung ausländischer Steuern (Art. 22 Abs. 1 lit. b sublit. aa DBA D/BG).[127] Folglich ist § 34c Abs. 1 Sätze 2–5, Abs. 2 EStG anzuwenden (§ 34c Abs. 6 Satz 2 Halbsatz 1 EStG)[128] und die in Bulgarien einbehaltene Kapitalertragsteuer ist auf die deutsche Einkommensteuer anzurechnen. Im Rahmen der Anrechnung ist einer Reihe von Besonderheiten Rechnung zu tragen. Die bulgarische Steuer ist maximal in Höhe des niedrigeren Betrags aus der tatsächlich gezahlten ausländischen Steuer[129] und dem ermittelten Anrechnungshöchstbetrag[130] (§ 34c Abs. 1 Satz 2 EStG) anzurechnen. Somit ist die Anrechnung auf die durchschnittliche tarifliche inländische Einkommensteuer, berechnet als tarifliche Einkommensteuer geteilt durch das zu versteuernde Einkommen, und bezogen auf die ausländischen Einkünfte beschränkt. Durch die Abstellung auf das zu versteuernde Einkommen werden persönliche Abzüge berücksichtigt.[131] Bei der Berechnung dieses Höchstbetrags ist zu beachten, dass die ausländischen Einkünfte nach den Vorschriften des deutschen Einkommensteuerrechts zu ermitteln sind (R 34c Abs. 3 S. 3 EStR).[132] In der Höchstbetragsberechnung ist nur der steuerpflichtige Teil der nach dem TEV besteuerten Dividende einzubeziehen (R 34c Abs. 3 Satz 5 EStR). Der den Anrechnungshöchstbetrag übersteigende Teil (Anrechnungsüberhang) ist auf die deutsche Steuer nicht anrechenbar und geht steuerlich verloren.[133] Nach Auffassung der Finanzverwaltung ist die ausländische Steuer trotz der als Folge des TEV zu 40 Prozent steuerbefreiten Einkünfte nicht zu kürzen (R 34c Abs. 2 Satz 3 EStR). Auf Antrag des Steuerpflichtigen kann die Quellensteuer alternativ im Rahmen der Ermittlung der Einkünfte in dem Umfang abgezogen werden, in dem sie nicht steuerlich befreit ist – im vorliegenden Fall also zu 60 Prozent (§ 34c Abs. 2 EStG, R 34c Abs. 5 i. V. m. Abs. 4. Satz 8 EStR).

Das DBA mit dem Ansässigkeitsstaat der Kapitalgesellschaft führt zu keiner Änderung bei der Besteuerung bzw. bei der unbeschränkten Steuerpflicht der Gesellschaft im Ausland bzw. in Bulgarien.[134] Das Besteuerungsrecht Bulgariens auf den Gewinn der Kapitalgesellschaft, auf dessen Basis die Gewinnausschüttung erfolgt, bleibt unberührt (Art. 10 Abs. 2 DBA D/BG). Die bulgarische Körperschaftsteuer auf den Gewinn der Kapitalgesellschaft kann auf Ebene des deutschen Investors nicht angerechnet werden, da es an einer rechtlichen und wirtschaftlichen Zugehörigkeit zum Einkommen des Gesellschafters fehlt.[135]

2.3.3 Qualifikationskonflikte

Im Fall der bulgarischen Kapitalgesellschaft entstehen keine Qualifikationskonflikte. Aus der Sicht Bulgariens wird bei der Abkommensanwendung analog zu der deutschen Würdigung der Dividendenartikel herangezogen. Bulgarien als der Sitzstaat der Kapitalgesellschaft hat das Recht, die Dividenden ebenfalls einer Besteuerung zu unterwerfen (Art. 10. Abs. 2 Halbsatz 1 DBA D/BG). Die Höhe der Quellenbesteuerung auf die Dividende ist jedoch durch das Abkommen auf 15 Prozent eingeschränkt (Art. 10 Abs. 2 lit. b DBA D/BG). Da das Abkommensrecht lediglich eine Schrankenwirkung auf die Besteuerungshöhe hat und die Besteuerungsrechte der Vertragsstaaten nicht erweitert, ist der fünfprozentige Quellensteuersatz nach bulgarischem Recht anzuwenden.

2.4 Zwischenfazit

Für den steuerrechtlichen Rahmen der Besteuerung von Direktinvestitionen am Standort Bulgarien lässt das vorstehend Erörterte die nachfolgend dargelegten Schlussfolgerungen zu.

Die Analyse hat gezeigt, dass sich unterschiedliche Steuerwirkungen ergeben, je nachdem, welche der drei Alternativen der Direktinvestition ausgewählt wird. Der Fall der Betriebsstätte scheint aufgrund der anwendbaren Freistellungsmethode (unter PVB) zu einer geringeren Besteuerung zu

[125] Vgl. Petkov (2015), Zeile 13ff.
[126] Vgl. Oberfinanzdirektion (OFD) Nordrhein-Westfalen (2020), Verfügung v. 9.4.2020, S 1301 - 1989/5000 - St 125, IStR 2020, S. 435, Nr. 11, Rn. 6.
[127] Vgl. Becker in: Wassermeyer (2025), Art. 22 DBA D/BG 2010, Rn. 38.
[128] Vgl. Wagner in: Brandis und Heuermann (2024), § 34c EStG, Rn. 133.
[129] Vgl. BFH-Urteil v. 15.3.2023, I R 8/20, BStBl 2024 II S. 264, Rn. 16.
[130] Vgl. Kraft in: Kanzler et al. (2025), § 34c EStG, Rn. 30. Vgl. auch H 34c Abs. 3 EStH.
[131] Vgl. Wagner in: Brandis und Heuermann (2024), § 34c EStG, Rn. 52. Siehe auch Kraft in: Kanzler et al. (2025) § 34c EStG, Rn. 30; H 34c Abs. 3 EStH, „Ermittlung des Höchstbetrags für die Steueranrechnung".
[132] So auch Weiss (2021), S. 170.
[133] Vgl. Wagner in: Brandis und Heuermann (2024), § 34c EStG, Rn. 17.
[134] Vgl. Spengel in: Jacobs (2023), S. 355.
[135] Vgl. Wagner in: Brandis und Heuermann (2024), § 34c EStG, Rn. 33.

führen als in den beiden anderen Fällen. Bei der Gründung einer bulgarischen Personengesellschaft entsteht aufgrund der entstehenden Qualifikationskonflikte durch das Zusammentreffen der beiden unterschiedlich ausgestalteten Steuersysteme von Bulgarien und Deutschland im Rahmen des internationalen Steuerrechts eine zusätzliche nicht anrechenbare steuerliche Belastung in Höhe von fünf Prozent. Dem Fall ist eine große Bedeutung beizumessen, weil im dessen Rahmen die Komplexität des internationalen Steuerrechts zumindest exemplarisch zum Vorschein kommt.

Im Rahmen der Kapitalgesellschaft vollzieht sich die Besteuerung auch nach deutscher Wertung auf zwei getrennten Ebenen: Einerseits unterliegt die Kapitalgesellschaft der Besteuerung in Bulgarien. Andererseits wird in Bezug auf die Dividende den beiden Steuerhoheiten das Recht gewährt, das Substrat zu besteuern, wobei die bulgarische Steuerlast unter der Anrechnungsmethode in Deutschland berücksichtigt wird.

In einem weiteren Gedankengang liegt selbst bei der Gegenüberstellung des flachen zehnprozentigen Steuertarifs Bulgariens im Vergleich zum progressiven Steuertarif in Deutschland (bis zu 45 Prozent) nahe, dass eine Direktinvestition in Bulgarien voraussichtlich steuerlich vorteilhafter sein könnte als eine rein inländische Geschäftsausweitung. In einer realen Konstellation spielen jedoch weitere komplexe Faktoren eine Rolle, wie zum Beispiel die Wirkung des Progressionsvorbehalts, die Progressivität des deutschen Tarifs, der inländische und ausländische Umgang mit Verlusten, insbesondere mit Verlustvorträgen und -rückträgen, sowie die steuerliche Zulässigkeit der Rückstellungen.

In diesem Kontext wird es deutlich, dass die durchgeführte rein qualitative steuerliche Analyse der Komplexität des Sachverhalts nicht gerecht wird. Sie erlaubt in diesem Fall nur pauschalisierte Aussagen, jedoch keine konkrete Quantifizierung der Steuerlast bzw. des steuerlichen Vorteils.

Folglich lässt sich die Frage nach der steuerlichen Vorteilhaftigkeit einer Direktinvestition in Bulgarien an diesem Punkt der Analyse unmöglich abschließend beantworten und bedarf einer weitergehenden quantitativen Würdigung.

3 Untersuchung der Investitionsalternativen auf steuerliche Vorteilhaftigkeit

Ausgehend von dem identifizierten Untersuchungsbedarf ist im Folgenden eine quantitative Analyse zur Beurteilung der steuerlichen Vorteilhaftigkeit des Auslandsengagements durchzuführen.

Diese hat zum primären Ziel, die Steuerbelastung der drei Direktinvestitionsalternativen in Bulgarien unter einer Vielzahl von möglichen Gewinn- und Verlustkonstellationen im Vergleich zu einer Geschäftsausweitung in Deutschland eindeutig zu quantifizieren. Hierzu soll ein möglichst realitätsnahes und branchenunabhängiges Modell aufgebaut werden. Angesichts der grundsätzlich breiten Definition von mittelständischen Unternehmen[136] und der damit verbundenen Schwierigkeit, auf eindeutige empirische Daten zuzugreifen, sollte in der Simulation auf die Generierung von Zufallszahlen zurückgegriffen werden, um die für die Steuerveranlagung relevanten Parameter zu erzeugen.

Im Anbetracht der oben genannten Anforderungen an die einzusetzende steuerliche Analyse erscheint es naheliegend, eine für die Fragestellung der Arbeit zugeschnittene Monte-Carlo-Simulation über fünf Veranlagungszeiträume zu konstruieren und durchzuführen.[137]

3.1 Darstellung der Methodik

Die Monte-Carlo-Simulation stellt eine computerbasierte Methode dar. Diese ermöglicht die Analyse von komplexen Sachverhalten basierend auf der Erzeugung von Zufallszahlen unter vordefinierten Verteilungen.[138] Ausgehend von diesen Zufallsvariablen[139] werden über viele Iterationen hinweg auf Basis eines vorgegebenen deterministischen Modells viele zufällige Zahlkombinationen generiert und daraus relevante Statistiken gesammelt.[140]

Für die Zwecke der vorliegenden Untersuchung ist die Monte-Carlo-Simulation in zwei Hauptszenarien – Gewinn- und Verlustszenario – unterteilt. Die Grundstruktur der zugrunde liegenden Hauptszenarien und Fälle der Simulation ist in der nachfolgenden Tabelle 1 veranschaulicht.

Beide Szenarien werden getrennt voneinander konstruiert und simuliert und beinhalten jeweils drei Fälle, die den drei im zweiten Kapitel abgebildeten Direktinvestitionsalternativen entsprechen. Zur Beurteilung der steuerlichen Vorteilhaftigkeit des Auslandsvorhabens im Vergleich zu einer rein inländischen Geschäftsausweitung werden das Verlust- und Gewinnszenario jeweils um einen Basisfall erweitert, der als Vergleichsbasis dient.

Da in der vorliegenden Masterarbeit der steuerliche Aspekt grundlegend ist, basiert die Simulation in den einzelnen Szenarien auf fünfperiodigen kasuistischen Veranlagungssimulationen. Dabei wurde die Veranlagungsmethode der Finanzbehörde[141] für die in Tabelle 1 dargelegten Szenarien und zugehörigen Fällen vorerst in einem rechnerischen Modell unter Einsatz mathematischer Funktionsformeln in Microsoft Excel nachgebildet. Auf Basis dieses erstmal deterministisch gehaltenen Modells wurden die für die Steuerlastberechnung relevanten Variablen identifiziert und als Zufallsvariablen definiert. Die in der stochastischen Simulation als Zufallsvariablen definierten Parameter und die

[136] Für den Begriff existiert keine allgemeingültige Definition. Gängige Definitionen bauen auf Kriterien der Jahresumsatzhöhe und der Anzahl der Angestellte auf, vgl. bspw. Institut für Marktforschung Bonn (2016), KMU-Definition.

[137] Die Monte-Carlo-Simulation wurde in Microsoft Excel unter Anwendung des Add-ins Risk Kit durchgeführt.

[138] Vgl. Gleißner (2020), Monte-Carlo-Simulation, HI1070758.

[139] In der Fachliteratur wird bei computergestützten Zufallszahlen der Begriff der *Pseudozufallszahlen* benutzt, vgl. Büchter und Henn (2007), S. 378. Im Folgenden wird bei der Verwendung des Begriffs der Zufallszahl bzw. Zufallsvariable von einer computererzeugten Pseudozufallszahl ausgegangen.

[140] Vgl. Brandimarte (2014), S. 3.

[141] Vgl. Haase et al. (2025), Besteuerung von Unternehmen III, Rn. 12.

Tabelle 1: Übersicht der Hauptszenarien und der zugehörigen zentralen Modellfälle

Szenario	Fälle	Definition
Gewinnszenario	Gewinnfall 1 BS	GmbH & Co KG mit Betriebsstätte in Bulgarien (s. Abbildung 1)
	Gewinnfall 2 PersG	Deutscher Investor gründet eine Personengesellschaft in Bulgarien (s. Abbildung 2)
	Gewinnfall 3 KapG	GmbH & Co KG gründet eine Kapitalgesellschaft in Bulgarien (s. Abbildung 3)
	Basisgewinnfall	GmbH & Co KG mit inländischer Betriebsstätte (s. Abbildung 10, S. XI)
Verlustszenario	Verlustfall 1 BS	GmbH & Co KG mit Betriebsstätte in Bulgarien (s. Abbildung 1)
	Verlustfall 2 PersG	Deutscher Investor gründet eine Personengesellschaft in Bulgarien (s. Abbildung 2)
	Verlustfall 3 KapG	GmbH & Co KG gründet eine Kapitalgesellschaft in Bulgarien (s. Abbildung 3)
	Basisverlustfall	GmbH & Co KG mit inländischer Betriebsstätte (s. Abbildung 10, S. XI)

diesbezüglichen Annahmen sind nachfolgend im Kapitel 3.2 näher erläutert.

Eine zentrale Zielgröße vom Interesse in Bezug auf die Steuerlastberechnung ist der Barwert der Gesamtsteuerbelastung. Dieser wird in allen Szenaren und Fällen im Rahmen einer einzelnen Iteration wie folgt berechnet:

Zum Zwecke einer ganzheitlichen Betrachtung und Vergleichbarkeit zwischen den Fällen wird auch im Rahmen der Konstellationen, in denen das Trennungsprinzip bei der Besteuerung greift, die Steuerbelastung auf Ebene des Anteilseigners und der Körperschaft zu einer Gesamtsteuerbelastung zusammengerechnet. Zudem wird in Gewinnjahren im Falle eines Auslandsengagements in Form einer bulgarischen Kapitalgesellschaft der zeitliche Aspekt der Gewinnausschüttung nicht berücksichtigt.[142] Aus Vereinfachungsgründen wird unterstellt, dass die Dividende jeweils zum Periodenende ausgeschüttet wird. Es wird stets davon ausgegangen, dass der volle Gewinn aus der Personengesellschaft entnommen bzw. im Rahmen der bulgarischen Körperschaften ausgeschüttet wird. Thesaurierungsbegünstigungen werden daher außer Acht gelassen.

Aus Vereinfachungsgründen zieht die Simulation als Diskontierungszinssatz den risikolosen Zinssatz in Höhe von 2,40 Prozent auf Basis der siebenjährigen deutschen Bundesanleihen[143] heran. Dadurch wird der Zeitwert des Geldes im Modell gewürdigt. Zusätzlich untersuchen die einzelnen Szenarien bei Bedarf weitere relevanten Zielgrößen aus der Simulation. Im Gewinnszenario ist bspw. die relative effektive Steuerbelastung aus dem Verhältnis der Gesamtsteuerbelastung zum Gesamtjahresüberschuss vor Steuern vom Interesse.

Bei jedem Simulationsdurchlauf werden neue Werte der Zufallszahlen auf Basis der gewählten Wahrscheinlichkeitsverteilungen in der zugrunde liegenden Veranlagungssimulation generiert und die Steuerbelastung bzw. die vorgehend genannten steuerlich relevanten Zielgrößen ausgerechnet.

3.2 Modellannahmen

In Bezug auf die simulierten zugrunde liegenden Konstellationen innerhalb der Szenarien werden die bis jetzt getroffenen Annahmen in der Masterarbeit fortgeführt. Der Steuerpflichtige plant daher eine aktive gewerbliche Geschäftstätigkeit und erzielt sowohl Einkünfte aus dem Inland als auch aus dem Ausland. Hintergrund dafür ist, die Güte des Modells zu erhöhen, indem die plausible Abbildung der Realität gewährleistet wird. Zudem ermöglicht die Annahme die Anwendung des PVB in den relevanten Fällen.

Die der Monte-Carlo-Simulation zugrunde liegenden Veranlagungssimulationen werden vereinfacht dargestellt und umfassen lediglich die für die Sachverhalte relevanten Aspekte. Es wird von einer Einzelveranlagung ausgegangen. Annahmegemäß macht der Steuerpflichtige keine Sonderausgaben, außergewöhnliche Belastungen oder Freibeträge für Kinder geltend.

Der Zeitraum der Simulation erstreckt sich auf fünf Veranlagungszeiträume, von 2025 bis 2029. Der ausgewählte Analysezeitraum orientiert sich am strategischen Planungshorizont von in der Regel drei bis sechs Jahren[144] und ermöglicht eine fundierte mehrperiodige Beurteilung der Steuerlast. Das handelsrechtliche Jahresergebnis wird durchgängig mit dem steuerlichen Ergebnis gleichgesetzt. Eine Ausnahme davon bildet nur die Berücksichtigung der Rückstellungshöhe in Bulgarien im Rahmen des Gewinnfalls, auf die in diesem Kapitel kurz eingegangen wird. Das bulgarische Ergebnis wird angesichts der voraussichtlich baldigen Einführung des Euros in Bulgarien[145] sowie des konstanten Wechselkurses der bulgarischen Währung zum Euro direkt in Euro ausgewiesen. Die Annahme lautet, dass der deutsche Investor nur Einkünfte aus Gewerbebetrieb erzielt und keine weiteren Einkünfte aus anderen Einkunftsarten im Sinne des § 2 Abs. 1

[142] Die Gewinnausschüttung muss im Falle einer bulgarischen Kapitalgesellschaft zunächst von der Hauptversammlung beschlossen werden (Art. 221 Nr. 7 HGB-BG).

[143] Vgl. Deutsche Finanzagentur (2025), Tabelle „Konventionelle Wertpapiere".

[144] Vgl. Sasse und Mink (2020), HI8669381.

[145] Mit dem Euro-Einführungstermin in Bulgarien ist das handelsrechtliche Ergebnis in Euro zu ermitteln, § 5 Abs. 1 i. V. m. § 60 Gesetz über die Einführung des Euro in der Republik Bulgarien.

$$\text{Barwert der Gesamtsteuerbelastung} = \sum_{t}^{T=5} \frac{\text{GStZ in } t}{(1+r)^t}$$

GStZ in t = Gesamtsteuerzahlung (in Deutschland und in Bulgarien) in Periode t
r = Diskontierungszinssatz
t = Periode (1 bis 5)

Abbildung 4: Formel für die Berechnung des Barwerts der Gesamtsteuerbelastung in der Simulation

EStG bezieht. Für alle Kalenderjahre wird der im Jahr 2025 geltende Einkommensteuertarif (s. Tabelle 11, S. XII) zugrunde gelegt. Die Freigrenzen des Solidaritätszuschlags werden in der Veranlagung nicht berücksichtigt.

Aus Vereinfachungsgründen werden bei der gewerbesteuerlichen Veranlagung in allen Fällen die Hinzurechnungen (§ 8 Nr. 1 GewStG) und Kürzungen (§ 9 Nr. 1 GewStG) außer Acht gelassen. Es wird vom Gewerbeertrag vor Verlustabzug, Abrundung und Freibetrag ausgegangen. Für die Gewerbesteuer wird der im Jahr 2023 in Deutschland durchschnittliche Gewerbesteuerhebesatz von 407 Prozent[146] herangezogen. Es wird unterstellt, dass keine Verlustvorträge aus vorangegangenen Zeiträumen bestehen und dass die Verlustvorträge in Bulgarien vom Steuerpflichtigen geltend gemacht wurden.

Sowohl im Gewinn- als auch im Verlustszenario wird davon ausgegangen, dass in Deutschland ausschließlich Gewinne erwirtschaftet werden. Diese Annahme trägt dem Umstand Rechnung, dass eine Ausweitung der Geschäftstätigkeit, sei es im In- oder Ausland, im Regelfall für bereits gewinnbringende Tätigkeiten in Deutschland in Frage kommt. Im Rahmen des Gewinnszenarios werden annahmegemäß von der jeweiligen Tätigkeit in Bulgarien ebenfalls Gewinne erzielt. Dabei wird die Summe der handelsrechtlichen Jahresüberschüsse aus Bulgarien und Deutschland im Anfangsjahr 2025 als Zufallsvariable $Gesamtjahresüberschuss_{2025}$ in der Monte-Carlo-Simulation modelliert. Diese stellt als Eingabeparameter der Ausgangspunkt für die Simulation dar. Es wird unterstellt, dass die Zufallsvariable zwischen einem Minimumwert von 200.000 und einem Maximumwert von 1.000.000 stetig gleichverteilt ist. Statistisch wird diese Zufallsvariable so definiert:

In jeder einzelnen Simulationsiteration wird eine Zufallszahl aus dem festgelegten Wertebereich generiert. Dabei besitzt jeder Wert die gleiche Wahrscheinlichkeit. Die Summe der Jahresüberschüsse aus Bulgarien und Deutschland in den vier Folgeperioden (2026–2029) basiert auf dem für das Jahr 2025 generierten Zufallswert unter Berücksichtigung einer aus Vereinfachungsgründen über den Perioden konstant gehaltenen Inflationsrate in Höhe von zwei Prozent pro Jahr. Dadurch werden unrealistische Schwankungen zwischen den Jahren ausgeschlossen und ein kontinuierliches, inflationsbedingtes Wachstum unterstellt. Der zugrunde gelegte Prozentsatz trägt dem Ziel der Europäischen Zentralbank Rechnung, eine Inflation in der Europäischen Union in Höhe von zwei Prozent mittelfristig zu gewährleisten.[147]

Die prozentuelle Verteilung der Zufallsvariable *Gesamtjahresüberschuss* zwischen Deutschland und Bulgarien wird in jeder Iteration variiert, jedoch über die vier Fälle und die fünf Jahre hinweg konstant gehalten. Zu diesem Zweck wird der Anteil des Inlandsjahresüberschusses am Gesamtjahresüberschuss als eine stetig gleichverteilte Zufallsvariable $Inlandsanteil_{2025\text{-}2029}$ mit einem Mindestwert von 40 Prozent und einem Höchstwert von 90 Prozent definiert:

Der Auslandsanteil bzw. der Anteil für Bulgarien ergibt sich entsprechend als komplementärer Wert zu 100 Prozent. Aus modelltechnischen Gründen wird im Basisfall der Inlandsjahresüberschuss betragsmäßig der inländischen Personengesellschaft und der Auslandsjahresüberschuss der inländischen Betriebsstätte zugewiesen.

Ferner ist mit Rückgriff auf die Ausführungen im zweiten Kapitel noch eine steuerliche Besonderheit Bulgariens im Vergleich zu Deutschland für die steuerliche Analyse beachtenswert, nämlich der Umstand, dass die Bildung von Rückstellungen nach bulgarischem Recht steuerlich unzulässig ist. Da es sich in Bezug auf die Rückstellungen um temporäre Differenzen handelt, wird das Szenario nicht separat ausgewiesen. Für die Zwecke der Untersuchung berücksichtigt das Modell nicht den Zeitpunkt des Differenzausgleichs. Untersucht wird der periodenspezifische Effekt aus der erhöhten bulgarischen Bemessungsgrundlage. Zur Isolierung des Effektes dient die Erweiterung und doppelte Replikation des Gewinnszenarios unter sonst gleichbleibenden Bedingungen und Zufallszahlen, mit dem einzigen Unterschied, dass in einem der beiden Modelle die Rückstellungshöhe in Bulgarien mitberücksichtigt wird. Somit sind die Rückstellungen in diesem Modell zum Jahresüberschuss in Bulgarien hinzuzurechnen. Es wird zudem davon ausgegangen, dass es sich um solche Rückstellungen handelt, die in Deutschland steuerlich anerkannt werden. Im inländischen Basisgewinnfall erfolgt daher keine Anpassung, da die Rückstellungen bereits gewinnmindernd in dem Jahresüberschuss berücksichtigt wurden. Die Rückstellungshöhe wird als Anteil am bulgarischen Jahresüberschuss definiert. Für die Zufallsvariable *Rückstellungshöhe* wird eine Gleichverteilung im Intervall zwischen 15 Prozent und 50 Prozent des jeweiligen bulgarischen Jahresüberschusses angenommen. Von einem stabi-

[146] Vgl. Statistisches Bundesamt (2025), Durchschnittlicher Gewerbesteuerhebesatz 2023.

[147] Vgl. Europäische Zentralbank (2021), Pressemitteilung vom 8.7.2021.

$$\textit{Gesamtjahresüberschuss}_{2025} \sim U\,[200.000;\ 1.000.000]$$

Abbildung 5: Wahrscheinlichkeitsverteilung der Zufallsvariable „Gesamtjahresüberschuss" im Jahr 2025

$$\textit{Inlandsanteil}_{2025–2029} \sim U\,[40\ \%;\ 90\ \%]$$

Abbildung 6: Wahrscheinlichkeitsverteilung der Zufallsvariable „Inlandsanteil" in den Jahren 2025–2029

len Geschäftsmodell ausgehend gilt es, die simulierte Höhe des Anteils über die Perioden (und Fälle) hinweg konstant zu halten. Zwecks Vergleichbarkeit wird in allen Iterationen und Fällen ein identischer Anteil zugrunde gelegt.

Das Verlustszenario trägt dem Umstand Rechnung, dass eine Direktinvestition potenzielle Anlaufverluste nach sich bringen kann. Darum werden im Verlustszenario in den ersten drei Perioden bzw. Jahren Verluste im Ausland (und in der inländischen Betriebsstätte im Basisfall) unterstellt. Im Basisfall wird analog zu dem Gewinnszenario das Inlandsjahresergebnis betragsmäßig der inländischen Personengesellschaft und das Auslandsjahresergebnis betragsmäßig der inländischen Betriebsstätte zugeordnet. Im Verlustszenario erfolgt die Modellierung des Jahresüberschusses in Deutschland im Jahr 2025 als Zufallsvariable $\textit{Jahresüberschuss}_{2025}$. Diese ist stetig gleichverteilt im Intervall zwischen 200.000 und 1.000.000 Euro und wird in den Folgeperioden (2026–2029) analog zum Gewinnszenario inflationsbedingt konstant mit zwei Prozent erhöht.

Für das Verlustszenario wurde aus dem deterministischen Modell der Veranlagungssimulation die Notwendigkeit ersichtlich, im Verlustszenario zwei separate Verlustunterfälle in der Monte-Carlo-Simulation zu konstruieren: einen negativen (N) und einen positiven (P). Im Verlustunterfall N wird die Gesamtsumme der Jahresergebnisse aus Bulgarien und aus Deutschland negativ, d. h., der bulgarische Verlust ist absolut höher als der in Deutschland erwirtschaftete Jahresüberschuss. Im Verlustunterfall P ist es umgekehrt: Es kommt zu einer positiven Gesamtsumme der Jahresergebnisse aus der ausländischen Direktinvestition und aus Deutschland. Das Verlustszenario wird gemäß der Übersicht in der Tabelle 2 um die beiden Verlustunterfälle ergänzt.

Tabelle 2: Erweiterte Übersicht des Verlustszenarios

Szenario	Verlustunterfall	Fälle
Verlustszenario	Verlustunterfall N: negatives Gesamtergebnis	Verlustfall 1 N: BS Verlustfall 2 N: PersG Verlustfall 3 N: KapG Basisverlustfall N
	Verlustunterfall P: positives Gesamtergebnis	Verlustfall 1 P: BS Verlustfall 2 P: PersG Verlustfall 3 P: KapG Basisverlustfall P

Die Verluste der drei Verlustperioden in beiden Unterfällen N und P werden als gleichverteilt in einem Intervall zwischen a und b simuliert, wobei $a < b$ gilt. Im Verlustunterfall N wird der untere Rand des Intervalls im Jahr 2025 durch den kleineren Wert aus zwei Größen bestimmt, nämlich –1.000.000 und der negativen Summe aus dem jeweils zufällig simulierten inländischen $\textit{Jahresüberschuss}_{2025}$ zuzüglich 1. Der obere Rand des Intervalls ergibt sich entsprechend als der Maximumwert dieser beiden Größen. Exemplarisch für das Jahr 2025 gestaltet sich das wie folgt:

Im Verlustunterfall P werden die jeweiligen Intervalle anders definiert, sodass bei der Generierung von Zufallszahlen die Bedingung eines positiven Gesamtergebnisses aus beiden Staaten (bzw. dem Stammhaus und der Betriebsstätte im Inland) erfüllt wird. Der untere Rand des Intervalls entspricht dem Minimum aus der Summe des negativen Werts des jeweiligen zufälligen $\textit{Jahresüberschusses}_{2025}$ in Deutschland zuzüglich 1 und –1. Der obere Rand ergibt sich aus dem größeren der beiden Werte.

Die Intervallauswahl stellt die Prämissen der beiden Unterfälle der Simulation sicher, nämlich eine positive bzw. negative Gesamtsumme aus In- und Ausland als Ausgangsgröße. Des Weiteren ist die systematische Einbeziehung von Minimum und Maximum im Intervall notwendig, damit eine mathematisch zulässige Definition des Intervalls stattfindet, die die Bedingung $a < b$ für die Gleichverteilung im Intervall $[a;b]$ erfüllt. Die Ermittlung der Zufallsverluste für die beiden folgenden Verlustperioden (Jahr 2026 und Jahr 2027) im Verlustunterfall N und P erfolgt nach den gleichen Wahrscheinlichkeitsverteilungen wie für das Jahr 2025. Dabei wird jedoch der inflationsbedingte angepasste Jahresüberschuss in Deutschland in den jeweiligen Intervallen berücksichtigt.

Für die Jahre 2028 und 2029 ergibt sich der handelsrechtliche Jahresüberschuss in Bulgarien in beiden Verlustunterfällen (N und P) als die Differenz zwischen der inflationsbedingt erhöhten Jahresobergrenze für die Jahre 2028 bzw. 2029 (basierend auf der Obergrenze von 1.000.000 € im Jahr 2025) und dem zufällig simulierten Jahresüberschuss in Deutschland (Zufallsvariablen $\textit{Jahresüberschuss}_{2028}$ bzw. $\textit{Jahresüberschuss}_{2029}$). Somit ist der Gesamtjahresüberschuss aus Bulgarien und Deutschland in den letzten zwei Jahren des fünfjährigen Zeithorizonts auf die im Gewinnfall festgelegte Höchstgrenze begrenzt. Dadurch werden unrealistisch hohe Gesamtgewinne in den letzten zwei Perioden ausgeschlossen. Ferner wurde innerhalb der einzelnen Verlustunterfälle P und N sichergestellt, dass über alle Verlustfälle hin-

$$\textbf{\textit{Auslandsverlust}}_{2025} \text{ im Unterfall N} \sim U\left[\text{Min}(-1.000.000;\ -(\textit{Jahresüberschuss}_{2025}+1));\ \text{Max}(-1.000.000;\ -(\textit{Jahresüberschuss}_{2025}+1))\right]$$

Abbildung 7: Wahrscheinlichkeitsverteilung der Zufallsvariable „Auslandsverlust" im Jahr 2025 im Verlustunterfall N

$$\textbf{\textit{Auslandsverlust}}_{2025} \text{ im Unterfall P} \sim U\left[\text{Min}(-1;\ 1-\textit{Jahresüberschuss}_{2025});\ \text{Max}(-1;\ 1-\textit{Jahresüberschuss}_{2025})\right]$$

Abbildung 8: Wahrscheinlichkeitsverteilung der Zufallsvariable „Auslandsverlust" im Jahr 2025 im Verlustunterfall P

weg sowie im Basisverlustfall in jeder Iteration die identischen Zufallsvariablen für die Berechnung der Veranlagungssimulation herangezogen werden.

Die Monte-Carlo-Simulation wurde in jedem Gewinn- und Verlustszenario jeweils 5 000[148] Mal durchgeführt. Bei diesen zahlreichen Iterationen wurden ebenso viele unterschiedliche Kombinationen von Zufallszahlen generiert und die Steuerbelastung in jedem Durchlauf berechnet. Dabei ist hervorzuheben, dass es sich bei einer derart hohen Anzahl an Simulationsdurchläufen eine gemäß dem empirischen Gesetz der großen Zahlen[149] genügend stabile Wahrscheinlichkeitsverteilung der möglichen Ergebnisse ergibt. Daher ist der statistische Standardfehler des geschätzten Mittelwerts[150] klein genug, um auf Basis dessen die Ergebnisse in den folgenden Abschnitten verlässlich zu beurteilen.

3.3 Darstellung und Auswertung der betrachteten Gewinnszenarien

Die Ergebnisse der Simulation bzw. der Verteilungen der Variable *Barwert der Gesamtsteuerbelastung* im Gewinnszenario sind in einem Box-Whisker-Plot grafisch veranschaulicht (s. Abbildung 9).

Der Box-Whisker-Plot ermöglicht den besseren Vergleich der Fälle und stellt die zentralen Tendenzen hinsichtlich Lage und Streuung der Daten dar.[151]

Der größere Interquartilsabstand bzw. die größere Box in den Box-Whisker-Plots für die Geschäftsausweitung im Inland und die bulgarische Kapitalgesellschaft indiziert eine größere Streuung der erhaltenen Ergebnisse aus der Simulation. Dieser Umstand lässt sich dadurch erklären, dass der progressive Einkommensteuertarif einen stärkeren Einfluss auf die Gesamtsteuerbelastung hat als in den beiden anderen Fällen, bei denen auf den Auslandsanteil des Gesamtgewinns die Freistellungsmethode und damit der lineare bulgarische Einkommensteuertarif Anwendung findet.

Der Box-Whisker-Plot-Vergleich zeigt, dass das Auslandsengagement in Bulgarien in allen drei Formen steuerlich günstiger ist als eine rein deutschlandweite Geschäftsausweitung. Die gestrichelte Linie in den Boxplots zeigt den Mittelwert der Variable *Barwert der Gesamtsteuerbelastung* an.

Die entsprechenden genauen Mittelwerte sind in der nachfolgenden Tabelle 3 dargestellt und werden im Folgenden weiterführend erörtert. In der zweiten Zeile der Tabelle sind die prozentualen Abweichungen der durchschnittlichen Barwerte der Gesamtsteuerbelastung im Vergleich zum Basisfall (deutschlandweite Geschäftsausweitung) dargestellt.

Die Errichtung einer Betriebsstätte in Bulgarien erweist sich auf Basis des durchschnittlichen Barwerts der Gesamtsteuerbelastung als die steuerlich vorteilhafteste Option. In diesem Fall setzt sich die Gesamtsteuerbelastung aus den deutschen Ertragsteuern auf die inländischen Gewinne und der zehnprozentigen bulgarischen Körperschaftsteuer auf den Betriebsstättengewinn zusammen (s. deterministisches Modell in Tabelle 13, S. XV). Die bulgarischen Einkünfte wirken sich dabei in Deutschland im Rahmen des PVB erhöhend auf den anwendbaren Einkommensteuersatz aus. Beispielsweise liegt die Erhöhung in der ersten Periode bei 1,97 % (Mittelwert des durchschnittlichen Steuersatzes ohne PVB: 39,17 % und mit PVB: 41,14 %).

Eine zentrale Erkenntnis aus den Ergebnissen ist, dass die Errichtung einer bulgarischen Betriebsstätte einen Vorteil von ca. 26,83 % auf Basis der durchschnittlichen Barwerte der Gesamtsteuerbelastung gegenüber einer inländischen Betriebsstätte aufweist.

Bei der Gründung einer bulgarischen Personengesellschaft liegt der durchschnittliche Barwert der Gesamtsteuerbelastung höher als bei der bulgarischen Betriebsstätte. Dies ist vor dem Hintergrund der vorangehend erfolgten Beurteilung des steuerrechtlichen Rahmens nicht überraschend, da die anfallende bulgarische Kapitalertragsteuer in Deutschland nicht anrechenbar ist und somit zu dieser Mehrbelastung beiträgt (s. deterministisches Modell in Tabelle 14, S. XVIII).

In dem Fall der bulgarischen Kapitalgesellschaft erhöht

[148] Aufgrund der begrenzten Kapazität der Software ist 5 000 die maximale mögliche Iterationsanzahl.

[149] Demnach führt die zunehmende Anzahl der Versuche zu einer Stabilisierung der relativen Häufigkeiten eines Ereignisses; weiterführend Büchter und Henn (2007), S. 174.

[150] Das liegt daran, dass zur Ermittlung des statistischen Standardfehlers die Standardabweichung der Stichprobe durch die Stichprobengröße (in diesem Fall 5 000) geteilt wird.

[151] Vgl. Becker et al. (2024), S. 65.

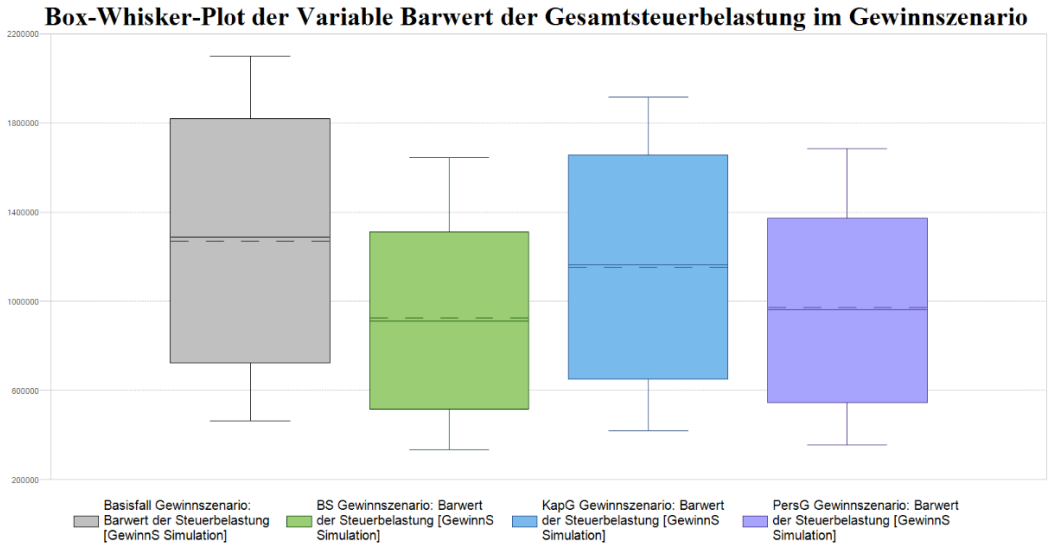


Abbildung 9: Box-Whisker-Plot der Variable Barwert der Gesamtsteuerbelastung im Gewinnszenario

Tabelle 3: Barwert der Gesamtsteuerbelastung im Gewinnszenario

Der prozentuale Unterschied wird wie folgt ermittelt: (Wert im jeweiligen Fall − Wert im Basisfall) / Wert im Basisfall × 100.

Gewinnszenario: Barwert der Gesamtsteuerbelastung					
Variable \ Fälle		Betriebsstätte	PersG	KapG	Basisfall
Barwert der Gesamtsteuerbelastung	$\bar{x}$	921.667	966.814	1.144.316	1.259.543
Abweichung vom Basisfall		–26,83 %	–23,24 %	–9,15 %	-

sich die durchschnittliche steuerliche Belastung im Vergleich zu den anderen Direktinvestitionsformen weiter. Das lässt sich damit begründen, dass die bulgarische Dividende auf Ebene des deutschen Investors zu 60 Prozent dem progressiven deutschen Tarif unterworfen wird (s. deterministisches Modell in Tabelle 15, S. XXI). Der prozentuale steuerliche Vorteil gegenüber dem Basisfall fällt somit zwar schwächer aus, ist aber dennoch vorhanden und beläuft sich auf ca. 9,15 %. Trotz der geltenden Abschirmung zwischen der Gesellschafts- und Gesellschafterebene und der damit verbundenen höheren Steuerlast bei Kapitalgesellschaften erweist sich die Rechtsform im Durchschnitt dennoch als steuerlich günstiger als der Basisinlandsfall.

Bei der inländischen Geschäftsausweitung vollzieht sich die Besteuerung nach den deutschen Steuervorschriften, d. h., der Gesamtgewinn unterliegt der Einkommensteuer, der Gewerbesteuer und dem Solidaritätszuschlag, wobei sich die Gewerbesteuerbelastung infolge der Ermäßigung nach § 35 EStG nur in geringem Maße auswirkt (s. deterministisches Modell in Tabelle 16, S. XXIII).

Ferner ist aus den Box-Whisker-Plots (s. Abbildung 9) ersichtlich, dass der minimale Barwert der Gesamtsteuerbelastung bei der Errichtung einer bulgarischen Betriebsstätte erzielt wird.[152] In diesem Zusammenhang ist es besonders relevant, das Minimum im identifizierten steuerlich günstigsten Fall der Betriebsstätte auf Basis der effektiven Steuerbelastung in Relation zum Gesamtjahresüberschuss unter den gegebenen Modellannahmen zu bestimmen (s. Tabelle 4).

Tabelle 4: Gewinnszenario: Relative effektive Steuerbelastung

Gewinnszenario: Relative effektive Steuerbelastung		
	Betriebsstätte	Basisfall
Minimumwert	21,41 %	38,32 %

Mit der Gründung einer bulgarischen Betriebsstätte lässt sich unter den getroffenen Modellannahmen eine effektive Steuerbelastung in Höhe von 21,41 % im Minimum erreichen. Im Vergleich dazu beträgt die minimale relative effektive Steuerbelastung bei einer inländischen Geschäftsausweitung 38,32 %. Diese Differenz ist ausschließlich auf den Effekt des progressiven Tarifs sowie auf einen geringeren Gewinn in der jeweiligen Iteration zurückzuführen.

[152] In diesem Fall ist der minimale Wert des Datensatzes durch den unteren Whisker des jeweiligen Box-Whisker-Plots dargestellt.

Vor diesem Hintergrund ist es für die Analyse relevant, wie sich der Auslandsanteil am Gesamtjahresüberschuss aus In- und Ausland auf die Variable *relative effektive Steuerbelastung* auswirkt. Um den Effekt zu isolieren, wird der Auslandsanteil – ceteris paribus –innerhalb definierter Bandbreiten über die Iterationen hinweg variiert. Daraus ergibt sich: Je höher der Auslandsanteil ist, desto geringer fällt die relative effektive Steuerbelastung aus. Dies ist in der folgenden Tabelle 5 auf Basis von simulierten Daten in den dargelegten Bandbreiten des Auslandsanteils quantitativ eindeutig belegt.

Selbst bei einem Auslandsanteil zwischen 20 und 40 % kann die Steuerbelastung mit knapp zehn Prozentpunkten durch die Gründung einer bulgarischen Betriebsstätte im Vergleich zu dem Fall einer deutschen Betriebsstätte gesenkt werden. Bei einem Auslandsanteil zwischen 40 und 60 % kann eine durchschnittliche effektive Steuerbelastung von 26,56 % erreicht werden, was mit knapp 17 Prozentpunkten unter dem Basisfall liegt.

Im Fall der Betriebsstätte zeigen die statistischen Auswertungen deutlich, dass eine Erhöhung der Auslandsaktivität zu einer noch niedrigeren durchschnittlichen relativen effektiven Steuerbelastung führt (s. Tabelle 5). Im Einklang mit dieser Aussage korreliert im Fall einer bulgarischen Betriebsstätte die Variable *relative effektive Steuerbelastung* stark negativ mit dem Auslandsanteil am Gesamtjahresüberschuss (linearer Korrelationskoeffizient: –0,907 bei einem Auslandsanteil von 40–60 %). Auch im Fall der Personengesellschaft ist die negative Korrelation weiterhin stark, fällt aber etwas schwächer aus als bei der Betriebsstätte (–0,881 bei einem Auslandsanteil von 40–60 %). Bei der bulgarischen Kapitalgesellschaft hingegen ist die negative lineare Korrelation mit –0,386 unter denselben Voraussetzungen deutlich schwächer ausgeprägt. Dies ist damit zu begründen, dass die ausländische Dividende dem TEV unterliegt und daher zu 60 Prozent dem Einkommensteuertarif unterworfen wird. Im Einklang mit den Annahmen des Modells korreliert der Auslandsanteil mit der relativen effektiven Steuerbelastung im Basisfall nicht.

Im Rahmen des Gewinnszenarios ist darüber hinaus der Effekt aus der Berücksichtigung der Rückstellungshöhe bei der Veranlagung auszuwerten.[153] Im Basisfall führt die Berücksichtigung der Rückstellungen zu keiner Änderung des Barwerts der Gesamtsteuerbelastung, da die Rückstellungen bereits gewinnmindert im handelsrechtlichen bzw. steuerlichen Gewinn erfasst sind. Die Ergebnisse zu dem Barwert der Gesamtsteuerbelastung sind in der folgenden Tabelle 6 dargelegt:

Bei Gegenüberstellung mit den zuvor analysierten Werten (s. Tabelle 3) zeigt sich die durchschnittliche Differenz infolge der Rückstellungsberücksichtigung am stärksten im Fall der bulgarischen Betriebsstätte (absoluter Rückgang der Abweichung vom Basisfall um 2,60 %). Dies liegt daran, dass die steuerliche Bemessungsgrundlage für die Körperschaftsteuer in Bulgarien durch die Zurechnung der Rückstellungen erhöht wird. Im Fall der Personengesellschaft fällt die Differenz im Vergleich zur Variante ohne Rückstellungsberücksichtigung etwas geringer aus (absoluter Rückgang um 2,56 %). Durch die Rückstellungsberücksichtigung erhöht sich zwar die bulgarische Körperschaftsteuer aufgrund der erweiterten Bemessungsgrundlage, gleichzeitig sinkt jedoch die ausschüttbare Dividende und somit auch die darauf erhobene Steuer. Die durchschnittliche Differenz zum Fall ohne Rückstellungseinbezug wird durch diesen Effekt beeinflusst.

Bei der Kapitalgesellschaft wirkt sich die Zusatzsteuerlast durch die steuerliche Nichtberücksichtigung der Rückstellungen in Bulgarien noch schwächer aus als in den anderen bulgarischen Gestaltungsfällen (absolute Differenz zum Gewinnfall ohne Rückstellungsberücksichtigung in Höhe von 1,87 %). Dies ist auf den oben genannten Effekt aus der Dividende im Fall der bulgarischen Personengesellschaft sowie auf die anwendbare Anrechnungsmethode (und nicht die Freistellung unter PVB) in Deutschland zurückzuführen.

Es wird evident, dass unter Berücksichtigung der Bemessungsgrundlagendifferenzen der durchschnittliche prozentuale Steuervorteil im Rahmen des Auslandsengagements gemindert wird. Jedoch ist eine Geschäftsausweitung in diesem Fall weiterhin steuerlich ungünstiger als die Direktinvestition in Bulgarien.

Es sollte an dieser Stelle betont werden, dass die Zusatzsteuerbelastung durch die Rückstellungsberücksichtigung temporär ist und in den künftigen Perioden in einer realen Konstellation ausgeglichen wird. Dennoch könnte diese frühzeitige Zahlung der Steuerbelastung auf Rückstellungen zu einer geringeren Liquidität in der betreffenden Periode führen. Dies kann insbesondere bei Geschäftsmodellen mit hohem Bestand an jährlich zu bildenden Rückstellungen von Bedeutung sein.

3.4 Darstellung und Auswertung der betrachteten Verlustszenarien

Nachfolgend wird das pessimistische Verlustszenario ausgewertet, in dem Anlaufverlusten bei der inländischen bzw. ausländischen Geschäftsausweitung berücksichtigt werden.

Zuerst werden die Ergebnisse des Verlustunterfalls N erörtert, bei dem das Gesamtergebnis aus der ausländischen Ausweitung und dem Inlandstätigkeit summarisch negativ ist. Auf Basis des durchschnittlichen Barwerts der Gesamtsteuerbelastung in den vier untersuchten Fällen ergibt sich im Vergleich zu dem vorangehend analysierten Gewinnszenario ein verändertes Bild. Führt die Summe der Jahresergebnisse aus der Investition im In- oder Ausland und der Haupttätigkeit in Deutschland zu einer negativen Summe, erweist sich die Geschäftsausweitung in Deutschland als steuerlich am günstigsten, denn sie führt zu dem geringsten Mittelwert des Barwerts der Gesamtsteuerbelastung. Die Ergebnisse sind in der folgenden Tabelle 7 dargestellt.

Aufgrund der direkten Verrechnung des Verlusts aus der inländischen Betriebsstätte mit dem Gewinn des Stammhau-

[153] Die deterministischen Modelle, auf denen die Simulation basiert, sind für den Betriebsstättenfall in Tabelle 17 (s. S. XXV), für den Fall der Personengesellschaft in Tabelle 18 (s. S. XXVII) und für die Kapitalgesellschaft Tabelle 19 (s. S. XXIX) dargestellt.

Tabelle 5: Änderung der relativen effektiven Steuerbelastung bei verändertem Auslandsanteil

Die geringwertigen Abweichungen im Basisfall sind lediglich ein Resultat der Streuung zwischen den Simulationen.

Änderung der relativen effektiven Steuerbelastung bei verändertem Auslandsanteil				
Fälle	**Betriebsstätte**	**PersG**	**KapG**	**Basisfall**
Auslandsanteil	**Relative effektive Steuerbelastung ($\bar{x}$)**			
10–20 %	38,15 %	38,83 %	41,39 %	43,10 %
20–40 %	33,16 %	34,51 %	39,67 %	43,09 %
40–60 %	26,56 %	28,81 %	37,41 %	43,10 %

Tabelle 6: Gewinnszenario: Barwert der Gesamtsteuerbelastung unter Rückstellungsberücksichtigung

Gewinnszenario: Barwert der Gesamtsteuerbelastung					
Variable \ **Fälle**		**Betriebsstätte**	**PersG**	**KapG**	**Basisfall**
Barwert der Gesamtsteuerbelastung	$\bar{x}$	954.323	997.836	1.167.806	1.259.543
Abweichung vom Basisfall		–24,23 %	–20,78 %	–7,28 %	-

Tabelle 7: Verlustunterfall N: Barwert der Gesamtsteuerbelastung

Verlustunterfall N: Barwert der Gesamtsteuerbelastung					
Variable \ **Fälle**		**Betriebsstätte**	**PersG**	**KapG**	**Basisfall**
Barwert der Gesamtsteuerbelastung	$\bar{x}$	764.338	802.590	1.493.395	609.037
Abweichung vom Basisfall		25,50 %	31,78 %	145,21 %	-

ses fallen im Rahmen des inländischen Einheitsunternehmens bzw. im Basisfall in den Verlustjahren keine Steuern an (s. deterministisches Modell in Tabelle 23, S. XLI). Erst in den nach den Modellannahmen simulierten folgenden Gewinnjahren wird der Gewinn besteuert, jedoch unter Abzug der akkumulierten gewerbe- und einkommensteuerlichen Verlustvorträge.

In der zweiten Zeile der Tabelle 7 ist die prozentuale Abweichung des durchschnittlichen Barwerts der Gesamtsteuerbelastung der drei Direktinvestitionsformen im Vergleich zum durchschnittlichen Barwert der Gesamtsteuerbelastung im Basisfall (der inländischen Geschäftsausweitung) aufgeführt. Dieser Wert liegt bei der bulgarischen Betriebsstätte mit etwa 25,50 % über dem des Basisfalls. Dennoch erweist sich in diesem Verlustunterfall N die Gründung einer bulgarischen Betriebsstätte als die steuerlich vorteilhafteste Form der Direktinvestition in Bulgarien. Die bulgarischen Verluste dürfen zwar in Deutschland nicht direkt berücksichtigt werden, diese entfalten aber ihre steuermindernde Wirkung auf den Einkommensteuersatz in Deutschland im Rahmen des negativen PVB. In den Verlustjahren bewirkt der negative PVB im jeweiligen Verlustjahr eine Ermäßigung des durchschnittlichen inländischen Einkommensteuersatzes von einem Mittelwert in Höhe von etwa 41,20 % auf 0 % (s. deterministisches Modell in Tabelle 20, S. XXXII). In Bulgarien wird der akkumulierte Verlustvortrag in den Gewinnjahren steuerlich berücksichtigt. Es ist jedoch zu beachten, dass die bulgarischen Verlustvorträge in den folgenden Gewinnjahren im Rahmen des PVB nicht berücksichtigt werden, sodass bei positivem handelsrechtlichem Ergebnis in Bulgarien der durchschnittliche Steuersatz auf das zu versteuernde Einkommen in Deutschland steigt. Im Vergleich zum rein inländischen Fall, bei dem im Verlustjahr keine Gewerbesteuer anfällt, entsteht im Falle der ausländischen verlustbringenden Betriebsstätte aufgrund der Nichtberücksichtigung der ausländischen Verluste ein positiver Gewerbeertrag in Deutschland, der besteuert wird.

Im Fall einer bulgarischen Personengesellschaft fällt der durchschnittliche Barwert der Gesamtsteuerbelastung höher aus als bei der bulgarischen Betriebsstätte. Dies ist auf die fünfprozentige in Deutschland nicht anrechenbare Kapitalertragsteuer zurückzuführen, die in den Gewinnjahren auf ausgeschüttete Dividenden erhoben wird. Die Besteuerung in Deutschland ist ansonsten identisch mit dem Betriebsstättenfall (s. deterministisches Modell in Tabelle 21, S. XXXV).

Bei der Gründung einer bulgarischen Kapitalgesellschaft werden dagegen aufgrund des Trennungsprinzips die in Bulgarien erzielten Verluste in Deutschland gar nicht berücksichtigt (s. deterministisches Modell in Tabelle 22, S. XXXVIII). In Deutschland erfolgt die Besteuerung des deutschen Gewinns, und in Bulgarien werden Verlustvorträge gebildet. In den Gewinnjahren erfolgt eine Gewinnausschüttung, die analog zum Vorgehen im Gewinnszenario besteuert wird. Aufgrund der akkumulierten Verlustvorträge fällt in Bulgarien bis zu deren vollständigen Ausschöpfung keine Körperschaftsteuer an. Diese Konstellation führt zu einer sehr hohen prozentualen Abweichung vom Basisfall, gemessen am durchschnittlichen Barwert der Steuerbelastung (s. Tabelle 7). Es sollte

an dieser Stelle beachtet werden, dass die bulgarischen Verlustvorträge zeitlich auf fünf Jahre begrenzt sind, was den Zeithorizont der Simulation übertrifft und deshalb als Effekt nicht berücksichtigt wird.

Des Weiteren ist der Verlustunterfall P zu untersuchen. Unter den getroffenen Modellannahmen führt die Summe der Ergebnisse aus der stets gewinnbringenden inländischen Haupttätigkeit und der inländischen Geschäftsausweitung bzw. ausländischen Direktinvestition summarisch zu einem positiven Gesamtergebnis.

Es soll an dieser Stelle betont werden, dass aufgrund der Modellannahmen die über alle fünf Perioden aggregierten Ergebnisse in den beiden Unterfällen P und N zwingend unterschiedlich ausfallen, so dass die absoluten Beträge der Steuerbelastung aus den Unterfällen nicht unmittelbar miteinander vergleichbar sind. Die Ergebnisse sind in der nachfolgenden Tabelle 8 veranschaulicht.

In diesem Verlustunterfall ändert sich die Reihenfolge der steuerlichen Vorteilhaftigkeit der Direktinvestitionsformen im Vergleich zum anderen Unterfall N. Auf Basis des durchschnittlichen Barwerts der Gesamtsteuerbelastung erweist sich die Gründung einer ausländischen Betriebsstätte als steuerlich am vorteilhaftesten. Die Abweichung vom Basisfall beträgt jedoch lediglich 1,11 %.

Dabei ist zu beachten, dass es sich um Mittelwerte auf Basis von 5.000 Iterationen handelt, die alle fünf betrachteten Perioden umfassen. Die konkrete Aussage, ob die ausländische Betriebsstätte oder Personengesellschaft im Gesamtbild vorteilhafter ist als die deutsche Geschäftsausweitung, hängt von den Beträgen im Einzelfall ab.

Die Nichtberücksichtigung der Verluste im Rahmen des Auslandsengagements wirkt sich auch in diesem Verlustunterfall steuerlich ungünstig aus. Die relative effektive Steuerbelastung im ersten Verlustjahr bei der Gründung einer ausländischen Betriebsstätte liegt im Durchschnitt bei 179,96 %, während sie im Fall der inländischen Geschäftsausweitung 35,68 % beträgt.

Ursächlich hierfür ist, dass im Fall der ausländischen Betriebsstätte trotz des negativen Ergebnisses in Bulgarien auf den in Deutschland erzielten Gewinn Steuern anfallen.[154] Im Rahmen des negativen PVB im Fall der bulgarischen Betriebsstätte bzw. Personengesellschaft wird der anwendbare Einkommensteuersatz lediglich ermäßigt. In der ersten Periode wird der Mittelwert des durchschnittlichen Einkommensteuersatzes, der auf das zu versteuernde Einkommen anzuwenden ist, von 41,20 % auf 34,22 % gesenkt. Die festzusetzende Einkommensteuer wird anschließend durch die Anrechnung der Gewerbesteuer gemindert.

Allerdings kann dieser nachteilige anfängliche Effekt aufgrund der hohen Steuerbelastung in Deutschland in den Gewinnjahren abgeschwächt werden, insbesondere wenn der Auslandsanteil am Gesamtjahresüberschuss in den Gewinnjahren prozentual höher ausfällt als der inländische Anteil.

Die Besteuerung erfolgt analog zum anderen Verlustunterfall N. Aufgrund der positiven Gesamtsumme der Ergebnisse aus In- und Ausland treten jedoch die im Folgenden kurz dargelegten Effekte auf. Analog zum Verlustunterfall N werden im Fall der inländischen Geschäftsausweitung die Verluste der inländischen Betriebsstätte mit den Gewinnen des Inlandsstammhauses verrechnet (s. deterministisches Modell in Tabelle 27: S. L). In diesem Fall ergibt sich ein um die Verluste gemindertes, aber dennoch positives steuerliches Ergebnis, das in Deutschland mit der Einkommen- und Gewerbesteuer besteuert wird. Daher werden keine Verlustvorträge gebildet, und in den beiden Gewinnjahren unterliegt der volle Gewinn der Besteuerung.

Bei der bulgarischen Personen- und Kapitalgesellschaft[155] wird in den Gewinnjahren auf die Gewinnrepatriierung eine Kapitalertragsteuer in Bulgarien erhoben. Bei der Personengesellschaft stellt diese eine definitive nicht anrechenbare Steuerbelastung dar. Die bei der bulgarischen Veranlagung aus den ersten drei Perioden akkumulierten Verlustvorträge führen dazu, dass die zehnprozentige körperschaftsteuerliche Belastung bis zu deren vollständigen Ausschöpfung ausbleibt. Der Verlustvortrag ist in der Regel noch im Rahmen des betrachteten Horizonts von zwei folgenden Gewinnjahren ausgeschöpft, sodass in Bulgarien ein steuerpflichtiger Gewinn im Rahmen des fünfjährigen Zeithorizonts gebildet und besteuert wird

Im Verlustszenario zeigt sich, dass die Abkehr vom Welteinkommensprinzip infolge der DBA-Freistellung, die im Gewinnszenario zu einem Steuervorteil führt, im Verlustfall für den deutschen Investor nachteilig ist. Es gilt zudem: Ist die Gesamtsumme der Jahresergebnisse aus der bestehenden Inlandstätigkeit und der geplanten Investition negativ, so fällt der prozentuale Unterschied im durchschnittlichen Barwert der Gesamtsteuerbelastung zum Basisfall gegenüber den Fällen des Auslandsengagements höher aus als in dem Fall eines positiven Gesamtergebnisses.

Es sollte jedoch bedacht werden, dass auf Basis der Verlustszenarien allein keine Gesamtaussage über die steuerliche Vorteilhaftigkeit des bulgarischen Auslandsengagements getroffen werden kann. Dies liegt daran, dass eine Direktinvestition stets mit Gewinnerzielungsabsicht erfolgt.

Aus dem Verlustunterfall P lässt sich zudem folgende Erkenntnis ableiten: In einem realen praktischen Fall mit einer erfolgreichen Direktinvestition werden auf mögliche Anlaufverluste Gewinne folgen, die diesen anfänglichen Effekt bei einem längeren Betrachtungshorizont voraussichtlich ausgleichen und langfristig den steuerlichen Vorteil bewirken, der sich im Gewinnszenario zeigt.

Vor diesem Hintergrund erscheint es für mittelständische deutsche Unternehmen empfehlenswert, im Rahmen der Entscheidungsfindung bzw. Steuerplanung eines Auslandsengagements Verlustszenarien einzubeziehen, sodass der steuerliche Vorteil über einen langfristigen Zeithorizont verlässlicher eingeschätzt werden kann.

[154] Siehe die deterministischen Modelle in Tabelle 24 in Bezug auf die bulgarische Betriebsstätte (s. S. XLIII) und Tabelle 25 für die bulgarische Personengesellschaft (s. S. XLVI).

[155] Siehe das deterministische Modell für den Fall der bulgarischen Kapitalgesellschaft in Tabelle 26 (s. S. XLVIII).

Tabelle 8: Verlustunterfall P: Barwert der Gesamtsteuerbelastung

Verlustunterfall P: Barwert der Gesamtsteuerbelastung					
Variable \ Fälle		Betriebsstätte	PersG	KapG	Basisfall
Barwert der Gesamtsteuerbelastung	$\bar{x}$	1.216.274	1.252.997	1.515.344	1.229.950
Abweichung vom Basisfall		–1,11 %	1,87 %	23,20 %	-

3.5 Grenzen der Simulation

Die durchgeführte Monte-Carlo-Simulation lässt bewusst unternehmensspezifische Merkmale außer Acht, um allgemeingültige und branchenunabhängige Aussagen zu ermöglichen. In der Praxis sollte jedoch ein deutscher Investor, der ein Auslandsengagement in Bulgarien in Betracht zieht, auf Basis seines Einzelfalls zusätzlich zu den Steuervorteilen auch weitere lokale, für die jeweilige Geschäftstätigkeit wesentliche Anreize im Rahmen der Entscheidungsfindung heranziehen. Darunter fallen unter anderem Kosten für die Produktion oder den Vertrieb (bspw. niedrigere Stromkosten, Personalkosten, Transportkosten), potenzielle staatliche Förderprogramme, der erleichterte Zugriff auf bestimmte Rohstoffe oder Absatzmärkte,[156] weitere ggf. relevante Steuerbegünstigungen (bspw. für nachhaltige Investitionen) und Ähnliches. Im Rahmen einer Kosten-Nutzen-Analyse sind diesen Standort- und Steuervorteilen die Kosten für Rechtsbeistand, Überwachung, Steuerberatung, Gründung sowie Investitionsausgaben usw. gegenüberzustellen.

4 Fazit

In der vorliegenden Arbeit konnte festgestellt werden, dass eine Direktinvestition in Bulgarien im Vergleich zu einer Geschäftsausweitung in Deutschland steuerliche Vorteile für deutsche mittelständische Unternehmen bietet.

Im Rahmen der Arbeit wurde der steuerrechtliche Rahmen der drei Gestaltungsformen einer Direktinvestition in Bulgarien dargelegt. Dadurch wurden die Besteuerungsunterschiede in Deutschland und Bulgarien sowie die Komplexität grenzüberschreitender Tätigkeiten exemplarisch verdeutlicht.

Die im Anschluss durchgeführte Monte-Carlo-Simulation konnte die Erkenntnisse dieser qualitativen Analyse des steuerrechtlichen Rahmens quantitativ bestätigen und erweitern. Dementsprechend führt die Gründung einer bulgarischen Betriebsstätte innerhalb von fünf Jahren zu dem geringsten durchschnittlichen Barwert der Gesamtsteuerbelastung im Vergleich zu den anderen Formen der Direktinvestition, nämlich der bulgarischen Personen- und Kapitalgesellschaft sowie der rein inländischen Geschäftsausweitung. Dabei konnte ein durchschnittlicher prozentualer steuerlicher Vorteil in Höhe von 26,83 % gegenüber einer inländischen Geschäftsausweitung festgestellt werden.

Die Analyse zeigt deutlich, dass in allen Fällen mit einem höheren Auslandsanteil am Gesamtjahresüberschuss aus der bestehenden inländischen Tätigkeit und der bulgarischen Direktinvestition eine niedrigere relative effektive Besteuerung erreicht werden kann. Bei der Berücksichtigung der Unterschiede in der steuerlichen Behandlung der Rückstellungen in beiden Ländern ändert sich die Gesamtaussage der Vorteilhaftigkeit nicht.

Im Szenario mit Anlaufverlusten aus der geplanten Geschäftstätigkeit zeigt sich ein verändertes Bild. Die Nichtberücksichtigung der ausländischen Verluste führt in den Verlustjahren zu einer höheren Steuerbelastung als bei einer inländischen Geschäftsausweitung. Der Effekt wird verstärkt, wenn das Gesamtergebnis aus der bestehenden Haupttätigkeit und der geplanten Geschäftsaktivität negativ ist.

Es sollte jedoch beachtet werden, dass eine Direktinvestition in der Regel mit einer Gewinnerzielungsabsicht erfolgt. Vor diesem Hintergrund kann auf Basis des Verlustszenarios keine Gesamtaussage über die Vorteilhaftigkeit der Direktinvestition in Bulgarien gemacht werden. Das Szenario stellt einen relevanten Hinweis dar, der im Rahmen einer realistischen Steuerplanung von deutschen Investoren gewürdigt werden sollte.

Es ist zu betonen, dass die durchgeführte Simulation unmöglich alle relevanten Faktoren bei einer Direktinvestition in Bulgarien abdecken kann. Im Rahmen der Analyse wurde ein allgemeines und branchenunabhängiges Modell angestrebt. Im Einzelfall sollte die Entscheidungsfindung stets unter Berücksichtigung der unternehmensspezifischen und der für das jeweilige Geschäftsmodell relevanten standortspezifischen Faktoren erfolgen.

Nichtsdestotrotz zeigt die Masterarbeit eindeutig das steuerliche Potenzial Bulgariens als attraktiven Investitionsstandort für mittelständische deutsche Unternehmen auf. Vor dem Hintergrund der Ergebnisse erscheint die in der Einleitung ausgeführte Tatsache erwähnenswert, dass Deutschland nach wie vor der meistgewählte Investitionsstandort mittelständischer Unternehmen ist. Es bleibt abzuwarten, ob der deutsche Mittelstand das Potenzial ausländischer Direktinvestitionen in Osteuropa künftig stärker erkennt und sich der Status quo, wie prognostiziert, in den kommenden Jahren tatsächlich verändern wird.

5 Summary

The master's thesis investigates the tax advantages that a direct investment in Bulgaria offers medium-sized German companies compared to a business expansion in Germany.

[156] Vgl. Pfaffermayr (1996), S. 7f.

The study begins with a qualitative analysis of the legal tax framework of the three forms of direct investment in Bulgaria. The analysis shows that different tax effects arise depending on which of the three direct investment alternatives is chosen and outlines the complexity of international tax cases, which partly stems from the tax systems and understanding of the countries involved.

A Monte Carlo simulation of two main scenarios over a five-year period was conducted to quantify the industry-independent tax potential of a direct investment in Bulgaria. The first simulation focuses on the taxation of profits, while the second examines the tax treatment in the case of initial losses within the early stages of the new business activity. In the simulation, the present value of the tax burden over a five-year period was determined for the three forms of direct investment and compared to a domestic German business expansion.

The simulation confirmed that a direct investment in Bulgaria is more tax-advantageous than an expansion within Germany. The most favourable case is achieved through a permanent establishment in Bulgaria, which results in an average tax advantage in relation to the domestic case of 26.83 %. However, this statement only holds true if the foreign activity generates profits. In the scenario with initial losses, the domestic business expansion leads to the lowest present tax burden during the loss years. This effect is stronger if the overall result from the existing main activity in Germany and the planned investment is negative. However, based only on the pessimistic loss scenario, no overall assessment can be made as a foreign investment is conducted with the goal of generating profits. It should only be used as a guidance for tax planning by medium-sized enterprises.

It should be emphasised that the simulation only highlights the general tax potential of Bulgaria as an investment location from the perspective of German medium-sized enterprises. In practice, such a decision should always be made on a case-by-case basis, considering the company- and location-specific factors.

In light of the results, the fact that Germany remains the most popular investment location among medium-sized companies seems surprising. It remains to be seen whether German medium-sized companies will become increasingly aware of the potential of foreign direct investment within Eastern Europe in the future and whether the current situation will change in the coming years as predicted.

Literatur

Abel-Koch, J. (2023). *Erschließung neuer Absatzmärkte wichtigster Grund für mittelständische Auslandsinvestitionen* (Fokus Volkswirtschaft Nr. 446). KfW Research.

Becker, T., Herrmann, R., Heumann, C., Pilz, S., Sandor, V., Schäfer, D., & Wellisch, U. (2024). *Stochastische Risikomodellierung und statistische Methoden: Angewandte Stochastik für die aktuelle Praxis* (2., überarb. und erweiterte). Springer Spektrum.

BMF. (2024). *Die wichtigsten Steuern im internationalen Vergleich 2023* (Monatsbericht des BMF). Bundesministerium der Finanzen.

Brandimarte, P. (2014). *Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics* (1.). Wiley.

Brandis, P., & Heuermann, B. (Hrsg.). (2024). *Ertragsteuerrecht* (174. EL). Franz Vahlen Verlag.

Büchter, A., & Henn, H.-W. (2007). *Elementare Stochastik: Eine Einführung in die Mathematik der Daten und des Zufalls* (2., überarb. und erweiterte). Springer-Verlag.

Bulgarische Nationale Steuerverwaltung. (2012). Erläuterung Nr. 24-96 v. 2.8.2012 bezüglich der steuerlichen Behandlung von Gewinnen, die durch eine Betriebsstätte gemäß dem Gesetz über die Körperschaftsteuer erzielt wurden.

Bulgarische Nationale Steuerverwaltung. (2014). Erläuterung Nr. 5304654 v. 16.10.2014 bezüglich der steuerlichen Behandlung von Dividenden-Einkünften des Eigentümers gemäß dem Umsatzsteuergesetz.

Bulgarische Nationale Steuerverwaltung. (2021a). Erläuterung Nr. 26-B-362-2 v. 10.6.2021 bezüglich der Anwendung des Doppelbesteuerungsabkommens zwischen Bulgarien und den USA.

Bulgarische Nationale Steuerverwaltung. (2021b). Erläuterung Nr. M-94-D-50#1 v. 2.2.2021: Besteuerung von Dividendeneinkünften, ausgeschüttet von der X GmbH zugunsten einer ausländischen juristischen bzw. natürlichen Person mit steuerlichem Wohnsitz im Vereinigten Königreich Großbritannien und Nordirland.

Bulgarische Nationale Steuerverwaltung. (2023). Erläuterung Nr. 07-00-502 v. 6.10.2023: Betreffend die Bestimmungen des Gesetzes über die Einkommensteuer natürlicher Personen und des Körperschaftsteuergesetzes.

Deloitte. (2024). CFO Survey Herbst 2024 Fokus Mittelstand. Verfügbar 17. Mai 2025 unter https://www.deloitte.com/de/de/services/deloitte-private/perspectives/aktuelle-wirtschaftliche-lage-im-deutschen-mittelstand.html

Detscher, S. (2012). *Direktinvestitionen in Mittel- und Osteuropa: Determinanten und Konsequenzen für den Transformationsprozess*. AV Akademikerverlag.

Deutsche Finanzagentur. (2025). Bundesanleihen. Verfügbar 18. Mai 2025 unter https://www.deutschefinanzagentur.de/bundeswertpapiere/bundeswertpapierarten/bundesanleihen

Dimitrova, H. (2024). Unternehmensbesteuerung in Bulgarien – Steuersubjekte und Kernpunkte der Besteuerung. *NWB IWB*, 376–383.

Elschner, C., Jajesniak-Quast, D., & Kudert, S. (Hrsg.). (2015). *Unternehmensbesteuerung in Ostmitteleuropa – Bestandsaufnahme, Entwicklungen und praktische Fragen nach 10 Jahren EU-Osterweiterung*. epubli.

Estrin, S., Richet, X., & Brada, J. C. (2000). *Foreign Direct Investment in Central Eastern Europe: Case Studies of Firms in Transition*. M.E. Sharpe.

Europäische Zentralbank. (2021). EZB-Rat verabschiedet neue geldpolitische Strategie. Verfügbar 20. Mai 2025 unter https://www.ecb.europa.eu/press/pr/date/2021/html/ecb.pr210708~dc78cc4b0d.de.html

Eurostat. (2025). Labour Cost Levels by NACE Rev. 2 Activity. Verfügbar 20. Mai 2025 unter https://ec.europa.eu/eurostat/databrowser/bookmark/34d5984c-272a-42d6-ac3d-39712fab9369?lang=en

Gleißner, W. (2020). Monte-Carlo-Simulation.

Gosch, D., Kroppen, H.-K., Grotherr, S., & Kraft, G. (Hrsg.). (2024). *DBA-Kommentar* (60. EL). NWB Verlag.

Haase, F. (2023). *Internationales und Europäisches Steuerrecht* (7., neu bearbeitete). C.F. Müller.

Haase, F., Nürnberg, P., & Bartsch, G. (2025). *Besteuerung von Unternehmen III* (4.). C. F. Müller.

Institut für Marktforschung Bonn. (2016). KMU-Definition des IfM Bonn. Verfügbar 20. Mai 2025 unter https://www.ifm-bonn.org/definitionen-/kmu-definition-des-ifm-bonn

Jacob, F., & Klein, M. (2014). S-Corporation die Zweite – Kernaussagen und Folgewirkungen des BFH-Urteils vom 26.6.2013, I R 48/12. *Internationales Steuerrecht (IStR)*, 121–130.

Jacobs, O. H. (Hrsg.). (2023). *Internationale Unternehmensbesteuerung: Deutsche Investitionen im Ausland. Ausländische Investitionen im Inland* (9., neu bearbeitete und erweiterte). C. H. Beck.

Kaminski, B. (1996). *Steuerliche Gestaltungsmöglichkeiten und deren Besteuerung bei der Verlagerung eines inländischen unternehmerischen Engagements in das Ausland*. Nomos Verlagsgesellschaft.

Kanzler, H.-J., Kraft, G., Bäuml, S. O., Marx, F. J., Hechtner, F., & Geserich, S. (2025). *Einkommensteuergesetz Kommentar Online* (10.) [Stand: 1.1.2025]. NWB Verlag.

Kaufmann, F., & Menke, A. (1997). *Standortverlagerungen mittelständischer Unternehmen nach Mittel- und Osteuropa – Eine empirische Untersuchung am Beispiel der vier Visegrád-Staaten*. Schäffer-Poeschel.

Kudert, S. (Hrsg.). (2007). *Investieren in Polen – Steuerliche und rechtliche Rahmenbedingungen für deutsche Unternehmen – mit Gestaltungsempfehlungen* (3., völlig neu bearbeitete und erweiterte). Erich Schmidt Verlag.

Kudert, S., Jarzynska, P., & Jamrozy, M. (2012). *Hybride Rechtsformen als Gestaltungsinstrumente für Outboundinvestitionen mittelständischer Investoren* (Discussion Paper Nr. 326). European University Viadrina Frankfurt (Oder), Department of Business Administration und Economics.

Looks, C., & Heinsen, O. (2017). *Betriebsstättenbesteuerung – Inboundinvestitionen, Outboundinvestitionen, Steuergestaltungen, Branchenbesonderheiten* (3.). C.H. Beck.

Makhavikova, H. (2018). *Determinants of FDI in Central and Eastern Europe: The Effects of Integration into the European Union*. Springer International Publishing.

Marinova, S. T., & Marinov, M. A. (Hrsg.). (2003). *Foreign Direct Investment in Central and Eastern Europe*. Ashgate.

Oberfinanzdirektion (OFD) Nordrhein-Westfalen. (2020). Verfügung v. 9.4.2020, S 1301 – 1989/5000 – St 125. *Internationales Steuerrecht (IStR)*, (11), 435.

OECD. (2019). *Model Tax Convention on Income and on Capital 2017 (Full Version)*. https://doi.org/10.1787/g2g972ee-en

Petkov, G. (2015). Besteuerung von Dividenden, die von inländischen bulgarischen juristischen Personen an ausländische Personen ausgeschüttet / ausgezahlt werden.

Pfaffermayr, M. (1996). *Direktinvestitionen im Ausland: Die Determinanten der Direktivestitionen im Ausland und ihre Wirkung für den Außenhandel*. Physica Verlag.

Querbach, T. (2024). Maßgeblichkeitsgrundsatz.

Rademacher-Gottwald, C. (2016). Internationale Steuerplanung: Betriebsstätte oder Tochtergesellschaft?

Rupp, T. (2025). Internationales Steuerrecht: Grenzüberschreitende Beteiligung an Personengesellschaften – Qualifikation und Abkommensberechtigung.

Sasse, A., & Mink, S. (2020). Unternehmensplanung und -steuerung in mittelständischen Unternehmen.

Scheffler, W. (2020). *Besteuerung von Unternehmen I: Ertrags-, Substanz- und Verkehrsteuern* (14.). C. F. Müller.

Schulz-Trieglaff, K. (2022). Klare Antwort des Europäischen Gerichtshofes auf die Frage nach den finalen Verlusten. *Steuer- und Bilanzpraxis*, 865–867.

Statistisches Bundesamt. (2025). Durchschnittlicher Gewerbesteuerhebesatz 2023. Verfügbar 20. Mai 2025 unter https://www.destatis.de/DE/Themen/Staat/Steuern/_Grafik/_Interaktiv/steuereinnahmen-hebesaetze-gewerbesteuer-laender.html

Tipke, K., & Lang, J. (2024). *Steuerrecht* (25., völlig überarb.). Verlag Otto Schmidt.

Vogel, K., & Lehner, M. (Hrsg.). (2021). *Doppelbesteuerungsabkommen: Kommentar auf der Grundlage der Musterabkommen* (7., völlig neu bearbeitete). C. H. Beck.

Wassermeyer, F. (Hrsg.). (2025). *Doppelbesteuerung: Kommentar zu allen deutschen Doppelbesteuetungsabkommen* (168. EL., 99.). C. H. Beck.

Wassermeyer, F., Andresen, U., & Ditz, W. (Hrsg.). (2018). *Betriebsstättenhandbuch* (2., völlig überarbeitete). Otto Schmidt.

Wassermeyer, F., Richter, S., & Schnittker, H. (Hrsg.). (2015). *Personengesellschaften im Internationalen Steuerrecht* (2.). Otto Schmidt.

Weiss, M. (2021). Vermeidung der Doppelbesteuerung durch die Anrechnungsmethode – Teil 2. *Praxis Internationale Steuerberatung (PIStB)*, 98–104.

Junior Management Science

www.jums.academy
ISSN: 2942-1861

CEO Narcissism and Stock Market Reactions: An Empirical Analysis

Paul Lehmann

Technische Universität Dresden

Abstract

This bachelor thesis examines how stock markets react to CEO turnover events involving narcissistic leaders. CEO signature size is utilized as an unobtrusive, archival proxy for narcissism, and data from 341 NASDAQ-listed firms over the period 2004–2024 is analyzed. An event study methodology is applied, with cumulative abnormal returns (CAR) calculated over a (-2, +2) day window surrounding CEO appointment announcements. Regression analyses with robust standard errors are conducted to assess the relationship between CEO narcissism and stock market reactions. The results indicate that no significant relationship exists between CEO narcissism and stock price reactions at the time of appointment. However, CEO duality—where the newly appointed CEO also serves as board chair, signifying concentrated power—is found to be a significant predictor of negative market reactions. These findings suggest that investors prioritize governance structures over individual personality traits when assessing CEO appointments. It is concluded that corporate governance considerations play a more decisive role than CEO characteristics in shaping shareholder sentiment during turnover events. Since shareholders, through general meetings or proxy statements, vote on the appointment and removal of directors who select the CEO, the thesis researches whether investor preferences influence the demand for narcissistic leadership by analyzing stock market reactions to CEO appointments.

Keywords: ceo narcissism; ceo turnover; shareholder value; upper echelons theory

> *"If I am a narcissist (which might be true), at least I am a useful one."*
>
> (Musk, 2018)

1 Introduction

Upper Echelons Theory, proposed by Hambrick and Mason (1984, pp. 193-195), posits that the attributes of an organization's top management team (TMT) play a crucial role in shaping its strategic decisions and organizational outcomes. Given the vast amount of information accessible within an[1] organizational framework, the theory contends that the personal histories and experiences of top executives significantly influence their cognitive biases and value systems. These factors subsequently impact how they interpret information and engage in decision-making processes. As a result, the organization often reflects the characteristics of its top managers. A particular personality trait that has garnered considerable academic attention in recent years is narcissism (Cragun et al., 2019, p. 909).

Narcissism is defined by traits including an inflated sense of self-importance, a sense of entitlement, a lack of empathy, and a persistent craving for admiration and validation from others (American Psychiatric Association, 2013). While both the concept and real-world manifestations of narcissism had captured the attention of scientific scholars, media and the public cultural domain (Miller and Campbell, 2010, p. 180; Silove et al., 2016, p. 115; Miller et al., 2021, p. 519), it wasn't until Chatterjee and Hambrick (2007, pp. 351-353) established an unobtrusive, archival measure for narcissism levels in CEOs that a growing body of leadership and management research has shown a wide array of consequences of CEO narcissism on corporate strategy decisions and organizational outcomes (Cragun et al., 2019, p. 909). For example, narcissistic CEOs are more likely to intensify mergers and acquisitions (M&A) activities (e.g. Chatterjee and Hambrick, 2007, p. 351), as well as command higher external audit fees (Judd et al., 2017, p. 34) and influence corporate social responsibility (CSR) measures (Petrenko et al., 2015, p. 263).

While there is a growing body of literature examining the effects of narcissism on various corporate strategies and outcomes (see Fig. 1), scholarly attention has been relatively sparse regarding how shareholders react to the appointment of a narcissistic CEO. Previous research has explored the implications of related personality traits and biases among CEOs, demonstrating that executive greed can adversely impact shareholder wealth (Haynes et al., 2016, pp. 36-38) and that overconfident CEOs may trigger negative market responses during merger announcements

[1] I am deeply grateful to Dr. Philipp Richter for his dedicated advisory role and invaluable insights. His precise feedback played a central role in shaping this thesis.

DOI: https://doi.org/10.5282/jums/v11i3pp513-531

© The Author(s) 2026. Published by Junior Management Science.
This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.

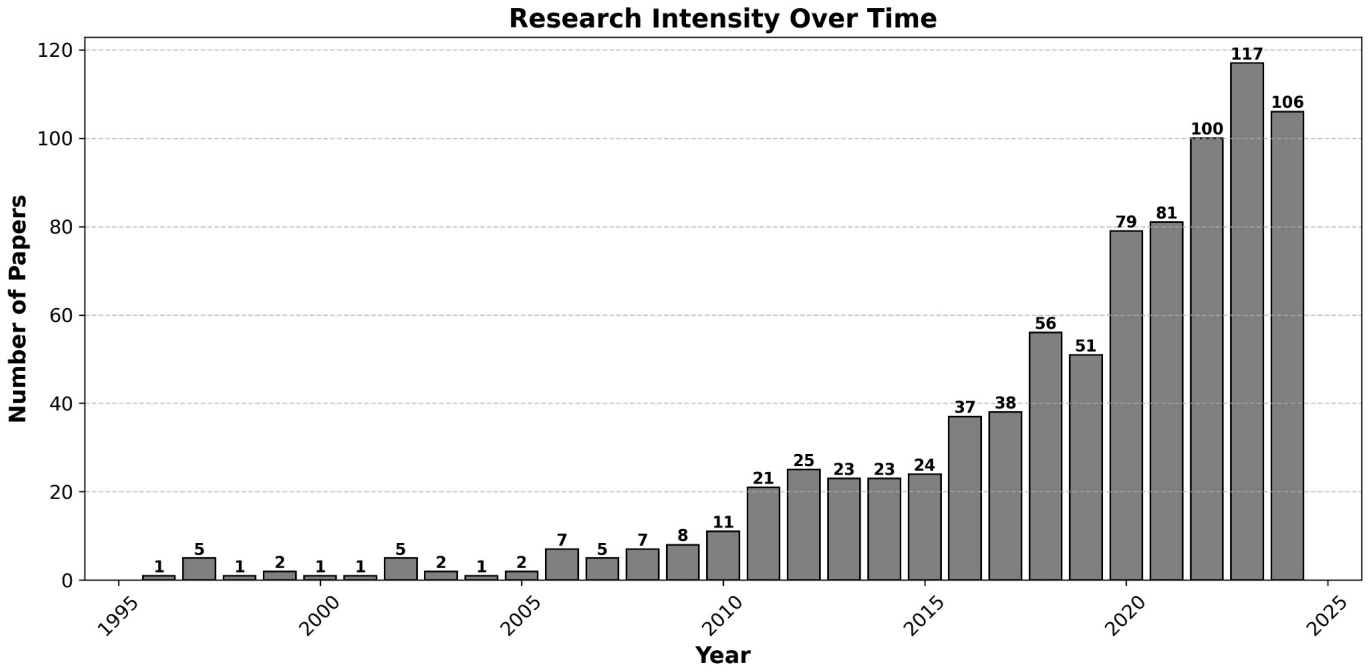


Figure 1: CEO Narcissism research intensity (own illustration)

(Malmendier & Tate, 2008, p. 21). Studies focusing on stock market reactions to CEO narcissism have largely centered on corporate events involving shareholders, such as earnings (Marquez-Illescas et al., 2018) and M&A announcements (Aktas et al., 2016).

Nonetheless, there persists a notable gap in literature regarding the effects of narcissism in the context of CEO turnover events and the associated stock market responses. This gap is particularly noteworthy for two reasons: first, narcissistic CEOs frequently make significant strategic decisions, for instance those related to internationalization efforts (Oesterle et al., 2016, p. 1), which can result in increased volatility in shareholder value (Jensen & Meckling, 1976); second, there is a growing consensus that the financial and legal risks linked to narcissistic behavior in CEOs present considerable drawbacks to this personality trait among top executives (O'Reilly et al., 2018, p. 367; Reina et al., 2014, p. 968).

This Bachelor thesis adopts signature size as a proxy for measuring narcissism (Ham et al., 2017, pp. 3-8) and analyzes financial data and historic stock prices from 341 NASDAQ-listed companies in the United States to investigate the relationships between CEO turnover events and the corresponding stock market reactions in the context of CEO narcissism using linear regression models.

The results of this empirical investigation contribute to the fields of management and leadership by evaluating capital market expectations regarding narcissism as a personality trait in CEOs at the time of their appointment, prior to the manifestation of subsequent effects resulting from narcissistic behaviors and decisions within the organization. This empirical framework offers valuable insights into the prior expectations and assessments of narcissistic CEOs by shareholders. In accordance with the suggestions of Cragun et al. (2019, p. 920), this Bachelor thesis aims to enhance the demand-side perspective on CEO narcissism by examining whether capital markets, which reflect investor sentiment, favor narcissistic CEOs.

This analysis has the potential to contribute a small part towards the reconciliation of findings in existing literature that underscore negative firm outcomes with the limited, yet strategically important advantages associated with narcissistic leadership (Rosenthal & Pittinsky, 2006, p. 617).

2 Theory and Hypothesis Development

This chapter explores the extant literature regarding the origins that precede narcissism research in contemporary management and leadership contexts. It lays out the definitions of narcissism that have been applied in research, as well as the definition of narcissism adopted for the empirical study in this Bachelor thesis. Relevant theoretical constructs that the research is embedded in are explored and detailed. Extant findings regarding the organizational outcomes of narcissistic CEO leadership and their downstream effects on firm performance are reviewed and evaluated. The chapter concludes with the offering of a hypothesis regarding the results of the empirical analysis. The hypothesis is grounded in the findings of extant research and literature.

2.1 Conceptual Origin and Definition

The tale of Narcissus, rooted in Greek mythology, was initially documented in Ovid's (2004) Metamorphoses: Narcis-

sus, a remarkably handsome youth, was foretold by the seer Tiresias to have a long life, as long as he never recognized his own reflection. He rejected all romantic advances, including those from Echo, a nymph condemned to only echo the words of others, showcasing profound levels of self-love and emotional detachment. As a result of divine retribution, Nemesis guided Narcissus to a pool of water, where he became entranced by his own reflection, unable to look away. This unattainable self-love ultimately led to his downfall, resulting in his death by either starvation or drowning, and his transformation into the narcissus flower. The myth serves as a cautionary tale about the dangers of excessive self-absorption outweighing the love of others and was an influence on the contemporary conceptualization of narcissism beginning at the turn of the 20th century (Grenyer, 2013).

Ellis (1898) was the first scholar to introduce the concept of narcissism into psychological literature using the Narcissus tale, linking it to autoeroticism, meaning a sexualized fixation on one's own mind and body, describing narcissism as a psychological condition where a person becomes increasingly preoccupied with their own body and image (Gerstner et al., 2013; Yakeley, 2018).

Freud (1914, 1957) broadened this concept, making it central to psychoanalytic theory and personality development. He distinguished between narcissism as a normal stage in childhood development and a secondary form which manifests later in life and is defined by the withdrawal of emotional attachment to others and redirecting it toward the self. In contrast to the primary form, which was seen as a normal stage in infancy, the secondary form appearing in adults was labeled as a pathological condition that impedes relationships and, as Freud suggested, could lead to mental disorders such as paranoia or depression.

It was not until the 1970s that the theory of narcissism was further studied and refined (Judge et al., 2006, p. 762), with scholars criticizing Freud's conception of narcissism, claiming a lack of distinguishability between healthy levels of self-esteem and pathological grandiosity as well as between narcissism and other mental disorders. Crucial in its refinement are the works of Kernberg (1975) and Spitzer (1980), which highlight the limitations of Freud's conceptualization and would make significant contributions to the formalized adoption of narcissism into the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III), authored by the American Psychiatric Association (APA, 1980), by making the disorder clinically distinct and empirically measurable. The definition was last updated in the DSM-5 in 2013 and entails a guideline to clinically assess narcissism as a binary (categorical) condition that is either present in an individual or not, highlighting distinct narcissistic tendencies: a "grandiose sense of self-importance", the belief that one is special and more important than others; "fantasies of unlimited success, power, [or] brilliance"; poor receptions of criticism; a "sense of entitlement"; the tendency to exploit others; a "lack of empathy"; and a constant and continuous need for "excessive admiration" (APA, 2013, pp. 669-671). This formal adoption provided the first standardized framework for the diagnosis of pathological narcissism by clinicians. The diagnosis was considered part of the so-called Cluster B disorders, which contain related diagnoses such as antisocial personality disorder (ASPD) or borderline personality disorder (BPD; APA, 1980).

In contrast, recognizing the potential for narcissistic traits to manifest in subclinical forms within the broader population, researchers sought to develop instruments that could measure these characteristics on a continuous spectrum instead of the binary classification used for clinical patients (Chatterjee & Hambrick, 2007). In line with those efforts, an alternative measure was conceived that would measure the severities of the different symptoms in the clinical definition of narcissism and make the scale applicable to non-pathological individuals: the Narcissistic Personality Inventory (NPI), first proposed by R. N. Raskin and Hall (1979), is a psychometric tool relying on self-reported questionnaire answers to gauge narcissism levels in otherwise healthy individuals.

While the original NPI contained up to 220 items (i.e. 220 questions), the index was rather quickly reduced and validated to an inventory of 40 items (R. Raskin & Terry, 1988) in an effort to improve usability by shortening the time needed for participants to answer all questions and avoid fatigue in respondents. In the same vein, the inventory of the NPI was later reduced to only 16 items and validated against earlier models (Ames et al., 2006).

Derived from the clinical definition, the NPI was designed to measure narcissistic traits on a continuum, reflecting the degree to which individuals embody these characteristics, rather than serving as a diagnostic tool for narcissistic personality disorder (NPD; Junge et al., 2024, p. 4).

An exploratory factor analysis (EFA) has determined four underlying structures of narcissism: "Leadership/Authority", meaning the confidence in one's decisions and superiority as well as a strong desire to be in a position of leadership; "Superiority/Arrogance", meaning a belief in one's naturally exceptional capabilities combined with looking down on others; "Self-Absorption/Self-Admiration", meaning a preoccupation with one's own abilities or bodily attractiveness; and "Exploitativeness/Entitlement", describing the willingness to manipulate others and the feeling of deserving special treatment (Emmons, 1987, p. 11). In addition to this four-factor model, R. Raskin and Terry (1988, p. 890) used principal components analysis (PCA) to identify three additional dimensions of narcissism, bringing the total to seven: "Self-Sufficiency", meaning the belief in one's independence and self-reliance; "Exhibitionism", meaning the urge to constantly be the recipient of praise and attention, putting one in the 'spotlight'; and "Vanity", describing the over-the-top, excessive concern with one's own personal appearance.

Management research literature has adopted these core traits as the basis for defining narcissism as a subclinical characteristic that is at the basis of a leader's actions and behaviors, specifically in the context of organizations (Cragun et al., 2019, p. 909).

Therefore, in this Bachelor thesis, a definition of narcissism is chosen that closely resembles the characteristics outlined in the DSM and NPI, positing that narcissism is a continuum of traits rather than a disorder, which expresses itself through a multitude of overt motives: 1) *Grandiosity* – a highly inflated sense of self-importance and an exaggerated belief in one's superiority over others, talents, and achievements; 2) *Admiration* – a constant desire for attention, validation, and praise from others in order to maintain one's self-esteem and social dominance, with narcissists enjoying the spotlight and being celebrated; 3) *Entitlement* – the belief that one is inherently deserving of special treatment, privileges, or rewards - regardless of merit - and leading to frustration and interpersonal conflicts when those expectations are not met; 4) *Empathy* – the difficulty in recognizing, understanding or caring about others' emotions, perspectives, and needs, resulting in interpersonal relationships that are founded in transactionality rather than a mutual understanding of each other; 5) *Manipulation* – the weaponization of a lack of empathy causing behavior that manipulates others for personal gain This conceptualization captures a narcissist's motivations and cognitive base.

2.2 Manifestation of Behavior

While the definition of narcissism provides insights into its core traits, it does not explain how these traits translate into real-world behaviors. Therefore, a theoretical framework must be applied to fully understand how narcissists regulate their behavior atop their cognitive base and maintain the positive self-image that is innate to them.

With that challenge in mind, R. Raskin et al. (1991) proposed that self-esteem regulation is central to narcissists and developed a *self-regulation model* of narcissism, arguing that narcissistic individuals are driven by their desire for grandiosity and use that desire as a primary strategy for bolstering and maintaining their self-esteem. This puts grandiosity as not just a personality trait but as the basis of an enduring strategy for self-enhancement that narcissists use to maintain their positive view of themselves. The authors further argue that this strategy for self-regulation is independent from social validation – meaning that instead of trying to fit in or gain approval from others, narcissists will simply create their own sense of superiority that they project onto others. This results in behaviors like bragging, seeking admiration for accomplishments, or downplaying their failures in order to avoid self-doubt and insecurity (R. Raskin et al., 1991, pp. 19-22). While the conceptualization of narcissism as a self-esteem strategy was a pivotal cornerstone in the development of a regulatory framework, the *self-regulation model* does not address how a narcissist sustains their grandiosity through interpersonal relationships – instead, the authors posit that narcissists do not need others for validation and that their self-esteem is constant regardless of situational context.

Morf and Rhodewalt (2001) build upon and significantly expand this model with the inception of the *dynamic self-regulatory processing model*. While they share the core premise that narcissism is a self-regulatory system aiming to enhance their inflated sense of self-esteem, they reject the notion that narcissists uphold their sense of self-esteem independently from others, introducing an interpersonal dimension into the self-regulatory system. This means that narcissism is considered an ongoing, dynamic process in which narcissists do not simply possess grandiosity; rather, they must constantly work to sustain it through social feedback. Another aspect of this interpersonal dimension is a narcissist's response to failures or threats: Morf and Rhodewalt suggest that narcissists are highly reactive to feedback, criticism, rejection or threats and will handle it by responding with aggression, shifting blame or retaliating against those who they perceive to have wronged them. Paradoxically, their constant need for admiration leads to self-defeating manifestations of behavior (e.g. exploitation, entitlement), which pushes people away and ultimately undermines their efforts to regulate their self-esteem (Morf & Rhodewalt, 2001, pp. 177-180). This model demonstrates how narcissistic leaders manipulate their social environment to gain external validation and sustain their self-image.

In a third iteration of this framework's theoretical progression, Campbell and Foster (2007) propose a systemic and dynamic view aimed at explaining how narcissists sustain and regulate their grandiosity by creating the *extended agency model of narcissism*, consisting of three components that interact with and reinforce each other: *"Interpersonal Skills"* entail the social abilities that narcissists leverage to attract admiration and maintain their grandiosity, such as charisma, confidence and assertiveness; *"Intrapsychic Strategies"* refers to a narcissist's internal cognitive mechanisms that drive them to sustain their self-esteem such as grandiose fantasies or selective memory recall, with these strategies allowing them to interpret reality in a way that reinforces their feeling of superiority by exaggerating own successes and downplaying or externalizing failures and rejecting criticisms, thereby ensuring that they stay shielded from self-doubt or insecurity; *"Interpersonal Strategies"* which are behaviors employed by narcissists to manipulate social interactions by outwardly promoting themselves and their successes or grandiose visions, ultimately serving to extract validation (Campbell & Foster, 2007, p. 123).

The systemic aspect of this framework stems from the postulation that these three components are interdependent, meaning that they get activated and deactivated together and reinforce each other. For instance, interpersonal skills such as charm or confidence help narcissists attain admiration and leadership positions. The subsequent social validation reinforces their grandiose self-image and strengthens their fantasies of superiority, which in turn increases their sense of entitlement and dominance, leading to more aggressive self-promotion and manipulation; this in turn secures further admiration, thereby restarting the cycle. These interconnected elements can also activate and deactivate together if they pass or fall below certain thresholds. For instance, in a situational context that offers opportunities for admiration, validation and recognition, the narcissistic system is activated. The same is true for competitive environments, which leads

the system to fully activate all three components in order to escalate their self-promotion and gain dominance. In contrast, if narcissists lose status or social validation, they may temporarily deactivate this system and decrease their sense of grandiosity until new sources of validation emerge (Campbell & Foster, 2007, pp. 122, 124).

Further, the *extended agency model of narcissism* framework introduces "narcissistic esteem" (Campbell & Foster, 2007, p. 121) as a concept that characterizes a narcissist's self-esteem differently from traditional self-esteem in non-narcissists. Narcissistic esteem is primarily tied to dominance, pride, and addictive qualities. This *narcissistic esteem* is the final result of the interconnected components working as intended, making the narcissist feel good. In a yet again cyclical nature, higher *narcissistic esteem* leads the system to perform better and prevent a shutdown. Lastly, the authors propose that there is no one overarching goal of narcissism and that it is simply a dynamic system that is self-reinforcing and can be activated and deactivated depending on external, situational factors (Campbell & Foster, 2007, p. 122).

Using this framework helps to understand how the core values, traits and needs that are innate to narcissists translate into real-world behaviors, shaping the outcomes of CEO narcissism in the organizations they lead and providing the necessary foundation to analyze and interpret the corresponding findings in the extant literature that are laid out and discussed in the following chapters as well as the empirical results of the study that follows.

2.3 CEO Narcissism Outcomes

The following chapter presents findings in the management and leadership literature regarding the outcomes of narcissistic CEO behavior on the firms they lead. It is important to note that contemporary research in leadership and management in the context of CEO narcissism has been significantly influenced by the seminal study by Chatterjee and Hambrick (2007), which has been acknowledged as such by other scholars (e.g. Buyl et al., 2017, p. 1383; Li et al., 2022, p. 3) due to their contribution of an archival, unobtrusive measurement for narcissism in CEOs (Chatterjee & Hambrick, 2007, p. 352), based on five observables that reflect narcissistic tendencies: 1) The size and placement of the CEO's *photograph* in the company's annual reports, reflecting the desire to be showcased; 2) The *frequency* with which the CEO's *name* is mentioned in company press releases, indicating desires of visibility; 3) The ratio between *first-person singular pronoun* (FSP) and plural pronoun usage by CEOs during interviews to gauge self-centeredness; 4) The CEO's *salary* compared to those of the second-highest-paid executive in the firm as a reflection of entitlement and superiority; 5) The CEO's non-cash compensation packages compared to those of the second-highest-paid executive (Chatterjee & Hambrick, 2007, pp. 363-365). Figure 1 captures this research trend beginning around 2007.

Because of the wide variety of firm outcomes that have been explored in the extant literature, the respective primary domains in which the CEO behavior manifests have been attempted to be categorized. However, while categorically distinct, they are not independent from each other. More specifically, strategic, ethical, and cultural outcomes interact with each other and have corresponding downstream effects on firm performance, as shown in Figure 2.

2.3.1 Strategic Outcomes

Narcissistic CEOs are more likely to initiate takeovers and lead negotiations faster than non-narcissistic CEOs (Aktas et al., 2016, p. 113). When firms led by narcissistic CEOs are the target of a takeover, they receive higher bid premiums (Aktas et al., 2016, p. 119). Similarly, Chatterjee and Hambrick (2011) showed that narcissist CEOs overpay for acquisitions (p. 202). To add, Zhu and Chen (2014b) found that narcissistic CEOs prefer appointing narcissists to the board in order to have their own proclivity towards risk-taking behavior reinforced (p. 2078).

Similarly, Narcissistic CEOs are more likely to both acquire and overpay for well-known, high-awareness brands to gain greater public visibility (Cao & Xu, 2022, p. 777). Similarly, Chatterjee and Hambrick (2007) found that narcissistic CEOs engage in more and larger acquisitions (p. 351), which was confirmed by Ham et al. (2018), who added that this relationship is more pronounced the higher the financial slack (i.e. high cash reserves and low debt) of the company (p. 235, p. 259). However, findings by Aabo et al. (2023) contradict this, showing no relationship between CEO narcissism and intensified corporate acquisition activities (p. 94). This might be due to sample differences: Aabo et al. (2023) relied solely on FSP as measurement for narcissism (p. 91), potentially not fully capturing CEO narcissism.

Narcissistic CEOs are more likely to announce repurchasing of shares in greater amounts but less likely to follow through. Weak governance structures exacerbate these tendencies (Boamah & Banerjee, 2024, p. 717).

In financial disclosures, narcissist CEOs use a more optimistic tone than what is justified by the firm fundamentals. They are then also more likely to invest heavily in research and development (R&D; Buchholz et al., 2018, p. 531).

In an empirical analysis of 92 U.S. commercial banks between 2006 and 2014, Buyl et al. (2017) found that CEO narcissism leads to riskier bank policies, with stock-based incentives exacerbating this effect (p. 1395). Similarly, narcissistic CEOs engage in significantly more risky strategic investments such as (R&D) and acquisitions - for which they overpay - and financial success or failure does not alter their risk-taking profile: conversely, positive media coverage emboldens them in their risk-seeking tendencies (Chatterjee & Hambrick, 2011, p. 202).

CEO narcissism predicts strategic dynamism, in which resources are frequently reallocated, corporate strategy experiences major shifts, and focus is set on high-risk decisions that promise a high reward instead of incremental changes (Chatterjee & Hambrick, 2007, p. 372). Kashmiri et al. (2017) provide similar evidence for radical product innovations and add

a link between CEO narcissism and the likelihood of product-harm crises, caused by insufficient long-term product safety investments (p. 633). Similarly, in a survey of 132 Chinese firms, You et al. (2022) found CEO narcissism to negatively influence innovation ambidexterity, i.e. a balanced approach to innovation strategy, focusing on both new and existing domains, with narcissistic CEOs showing a proclivity to focus on either of the two, leading to either over-innovation or stagnation (p. 189).

Research conducted by Gerstner et al. (2013) examined pharmaceutical companies' reactions to the emergence of biotechnology, finding that CEO narcissism predicted a more aggressive adoption strategy of disruptive technologies while disregarding uncertainty behind them, with audience engagement (e.g. media and industry) strengthening this link (p. 257).

CEO narcissists are more likely to expand firms internationally in order to enhance their status rather than out of economic considerations (Oesterle et al., 2016, p. 1114).

Curiously, the combination of narcissism and humility was shown to significantly improve product innovation in a sample of Chinese firms by balancing self-confidence with a willingness to learn, encouraging risk-taking while maintaining team cohesion, and displaying their charisma to inspire and unite their employees behind a shared vision (Zhang et al., 2017, p. 585, pp. 593-599). Crucially, the authors note that this balance between the two examined traits is potentially more prevalent in cultures more closely linked to Confucian principles, indicating that this combination of traits is less prevalent in the Western world context (Zhang et al., 2017, pp. 590-591).

In summary, extant research literature shows that narcissistic CEOs tend to be more aggressive in strategic decisions, quickly pursuing acquisitions and often overpaying for trendy brands to boost visibility. They attract high takeover offers, engage in risky investments like R&D, and frequently change corporate strategies. Their inclination to hire similar board members promotes risk-taking, and inadequate governance supports this behavior. They also issue overly optimistic earnings reports, pursue internationalization for status rather than economic reasons, and gain confidence from positive news. Their risk-taking continues regardless of economic outcomes.

2.3.2 Ethical Outcomes

Narcissistic CEOs engage in earnings management outside the Generally Accepted Accounting Principles (GAAP; Capalbo et al., 2017, p. 210). An analysis of 923 CEOs at 716 firms found narcissism to be positively linked to the frequency and size of expense exclusions from non-GAAP earnings, and negatively linked to reporting quality (i.e. excluding recurring expenses; Abdel-Meguid et al., 2020, p. 3). These findings are confirmed by Buchholz et al. (2019), who found narcissistic CEOs to engage in Accrual-Based Earnings Management (ABEM; i.e. delaying of expenses or early recognition of revenue; p. 663). However, they also found evidence for income-decreasing earnings management, indicating that narcissistic CEOs are not only driven by short-term success motivations; rather, they will at times strategically overstate expenses or understate revenues to create a reserve of hidden earnings that can be used in the future to smooth performance (Buchholz et al., 2019, pp. 667-668). Further, research has linked CEO narcissism to tax avoidance and tax sheltering (García-Meca et al., 2021, p. 223; Olsen and Stekelberg, 2016, p. 1). In support of these results, narcissism in CEOs has been found to increase audit fees due to internal control weaknesses, thus highlighting the importance of audit oversight (Judd et al., 2017, p. 33). In experimental settings, Majors (2015) found that range disclosures reduce aggressive financial reporting in narcissistic managers, suggesting that regulatory oversight can mitigate opportunistic reporting (pp. 973, 989). Findings by Mutschmann et al. (2022) conclude that narcissistic CEOs not only engage in earnings management, but outright fraud, extending the findings of Capalbo et al. (2017) and Buchholz et al. (2019). However, in contrast to the conclusions of García-Meca et al. (2021), they showed that mitigating these tendencies necessitates outsourced audit teams that are outside of the narcissistic CEO's sphere of influence and power (Mutschmann et al., 2022, p. 763). Findings by O'Reilly et al. (2018) support the notion of increased fraud as a result of CEO narcissism, concluding that a narcissistic CEO will be more likely to involve his firm in lawsuits and less likely to settle said lawsuits as their grandiosity and arrogance prohibits them from admitting defeat (p. 365). Findings by Rijsenbilt and Commandeur (2012) confirm ties between CEO narcissism and corporate fraud, most frequently in the form of earnings manipulation, motivated by overconfidence to get away with it (pp. 422-423).

Narcissistic CEOs use environmental, social, and governance (ESG) disclosures, particularly social and governance reporting (e.g. workplace diversity, board transparency), to generate media attention and improve their image (Dabbebi et al., 2022, pp. 3147-3148). Similarly, B. Kim et al. (2018) found that narcissistic CEOs engage in CSR measures even if their firms are based in countries with high levels of uncertainty avoidance (UA; i.e. a cultural preference for risk averse, long-term focus), suggesting that narcissistic CEOs use CSR initiatives as a personal branding tool regardless of cultural norms that would generally discourage their implementation (p. 203). This assumption is supported by Petrenko et al. (2015), who found that CSR initiatives implemented by narcissistic CEOs do not improve firm performance and suggested that it is rather a tool they use to enhance their personal image (p. 262). The effect of CEO narcissism on CSR initiatives is distinct from CEO hubris (Tang et al., 2018, p. 1370).

To summarize, narcissistic CEOs have been found to engage in earnings manipulation, tax avoidance, corporate fraud, and excessive litigation, often driven by overconfidence and a need for self-enhancement. They inflate earnings, exploit tax shelters, and manipulate financial reports, increasing audit fees and regulatory scrutiny. Their

resistance to legal settlements prolongs lawsuits, while their use of ESG and CSR initiatives serves as self-promotion rather than genuine corporate responsibility. These behaviors undermine ethical business practices and call attention to the need for stronger governance and external oversight to mitigate the ethical risks associated with narcissistic leadership.

2.3.3 Cultural Outcomes

CEO personality traits shape the culture of the organizations they lead, which subsequently impacts the company's performance outcomes (Berson et al., 2007, pp. 626-628). Indeed, in a study of 173 CEOs and corresponding firm performance data, Wales et al. (2013) identified organizational culture as a partially mediating factor between CEO narcissism and the variance in firm performance, suggesting that narcissistic CEOs drive their organizations toward high-risk strategic behavior with volatile long-term performance outcomes; more specifically, the authors find that narcissistic CEOs promote a strong entrepreneurial orientation (EO), characterized by a focus on innovation (p. 1041). Their extreme overconfidence leads them to endorse high-risk, high-reward corporate strategies, effectively forcing their own tendencies and biases onto the organization through firm culture (Wales et al., 2013, p. 1045). This is in line with findings from Kashmiri et al. (2017) which demonstrate that narcissistic CEOs favor radical innovation in product management activities and foster an aggressive organizational culture aimed at outpacing competitors (p. 633). The potential self-selection of narcissists to firms already orientated towards high-risk, high-reward strategies (Young et al., 2016, p. 49) and findings suggesting that organizations drive demand for narcissists in times of uncertainty (Nevicka et al., 2013, p. 370) raise the question of reverse causality.

In an empirical analysis of 641 CEOs from 399 U.S. firms in the time period between 2007-2017, a U-shaped relationship between CEO narcissism and TMT age was discovered, suggesting that CEOs lower in narcissism tend to appoint younger TMT members in an effort to exert dominance and position themselves as the most accomplished person in the room in order to protect their self-view, while CEOs on the upper end of the narcissism spectrum tend to appoint older TMT members, who they perceive to be high in status due to their age, in order to extract admiration and validation from them (see also Chatterjee and Pollock, 2017, p. 703). TMT members appointed by narcissistic CEOs were also found to have shorter tenures, suggesting a dysfunctional relationship in which CEO narcissists may dispose of TMT members that defy them or in which TMT members leave their roles voluntarily as a result of difficulties in cooperating with the narcissistic CEO (Bredlich et al., 2021).

Additionally, in a sample of 145 Spanish publicly listed firms, CEO narcissism was found to negatively influence the amount of female TMT members. This effect was more pronounced the more power the CEO yields (i.e. longer tenure, CEO-chair duality). Curiously, female narcissistic CEOs exhibit an even stronger aversion to the appointment of female TMT members. The findings suggest that narcissistic CEOs prefer TMT compositions that reinforce their self-view, and that they might perceive women in leadership roles as a threat to their dominance (Martínez-Ferrero et al., 2023).

Junge et al. (2024), employing a novel, unobtrusive measure for executive narcissism by examining social media profiles on the platform LinkedIn, found CEO narcissism to be positively associated with TMT turnover, confirming the findings of Bredlich et al. (2021). Further, the authors found that this association is moderated by TMT narcissism, suggesting that narcissistic TMT members may clash with a narcissistic CEO more frequently, leading to even higher turnover (Junge et al., 2024, p. 25). Paradoxical on the surface, the authors also found that narcissistic CEOs prefer to hire narcissistic executives to their TMT in the first place (Junge et al., 2024, p. 21), suggesting that they prefer to hire narcissists to their leadership teams but are unable to work with them. The finding confirms prior research by Bachrach et al. (2023, p. 1).

In summary, CEO narcissism shapes organizational culture by developing high-risk, high-return approaches, with a priority for innovation, and for an assertive competitive stance. The TMT composition is impacted by narcissistic CEOs with the tendency to hire older members for admiration or young members to dominate, leading to higher turnover through poor working relationships. They are also resistant to female leaders, with female narcissistic CEOs being more averse. While narcissistic CEOs prefer the hiring of narcissistic executives, it becomes challenging for them to work with them, leading to TMT instability.

2.3.4 Firm Performance Outcomes

While firms led by narcissistic CEOs experience greater fluctuations in performance outcomes as measured by return on assets (ROA), the average firm performance remains comparable to that of companies led by non-narcissistic CEOs, the seminal study by Chatterjee and Hambrick (2007) found, analyzing 111 CEOs from 105 unique U.S. firms in the computer hardware and software industries (pp. 372-373).

Hou et al. (2024) found that CEO narcissism is associated with lower credit ratings, confirming their hypothesis derived from observed manipulative tendencies and propensity for risk-taking in narcissistic CEOs, leading to higher leverage and increased earnings manipulation activities in the firm, causing external agents such as credit rating agencies (CRA) to significantly lower their ratings in adjustment to narcissistic CEO behavior outcomes. This effect was moderated by several factors such as the financial constraints of the firm, which causes CEO narcissists to act more conservatively in the face of limited cash reserves or external funding, the degree of industry competitiveness, suggesting that narcissistic CEOs recognize market pressures and lower their affinity for excessive risk accordingly, and the degree of institutional ownership, suggesting that large investors (e.g. funds) - who monitor CEO activity more closely and rigorously - lead to restrained behavior in said activities (Hou et al., 2024, p. 39, pp. 57-58).

CEO narcissism is tied to increased debt financing costs, meaning companies with narcissistic CEOs have higher bond yield spreads, which is the difference between corporate bond yields and safer U.S. Treasury bonds. This is because narcissistic CEOs undertake riskier financial actions such as excessive leverage, risky takeovers, and earnings management, which makes the companies more likely to encounter financial issues since if narcissist CEOs are provided with more money, they tend to invest it into high-risk endeavors. Bondholders, being the principals in loan agreements, are aware of the self-serving actions and propensity for risk of narcissistic CEOs (agents) and thus demand higher yields in order to compensate for the increased agency costs. The research also finds that the effect persists when control mechanisms such as strong boards and institutional ownership are in place, which contradicts the findings by Hou et al. (2024) in regard to the interactions between CEO narcissism, internal control mechanisms, and external stakeholders/agents (J. H. J. Kim & Anderson, 2024, pp. 1-2, p. 7, p. 14).

Findings by Olsen et al. (2013) suggest that CEO narcissism is positively linked to higher earnings per share (EPS) and stock prices (p. 243). Narcissistic CEOs inflate reported EPS by overproducing goods to reduce cost per unit, offering excessive sales discounts and lenient credit terms in order to boost short-term revenue, and by cutting back on discretionary expenditures (e.g. R&D and advertising) to enhance short-term profitability (Olsen et al., 2013, p. 260-261). This suggests a short-term orientation in narcissistic CEOs to immediately attain attention and praise by increasing EPS and eliciting a positive market response in the form of higher stock prices, at the cost of potential future operational inefficiencies. Notably, these practices lead narcissistic CEOs to beat analysts' forecasts, further resulting in surging stock prices and fueling their *narcissistic esteem*, yet also highlighting that analysts may be ill-equipped to accurately predict CEO narcissism behavior: reliant on historical performance data to foresee short-term changes in earnings management activities, unable to directly observe the decision-making of the narcissistic CEO inside the company, and possibly being influenced by *CEO impression management* (i.e. self-promotion through optimistic language, grandiose claims, or media presence), analysts apparently struggle to anticipate the impact of narcissistic CEOs on earnings forecasts (Olsen et al., 2013, pp. 254-265).

In a study by Peterson et al. (2012), organizational identification, i.e. the degree to which a CEO emotionally identifies with the company, was found to mediate the relationship between CEOs and servant leadership, suggesting that CEOs who identify strongly with their organization are more likely to display servant leadership, which was found to positively affect ROA. Notably, narcissistic CEOs are less likely to identify with their organization and thus display servant leadership behavior (Peterson et al., 2012, p. 565, pp. 585-591). However, narcissistic CEOs that *do* identify with their firm were found to positively influence TMT behavioral integration (i.e. the degree to which TMT members act collaboratively, exchanging information and jointly make decisions), leading to better firm performance (Reina et al., 2014, p. 958, pp. 967-969). Reina et al. (2014) also support the findings of Peterson et al. (2012) by showing that narcissistic CEOs with low organizational identification negatively impact ROA through poorer TMT behavioral integration outcomes (p. 966).

Using signature size as a novel proxy to measure narcissism but not overconfidence, Ham et al. (2018) found that - as a result of overinvestment in R&D and M&A activities – narcissistic CEOs are linked to lower profitability and operating cash flow (OCF) but, curiously, increased compensation for the narcissistic CEO (Ham et al., 2018, p. 234).

Furthermore, Narcissistic CEOs have been found to cause their firms to suffer more losses during economic downturns, but also to recover faster in the subsequent recovery phase, suggesting that they ignore potential risks while focusing on desirable outcomes, which affects firms differently depending on their financial health (Patel & Cooper, 2013, p. 1528, p. 1537). These results highlight the potential double-edged nature of CEO narcissism and its impacts on firm performance, emphasizing the different situational factors and contexts in which CEO narcissism could be either useful or detrimental.

In conclusion, firm performance outcomes of narcissistic CEO behavior are characterized by higher volatility, greater financial risk, and short-term profits at the expense of long-term stability. While firms experience higher volatility in their profitability, the evidence on average ROA is mixed. As narcissistic CEOs are more prone to manipulating earnings, undertake risky strategic decisions, and excessively engage in M&A and R&D activities, credit ratings suffer. Due to their self-serving nature, measures aimed at improving short-term performance also increases short-term stock performance, which poses questions about the long-term alignment of interests between the CEO (agent) and shareholders (principal). The degree to which narcissistic CEOs identify with their organizations appears to determine whether narcissistic CEOs will exert dysfunctional, centralized decision-making or increase TMT behavioral integration. Similarly, the manifestations of CEO narcissism may depend on situational and environmental factors that determine success or failure. These findings provide the backdrop against which the following chapter examining stock market reactions to narcissistic CEO behavior outcomes is elaborated upon.

2.3.5 Stock Market Reactions

Haynes et al. (2016) found executive greed, defined as the excessive pursuit of material wealth, to negatively impact shareholder returns (p. 555), measured as total shareholder return (TSR, i.e. the change in stock price plus dividends over one year). The concept of greed is closely related to that of narcissism, in that narcissistic individuals will exhibit greedy behaviors (Kernberg, 1980, p. 136) in order amass personal material wealth, reinforcing their own perception of superiority and attracting praise and admiration. This results in a misalignment between CEO and shareholder prior-

ities, increasing agency costs and thus reducing shareholder value (Haynes et al., 2016, pp. 569-570). The effect of CEO greed on shareholder returns was moderated by tenure (i.e. as greedy CEOs age and gain status through experience, they tend to engage in less self-serving behavior), governance (i.e. powerful, independent boards), and oversight (e.g. disclosure requirements; Haynes et al., 2016, p. 555).

Similarly, research by Malmendier and Tate (2008) linked CEO overconfidence to negative cumulative abnormal returns (CAR) in the 3-day window of merger announcements (p. 20, p. 34). Overconfidence is the individual's false belief to judge potential outcomes more accurately than they actually do, which is rooted in an attribution bias of control over the situation (Haynes et al., 2016, p. 567). Seeing as narcissism involves the belief in one's intellectual superiority, the findings by Malmendier and Tate (2008) lend itself to the derivation of a connection between overconfidence and narcissism. Additionally, Boamah and Banerjee (2024) show that the market reaction to repurchase announcements is lower for narcissistic CEOs, suggesting that investors anticipate a low follow-through rate because narcissists view their firms as undervalued even when they are not, and using announcements as a tool in an attempt to signal market confidence (p. 717, pp. 719, 750). Supporting these findings, in examining 146 M&A deals from 2002 to 2006, Aktas et al. (2016) show that acquirer shareholders react more negatively to deals involving target CEOs with higher levels of narcissism (p. 113). This could suggest that investors expect aggressive negotiations with the target CEO, anticipating the acquiring firm to overpay, the deal to collapse, or the target firm's leadership to integrate poorly into the acquiring firm due to refusal by the acquiring CEO to relinquish part of his power and influence. Similarly, narcissism in CEOs - by virtue of overvaluing brand assets - induces negative stock market reactions when acquiring brands, as investors suspect the CEO to overvalue the asset they are trying to acquire, thereby inflating the price of the deal. Conversely, stock prices of the selling firm increase (Cao & Xu, 2022, p. 777, pp. 794-797).

Investors appear to be able to recognize and adjust to the positive tone biases employed by narcissistic CEOs in earnings disclosures (Marquez-Illescas et al., 2018, p. 401). Similarly, CEO narcissists voluntarily provide a higher amount of financial information - such as earnings guidance (Marquez-Illescas & Zhou, 2023, p. 221) - in an effort to seek attention and gain validation (Olsen et al., 2013, p. 264), allowing analysts to more accurately forecast earnings of companies led by narcissistic CEOs (Marquez-Illescas & Zhou, 2023, p. 236). Similarly, CEO Narcissism has been linked to lower stock recommendations from analysts, suggesting that analysts perceive narcissistic CEOs as self-serving and thus misaligned with shareholder priorities (Turner et al., 2024, pp. 1-3, pp. 14-16).

Investors and analysts are potentially misled by a CEO narcissist's earnings management activities, particularly in the domain of opportunistically excluding recurring items from non-GAAP earnings - which is at the discretion of the CEO - as suggested by Abdel-Meguid et al. (2020, p. 11).

In summary, stock markets tend to react negatively to performance outcomes of CEO narcissism, particularly in financial decision-making, M&A, and earnings management. Investors anticipate misaligned priorities, overconfidence, and self-serving behaviors, leading to weaker stock reactions to repurchase announcements, negative CARs following acquisitions, and lower analyst stock recommendations. Narcissistic CEOs frequently overvalue acquisitions, manipulate earnings through non-GAAP exclusions, and use financial disclosures as tools for self-promotion rather than transparency. While they provide more financial information, investors and analysts adjust for their biases, recognizing patterns of exaggeration and opportunism. Although narcissistic CEOs drive aggressive expansion and market visibility, their decision-making increases volatility, raises agency costs, and heightens long-term risk, reinforcing a generally cautious stance from investors.

2.4 Hypothesis

Narcissistic CEOs possess a strong, persistent need for admiration and approval, born of their inherently fragile self-esteem. This need drives them to short-term strategic decisions for immediate exposure rather than longer-term firm success. Empirical findings consistently demonstrate that narcissistic CEOs actively pursue high-risk investments, such as extravagant acquisitions, excessive R&D spending, and aggressive foreign expansion, often in the pursuit of personal aggrandizement rather than the interest of increased firm success. Their tendency to manage earnings reports and financial disclosures further fuels investor concerns by increasing agency costs and obscuring genuine corporate performance.

Narcissistic CEOs also tend to shape organizational culture to be permissive of their behavior, either by installing compliant or like-mindedly narcissistic board members, which diminishes critical oversight and increases organizational instability. Market participants recognize these behavioral proclivities and price their assessments accordingly, anticipating greater financial volatility, strategic misalignment, and ethical transgressions. This recognition is manifested as investor skepticism and negatively impacts stock valuations, especially when CEOs publicly announce strategic plans or leadership succession.

With investors' sensitivity to these destructive proclivities and their long-term implications on firm stability and profitability, market reactions at the time of CEO appointment announcements are assumed to be cautious, adjusting for anticipated agency misalignment through self-interested behavior. It is therefore hypothesized that during CEO turnover events:

> *H1: CEO narcissism is negatively related to stock price at the time of their announcement.*

3 Data and Methodology

The empirical study examines 341 CEO turnover events from 341 unique NASDAQ-listed firms, covering the time period from 2004 to 2024. The dataset is cross-sectional, meaning each firm appears only once, and captures a snapshot of market reactions at the time of a CEO turnover event rather than tracking firms over multiple years.

Financial data is gathered using the database service Datastream/EIKON, accessed through institutional resources, while public documents filed with the United States Securities and Exchange Commission (SEC) provided CEO-level data, such as signature size and duality status (i.e. CEO and Chairman role), and was used to compute data that was inaccessible in Datastream/EIKON. To ensure reliability of the analyses, narcissism measures are winsorized and log-transformed to adjust for skewness and maintain statistical robustness (see Chapter 4: Results).

To test the relationship between CEO Narcissism and CAR, OLS regression models are used. Additionally, the sample is then split into categorized low- and high-narcissism groups to explore non-linearity affects on CAR through subgroup analysis.

3.1 Independent Variable

Levels of narcissism on the subclinical spectrum are measured using signature size, a novel archival and unobtrusive instrument that has been used to measure CEO narcissism in prior research (e.g. Boamah and Banerjee, 2024, p. 719; Hou et al., 2024, p. 42), and validated as a tool to measure underlying narcissism levels (Ham et al., 2018, p. 234, p. 241; Junge et al., 2024, p. 16).

CEO signatures are extracted from DEF 14A proxy statements filed annually with the SEC. Proxy statements are searched stepwise, beginning with the first filing following the CEO announcement and continuing until either a signature is found, or the CEO's tenure comes to an end. The total signature area is determined by calculating the rectangular bounding box enclosing the CEO's handwritten signature as formulated in Equation (1):

$$A_{signature} = H \times W \tag{1}$$

Where:
$A_{signature}$ total enclosed area of the CEO's signature
H height of the signature bounding box
W width of the signature bounding box

To account for variation in CEO name length, a normalized signature size is calculated in (2):

$$A_{normalized} = \frac{A_{signature}}{N_{letters}} \tag{2}$$

Where:
$A_{normalized}$ signature size adjusted for length of name
$N_{letters}$ number of characters in the CEO's full name

Since signature sizes tend to exhibit skewness, winsorization is applied according to Equation (3): Outliers at the 1st and 99th percentiles are replaced with predefined percentile values. Extreme observations have their influence reduced while the distribution's overall structure is preserved. It is computed as:

$$A_{winsorized,i} = \begin{cases} P_{\alpha} & \text{if } A_{normalized,i} < P_{\alpha} \\ A_{Normalized,i} & \text{if } P_{\alpha} \leq A_{normalized,i} \leq P_{1-\alpha} \\ P_{1-\alpha} & \text{if } A_{normalized,i} > P_{1-\alpha} \end{cases} \tag{3}$$

Where:
$A_{winsorized,i}$ winsorized signature size for CEO i
$A_{normalized,i}$ signature size per letter for CEO i, before winsorization
P_{α} lower percentile threshold (1st percentile)
$P_{1-\alpha}$ upper percentile threshold (99th percentile)
α winsorization level (1%)

3.2 Dependent Variable

Cumulative Abnormal Returns (CAR) have been widely used in CEO narcissism research as a performance measure of stock price around corporate event studies (e.g. Aktas et al., 2016; Malhotra et al., 2017; Marquez-Illescas et al., 2018). In order to assess market reactions surrounding CEO turnover events, Cumulative Abnormal Returns are computed using the market-adjusted model. The event window was set at (-2, +2) days and an estimation window of (-180, -30) days relative to the turnover event was chosen. Daily stock returns are computed using arithmetic returns, defined in (4) as:

$$R_{it} = \frac{P_{it} - P_{it-1}}{P_{it-1}} \tag{4}$$

Where:
R_{it} return for firm i on day t
P_{it} closing stock price of firm i on day t
P_{it-1} closing stock price of firm i on day $t-1$

In Equation (5), using the market-adjusted model, expected returns are calculated against the S&P500 index as a broad market benchmark. Therefore, expected returns for the firm at day t are derived from the S&P500 returns on day t :

$$E(R_{it}) = R_{mt} = \frac{M_t - M_{t-1}}{M_{(t-1)}} \tag{5}$$

Where:
M_t closing price of S&P500 on day t
M_{t-1} closing price of S&P500 on day $t-1$
$E(R_{it})$ expected return for firm i on day t
R_{mt} return of the S&P500 on day t

The abnormal return (AR) for each day within the event window is calculated in (6) by subtracting $E(R_{it})$ from R_{it}. Summing the abnormal returns over the event window results in the value for CAR:

$$CAR_i = \sum_{t=-2}^{t=+2} AR_{it} \tag{6}$$

where CAR_i is the cumulative, (i.e. total) abnormal (i.e. not captured by S&P during time t) return for firm i. Stock market data is extracted from the Datastream/EIKON database.

3.3 Controls

CEO duality is frequently used as a control measure of how much power and influence the CEO possesses within the company, particularly in terms of corporate governance, strategic decision making and power centralization (e.g. Buyl et al., 2017; Dabbebi et al., 2022; Zhu and Chen, 2014a), affecting market perceptions. DEF 14A proxy statements entail this information with very little exceptions. CEO duality is encoded as a dummy variable (i.e. 1 if the CEO is also the chairman, and 0 otherwise).

Earnings per share (EPS), a firm level profitability metric for publicly listed companies, is calculated in Equation (7):

$$EPS = \frac{\text{Net Income}}{\text{Weighted Average Number of Common Shares Outstanding}} \tag{7}$$

Where *Net Income* describes the total earnings of the company after taxes and expenses, and *Weighted Average Number of Common Shares outstanding* refers to the number of shares outstanding adjusted (weighted) over the reporting period. EPS is closely watched by market observers, business media, and investors (Olsen et al., 2013, p. 247). It is directly reported in and extracted from the last filed annual report (10-K) that the company released before the CEO turnover event. As annual reports typically contain further historical data on EPS, the EPS figures from the past three fiscal years before the CEO turnover event are collected and averaged out, combining to paint a broader picture of the company's profitability prior to the announcement of a new CEO.

Total Assets, i.e. the summed of a company's – material as well as intangible – assets, is an accounting item that is frequently used in business research to control for firm size using the natural logarithm transformation (e.g. Cao and Xu, 2022; Dabbebi et al., 2022; García-Meca et al., 2021; J. H. J. Kim and Anderson, 2024; Yook and Lee, 2020). Its value is extracted from 10-K annual reports from the last year before the CEO turnover event and logarithmized in Equation (8) in accordance with prior research to stabilize variance and reduce skewness:

$$\ln(TotalAssets_i) = \log_e(TotalAssets_i) \tag{8}$$

Where:

$\ln(TotalAssets_i)$ natural logarithm of total assets for firm i

$\log_e$ natural logarithm function with base e

Industry is introduced as a category to control for in order to prevent industry-specific trends from biasing the results, such as the competitive environment, regulatory guidelines, innovation rates, and resulting market expectations. The 2-digit Standard Industrial Classification (SIC) numbers are used to encode all nine industry types in categorical dummies numbered 0-8 in order to account for industry fixed effects in the regression model analysis. Companies' industry codes are retrieved from the SEC database.

The *Year* of the CEO turnover event is extracted from every observation and encoded as dummy variables similar to *industry* in order to account for broad market trends that are not captured by the market-model of CAR. Macroeconomic trends, such as economic cycles, financial crises, interest rates, or major policy shifts are accounted for by running regressions using fixed year effects.

External hire is a binary control variable that indicates whether the newly appointed CEO was hired from outside the organization or promoted internally. External hires can influence investor perceptions as they represent a shift in corporate strategy and are coded as 0 if the appointed CEO's name appears in 10-K, 10-Q or DEF 14A filings in the year prior to the CEO turnover event, and as 1 otherwise.

The *Earnings* dummy variable captures whether an annual (10-K) or quarterly (10-Q) earnings announcement took place within the event window of the CEO turnover announcement, which could obscure the isolated impact on CAR and is controlled for by coding 1 for coincidental earnings announcements, and 0 otherwise.

3.4 Regression

Ordinary Least Squares (OLS) regression is the most widely used statistical method for estimating relationships between a dependent variable and one or more independent variables. It minimizes the sum of squared residuals to provide the best linear fit, making it both easy to interpret and computationally straightforward. Due to its robustness and transparency, OLS is employed as a regression model to detect linear dependencies in the observed data. Two data models are used: first, a baseline approach in (9), regressing CAR as the dependent variable on CEO Narcissism as the independent variable:

$$CAR_i = \beta_0 + \beta_1 CEONarcissism_i + \epsilon_i \tag{9}$$

Where:

β_0 intercept term (expected value of CAR_i when CEO Narcissism is zero

β_1 regression coefficient (change in CAR_i for one-unit increase in CEO Narcissism

$CEONarcissism_i$ log-transformed and winsorized mea-

sure of CEO Narcissism of the CEO for firm i
ϵ_i error term (residual)

Subsequently, a second regression is performed in (10), adding controls and fixed effects to enhance the model's explanatory power (*Full Model*):

$$\begin{aligned} CAR_i = {} & \beta_0 + \beta_1 CEONarcissism_i + \beta_2 ExternalHire_i \\ & + \beta_3 Earnings_i + \beta_4 CEODuality_i + \beta_5 EPS_i \\ & + \beta_6 \ln(TotalAssets_i) \\ & + \sum_{j=1}^{8} \gamma_i Industry_{ij} + \sum_t \delta_t Year_{it} + \epsilon_i \end{aligned} \tag{10}$$

The raw distribution of CEO Narcissism values exhibits characteristics that suggest the need for a transformation along the logarithmic scale. Values of CEO Narcissism are heavily right-skewed (see Fig. 4), with a small number of extreme values reaching far beyond the majority of observations (see Fig. 3). This "tail" can also be observed in the heavy departures from normality in the Q-Q plot (see Fig. 4). This suggests that the assumption of normally distributed residuals may not hold, resulting in distorted linear relationships captured in OLS regression. Log transforming using Equation (11) extends the distance between smaller values and reduces it for outliers on the right tail of the distribution, making it more symmetrical:

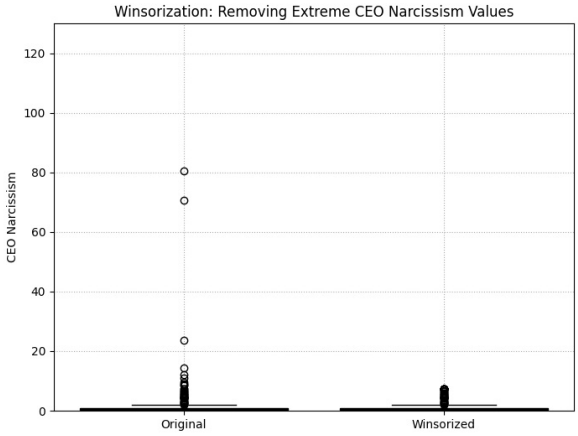


Figure 2: Winsorization of independent variable (own illustration)

4 Results

$$CEO'_{Narcissism} = \log(1 + CEO_{narcissism}) \tag{11}$$

The log transformation reduced skewness and improved the normality of the distribution while preserving relative differences. Figure 5 shows a more evenly spread range of CEO Narcissism values.

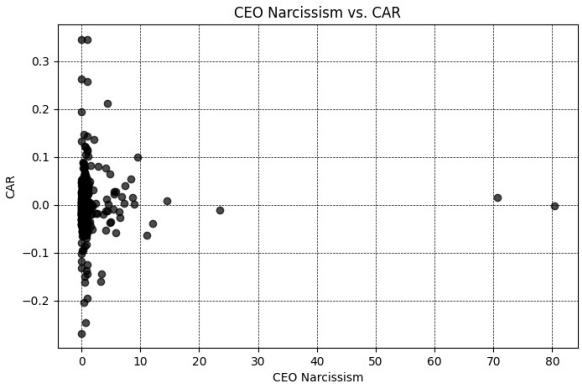


Figure 3: IV/DV scatter plot (own illustration)

Following these adjustments, descriptive statistics (Table 1) of key continuous variables was computed, showing that firms in the sample experienced an average of 6.5% abnormal stock price deviation during the event window while containing both very high (34.6%) and low (-26.8%) observations at the extremes. CAR skewness (0.86) and kurtosis (7.43) confirm pooling of values around the mean as well as few but stark outliers.

Table 1: Descriptive Statistics (own illustration)

	count	mean	std	min	25%	50%	75%	max
CEO Narcissism	341	0.516	0.528	0.005	0.143	0.397	0.630	2.122
CAR	341	0.000	0.065	-0.268	-0.028	-0.003	0.024	0.346
EPS	341	0.870	2.738	-18.000	-0.350	0.780	2.160	11.900
Total Assets	341	21.042	2.385	11.608	19.659	21.045	22.738	27.777

A Pearson matrix was generated to examine preliminary relationships between key continuous variables. CEO Narcissism showed moderate correlations with key financial variables, including EPS and Total Assets, but no evidence of multicollinearity is apparent (see Table 2).

Table 2: Correlation Matrix (own illustration)

	CEO Narcissism	CAR	EPS	Total Assets
CEO Narcissism	1.000	0.037	-0.006	-0.030
CAR	0.037	1.000	-0.058	-0.107
EPS	-0.006	-0.058	1.000	0.340
Total Assets	-0.030	-0.107	0.340	1.000

To further assess multicollinearity possibilities, Variance Inflation Factor (VIF) analysis shows values well below accepted thresholds (<10, <5), confirming that multicollinearity is of no concern in the model (see Table 3).

Residual diagnostics reveal signs of heteroskedasticity in the residual plot (Fig. 6), violating the OLS assumption of constant variance in error terms. Specifically, residual variance increases at certain fitted values, suggesting potential bias in traditional OLS standard errors. To address this, the

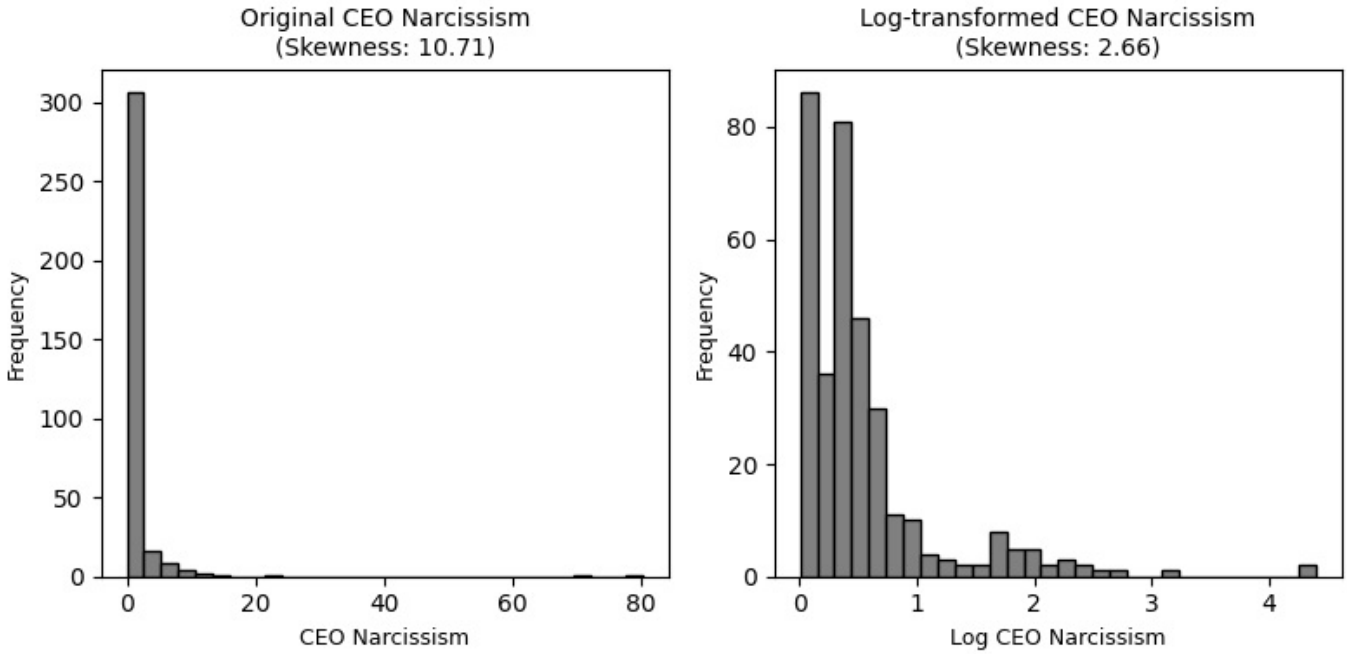


Figure 4: Histogram differences after transformation (own illustration)

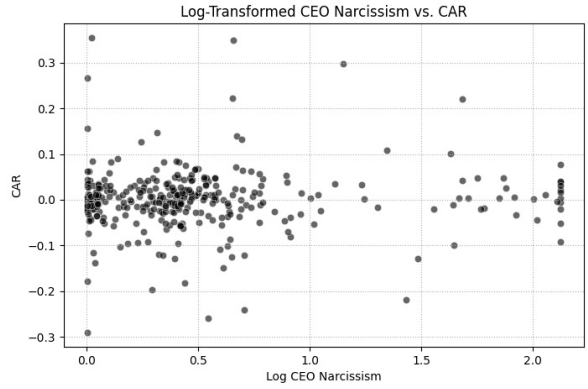


Figure 5: Scatter plot after transformation (own illustration)

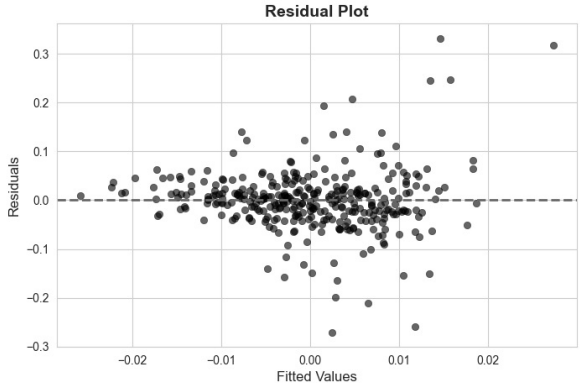


Figure 6: Residual plot (own illustration)

Table 3: VIF Analysis (own illustration)

Variable	VIF
CEO Narcissism	1.931000
CAR	1.005000
EPS	1.133000
Total Assets	2.082000

model was estimated using HC1 heteroskedasticity-robust standard errors (White-corrected errors). This adjustment corrects the variance-covariance matrix, ensuring that standard errors remain valid, and significance levels accurately reflect the presence of heteroskedasticity in the data. Thus, to ensure statistical validity given heteroskedasticity, all standard errors reported in Table 4 are HC1 robust standard errors.

Table 4 presents the results of the Ordinary Least Squares (OLS) regression estimating Cumulative Abnormal Returns (CAR) as a function of CEO Narcissism and control variables, while incorporating industry and year fixed effects (not included in the table for brevity).

The primary independent variable, CEO Narcissism, does not exhibit a statistically significant relationship with CAR in the full model (β = 0.0097, SE = 0.0065, p = 0.151). The 95% confidence interval (-0.003, 0.022) includes zero, confirming that the effect is indistinguishable from no impact. Among control variables, CEO Duality is the only statistically significant predictor (β = -0.0166, SE = 0.0081, $p < 0.05$), indicating that firms where the CEO also serves as board chair

Table 4: Regression Results (own illustration)

Standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

	Dependent variable: CAR	
	Baseline Model	Full Model
CEO Narcissism	0.0045	0.0097
	(0.0061)	(0.0065)
CEO Duality		-0.0166**
		(0.0081)
EPS		-0.0005
		(0.0017)
Total Assets		-0.0012
		(0.0023)
External Hire		-0.0142
		(0.0108)
Earnings		-0.0189
		(0.0096)
Year Fixed Effects		Yes
Industry Fixed Effects		Yes
R-squared	0.0014	0.1286
Adj. R-squared		0.0381
F-Statistic	0.5566 (df=1; 339)	10.68*** (df=32; 308)
Observations	341	341
Residual Std. Error	0.0646	0.0633

experience significantly lower CAR. The confidence interval for CEO Duality (-0.034, -0.002) does not include zero, reinforcing its significance. Earnings, EPS, External Hire, and Total Assets remain insignificant, with confidence intervals spanning zero, suggesting they lack explanatory power for CAR in this context. The model fit improves substantially with the inclusion of industry and year fixed effects. The R^2 increases from 0.0014 (Baseline) to 0.1286 (Full Model), indicating that fixed effects explain a notable portion of the variance in CAR. However, the Adjusted R^2 (0.0381) remains relatively low, suggesting that the model accounts for only a limited fraction of CAR variation. The F-statistic (10.68, $p < 0.01$) confirms that the model as a whole is statistically significant, even though individual predictors remain weak.

5 Discussion

This empirical study examines whether CEO Narcissism has a significant impact on Cumulative Abnormal Returns following firm announcements of CEO turnover events. Contrary to expectations, the results suggest that CEO Narcissism does not have a statistically significant effect on CAR. Hence, Hypothesis 1 is rejected. This indicates that investor reactions may not be systematically influenced by narcissism as a personality trait in CEOs at the time of their employment announcement. Instead, results show that CEO Duality, a measure of CEO power, emerges as a significant predictor, with firms experiencing lower returns when the CEO successor also serves as the chairman of the board. The findings challenge prior assumptions that narcissistic leadership directly shapes market reactions, instead suggesting that shareholders may be prioritizing governance factors over CEO personality traits in their assessments of CEO turnover events. In discussing the emergent results, findings are attempted to be contextualized and reconciled with those in the extant literature, exploring potential explanations for the observed results and outlining the implications and limitations of the methodology that was chosen.

5.1 Analysis

Stock markets may not price in narcissistic traits upon a CEO's appointment as such personality trait information is not easily observable or quantifiable at the time of the announcement, which typically highlights the CEO's experience and accomplishments rather than psychological profiles, meaning investors lack concrete data on narcissism to evaluate and trade on. In efficient markets, readily available public information gets absorbed quickly and is reflected in moving stock prices - if a CEO's narcissism is not explicitly disclosed or overtly visible, shareholders and investors have little to no basis to price it in. Unless the CEO has a notorious public persona or a prior track record that strongly hints at narcissistic behaviors, the underlying trait could remain hidden until after the announcement of a CEO turnover event.

Similarly, even if information on narcissistic traits were available to the stock market during the announcement, it is not necessarily always rational to price it in immediately: while extant literature has linked CEO narcissism to negative outcomes on firm performance, it's the narcissists behavioral manifestations that are affecting these results (e.g. by way of altering a company's strategic approach or cultural positioning), not the trait alone. Investors and analysts might only start evaluating a new CEO once they start employing strategy shifts or communicate with stakeholders directly through interviews, corporate filings or announcements early into their tenure - but not based on initial impressions at the time of the CEO turnover announcement. At the appointment stage, personality traits like narcissism may not be getting priced in as the market awaits behavioral confirmation (or disconfirmation) of whether the new CEO's leadership is a) genuinely influenced by narcissistic manifestations of behavior and b) helpful or detrimental to the firm. As prior empirical findings suggest, narcissistic CEOs reshape the strategic, cultural and ethical directions of their firms. Assuming these processes unfold over time rather than overnight, the immediate risk of these potential outcomes might not yet be apparent to shareholders at the time of the CEO succession announcement.

Alternatively, non-significant findings could be explained by cognitive biases that lead to stock market inaction during CEO turnover events: as investors are presented with new information, they are slow to update their prior beliefs (conservatism) and initially underreact to the true detrimental risk of narcissistic CEOs, adjusting their sentiment over time instead of immediately following the announcement (Barberis et al., 1998, pp. 314-316). For example, if investors exhibit neutral optimism towards a new CEO *a priori*, they could be underweighting intangible traits like narcissism and focus on hard information (e.g. past performance, current financial

shape of the company) out of cautious optimism or simply uncertainty. Only once the CEO narcissist's actions provide a signal – for instance, a first earnings call that is absurdly optimistic in tone or the rushed initiation of a large acquisition process – do investors update their evaluation on narcissism as a leadership trait and adjust their risk perception of increased agency costs in narcissistic CEOs, forcing a negative stock market reaction (e.g. Buchholz et al., 2018; Marquez-Illescas et al., 2018).

CEO Duality negatively influences stock prices during CEO succession announcements. As occupying both the position of CEO and chairman of the board grants the CEO more centralized and unchecked power, suggests weak governance mechanisms, and harms shareholders if that power is wielded in the CEO's self-interest as opposed to the best interests of shareholders, leading to increased agency costs – this effect would then be more pronounced if the CEO possesses narcissistic tendencies, as those can confidently assumed to be self-serving by investors (Lassoued & Khanchel, 2022). During CEO turnover announcements, duality of top executive positions is explicitly disclosed in a straightforward and direct manner (hard information), which makes it easy for stock markets to digest and react to. Conversely, personality dimensions like narcissism are harder to gauge, especially if the successor CEO has not been in the spotlight prior to the turnover event. The market may absorb concrete, publicly available information – like duality – more swiftly, while more nebulous cues about the CEO's personality are slower to be priced in. The immediate announcement window is too short to capture the market's collective judgement on narcissism as opposed to duality, leading to the observed results.

In their initial assessment of narcissistic characteristics in a CEO successor, stock markets may conflate narcissism with related but distinct personality traits: overconfidence, i.e. the tendency to have excessive confidence in one's own judgement, is a trait overtly similar to narcissism (Campbell et al., 2004). Both overconfident and narcissistic CEOs are linked to more positive tone in disclosures and announcements (Schrand & Zechman, 2012, p. 311). However, mere overconfidence, unlike narcissism, is not linked to increased agency costs for shareholders due to a misalignment of interests. Similarly, extraversion helps narcissists emerge as leaders (Grijalva et al., 2014, p. 1), although it represents only a small part of the broad spectrum of traits that are inherent to narcissism. Perhaps as part of their neutral optimism, investors conflate a narcissist's assertive communication and self-assuredness and with overconfidence, extraversion, or simple charisma and charm. The short event window around CEO succession announcements may not be sufficiently suited to capture more than investors' initial assessment of CEO characteristics, making it difficult to properly distinguish narcissism and its self-serving features from a bold or optimistic CEO who is in alignment with shareholder interests.

5.2 Limitations

The study comes with several limitations. First, CEO signature size serves as a practical yet potentially limited proxy for narcissism: its one-dimensional nature may capture only some facets of the broader, multidimensional narcissism spectrum such as grandiosity, ego, hubris, or dominance (Cragun et al., 2019, p. 915). Further, signature size may be reflective of unrelated factors like physical traits, handwriting habits/instructions, or cultural norms rather than narcissism alone (Hou et al., 2024, p. 58). These factors may introduce measurement noise, raising validity concerns. Indeed, Abdel-Meguid et al. (2020) confirmed their hypotheses linking CEO narcissism to increased earnings engagement using a composite narcissism measure which included relative cash compensation, relative noncash compensation, and CEO picture size and prominence in 10-K annual reports; however, none of their hypotheses could be confirmed using signature size as a singular measure (p. 39). While practical, signature size's indirectness, susceptibility to unrelated influences, and difficulty distinguishing between narcissism from related traits highlight its methodological limitations.

Several further methodological limitations must be acknowledged when interpreting the results of the study. A potential key variable that was not controlled for is the type of CEO turnover in the context of a forced vs. voluntary departure: stock markets may react differently to CEO turnover events depending on whether the incumbent CEO departed the firm in a planned measure or was forced to leave due to internal board pressures as a result of poor performance. In forced turnovers, stock reactions may be positive, signaling relief or optimism in anticipation of a turnaround in performance, or negative, indicating that the CEO departure is merely a symptom of larger problems rooted more deeply within the organization. In voluntary turnovers, no such problems might be apparent since the departure is planned and not a broader signal of distress within the company, leading to more muted market reactions. By not controlling for this, the model runs the risk of conflating the effects of successor CEO personality traits with the nature of the departure. For example, if highly narcissistic CEOs tend to be hired when a company is in crisis due to the need for a strategic turnaround and the alignment of risk profiles between the narcissistic CEO and a firm experiencing financial distress (Kliger & Tsur, 2011, p. 29), the resulting positive CAR could be mistakenly misattributed to narcissism by the model. Conversely, if a long-tenured and successful CEO retires and is succeeded by a narcissistic CEO outside hire, the market may react negatively due to fear of digression from the established and successful company strategy, again confounding the trait effect with turnover type. As such, the omission of a forced-turnover indicator presents a methodological gap that could introduce omitted variable bias, meaning the coefficient of CEO narcissism would partly reflect unobserved turnover circumstances. External hire dummies have been included in the model, which may be able to partially capture forced turnovers (as those are more likely to be accompanied by an outsider CEO getting hired), but ultimately only serving as

an imperfect proxy for the latter.

Further, while the average three-year EPS is included as a profitability measure in the model, this variable may not fully capture recent performance trajectory. For instance, a firm could have a declining performance trend even if the three-year average is positive. In addition, the omission of further performance metrics narrows the view of firm performance and makes it more difficult to capture the overall financial state the company is in at the time of a CEO turnover event. For instance, Tobin's Q is a forward-looking indicator that reflects investor expectations of future growth, while return on assets (ROA) captures operational efficiency. Both variables are frequently used in existing research. Furthermore, leverage and liquidity ratios signal financial risk and stability; by not including these metrics as variables, the applied model may miss important cues about firm conditions and investor expectations, resulting in a less comprehensive understanding of firm health and profitability, which stock markets use to adjust their evaluations of a company.

Since the sample includes strictly NASDAQ-listed corporations, the findings are limited to companies that are large, U.S. based, and public. Smaller or privately held firms may have different governance structures and stakeholder constellations, resulting in different CEO turnover dynamics. Firms outside the U.S. might operate under distinct conditions in the regulatory or cultural context. As such, the findings lack external validity and are outside the scope of mid-sized companies or geographically different markets.

By using industry fixed effects rather than examining specific industrial sectors, the model assumes that the impact of CEO narcissism is uniform across industries. However, the regulatory and competitive conditions inherent to certain industries may influence the effect of CEO narcissism on firm outcomes (and, by extension, investor sentiment). For instance, some industries may amplify a narcissistic leader's impact – high-tech or entertainment sectors may value bold visions and risky, grandiose strategy impulses differently than the real estate and construction industry. As a result, this study may be overlooking the variance of CEO narcissism effects on different industry conditions.

Finally, this study is limited in that it does not account for the possibility that narcissistic CEOs either self-select or are deliberately selected into certain types of firms or firm contexts, raising concerns about endogeneity and reverse causality. Boards might deliberately seek overconfident, narcissistic leaders within specific circumstances (e.g. underperformance or emerging industries dependent on innovation). If such CEOs are disproportionately found in certain environments, the observed market reaction might be reflecting the underlying firm characteristics or conditions rather than the CEO's personality traits alone: firms that attract narcissistic CEOs could inherently have trajectories or issues that determine investor reactions. Limited insight on self-selection biases makes causal inference of observed correlations difficult.

5.3 Future Research

Extant research on CEO narcissism has primarily focused on downstream effects of narcissistic CEO behavioral outcomes on the firms they lead. However, it is still unclear under which specific conditions demand for narcissism as a personality trait in leaders emerges. While this study was not able to provide evidence for shareholders driving demand for appointments of narcissistic CEOs through the board of directors, it is crucial to understand why boards might opt to prefer narcissism as a character trait to advance company success (or avoid failure) – as such, the situational contexts in which narcissism may thrive have to be better understood. Boards may view narcissism as a helpful tool to advance transformational efforts within a struggling company, signal decisiveness to stakeholders, or more aggressively pursue growth and expansion strategies to secure market dominance. Yet, boards may overestimate the short-term benefits of narcissistic leadership (e.g. charisma, media attention, or anticipated shareholder optimism) while greatly underestimating the long-term risks that previous research has highlighted. Future research should consider the economic environments as a potential driving force behind demand for narcissism: for instance, highly regulated industries should perceive the narcissistic propensity for rule-breaking and excessive risk-taking differently than highly competitive industries, where disruption and innovation are critical (e.g. finance, biotech).

Moreover, it is critical to advance measurement techniques that reliably predict narcissism. Archival proxies, such as signature size or first-person pronoun use, remain widely used but face validity concerns (Van Scotter, 2019, p. 631). With the advancement of computational capabilities and the emergence of large language models (LLM), machine learning and sentiment analysis could provide a promising approach to developing novel narcissism measures by analyzing earnings call transcripts, CEO interviews, or shareholder letters. Expanding measurement to social media activity (e.g. LinkedIn; Junge et al., 2024, pp. 11-13) could provide a richer understanding of narcissistic self-promotion not only in CEOs but also in the TMT members they assemble.

6 Conclusion

This thesis set out to examine whether capital markets react systemically to the appointment of narcissistic CEOs by analyzing cumulative abnormal returns surrounding CEO turnover events. Drawing on Upper Echelons Theory and employing signature size as a measurement proxy for narcissism, the study investigated whether shareholders adjust their expectations in anticipation of the behavioral manifestations and strategic implications associated with narcissistic leadership.

The results provide no statistically significant evidence that successor CEO narcissism influences stock markets reactions at the time of a CEO turnover event. Though prior literature has demonstrated that narcissistic CEOs drive

their firms toward more aggressive strategies, risk-taking, and volatility in performance, the findings suggest that stock markets do not immediately factor these characteristics into their assessments upon CEO appointments. CEO duality – meaning the incoming CEO also assumes the role of chairman of the board – was instead found to negatively affect stock prices, highlighting the importance of governance structures and distributed power at the top executive level.

The findings contribute to the growing body of research on CEO personality, specifically narcissism, by offering a demand-side exploration on how markets perceive and evaluate narcissistic leadership at $t = 0$ of the CEO's tenure. Contrary to expectations that investors would preemptively discount companies that appoint narcissistic CEOs due to their documented risk profiles and negative downstream effects on firm performance, the results indicate that markets might instead prefer to focus on tangible, hard information (e.g. CEO duality as a corporate governance signal) rather than speculative assessments of personality traits.

While the study conducted in this bachelor thesis provides valuable insights, it is limited in its methodological approach. The use of signature size as an indirect, one-dimensional measure of narcissism, while validated in prior research, may not fully capture the complexity of narcissism as a personality trait. In addition, the event study methodology inherently captures short-term market reactions and does not account for the gradual nature in which personality traits translate and transform into real, tangible actions and behaviors that markets can more reliably assess. Future research could explore alternative measures of CEO narcissism, investigate long-term shareholder wealth effects, or more deeply examine the environmental factors that may drive demand for the appointment of narcissistic CEOs.

This research enhances the understanding of the interplay between executive personality traits and shareholder expectations. While narcissistic CEOs have been linked to strategic, cultural, and ethical risks, their appointment alone does not appear to trigger immediate market reactions. Rather, the findings suggest corporate governance structures to be the primary focus for investors when evaluating executive leadership transitions.

References

Aabo, T., Hanousek, J., Jr., Pantzalis, C., & Park, J. C. (2023). CEO Personality Traits and Corporate Value Implication of Acquisitions. *Journal of Empirical Finance*, *73*, 86–106. https://doi.org/10.1016/j.jempfin.2023.05.006

Abdel-Meguid, A., Jennings, J. N., Olsen, K. J., & Soliman, M. T. (2020). The Impact of the CEO's Personal Narcissism on Non-GAAP Earnings. *The Accounting Review*, *96*(3), 1–25. https://doi.org/10.2308/tar-2017-0612

Aktas, N., de Bodt, E., Bollaert, H., & Roll, R. (2016). CEO Narcissism and the Takeover Process: From Private Initiation to Deal Completion. *The Journal of Financial and Quantitative Analysis*, *51*(1), 113–137. http://www.jstor.org/stable/43862310

Bachrach, D. G., Yong Kim, K., Patel, P. C., & Harms, P. D. (2023). Birds of a Feather?: Firm Sales Growth and Narcissism in the Upper Echelons at the CEO-TMT Interface. *The Leadership Quarterly*, *34*(2), 101621. https://doi.org/10.1016/j.leaqua.2022.101621

Barberis, N., Shleifer, A., & Vishny, R. (1998). A Model of Investor Sentiment. *Journal of Financial Economics*, *49*(3), 307–343. https://doi.org/10.1016/s0304-405x(98)00027-0

Berson, Y., Oreg, S., & Dvir, T. (2007). CEO Values, Organizational Culture and Firm Outcomes. *Journal of Organizational Behavior*, *29*(5), 615–633. https://doi.org/10.1002/job.499

Boamah, E. O., & Banerjee, S. (2024). The Beguiling Behaviour of Narcissistic CEOs: Evidence from Repurchase Announcements. *Journal of Business Finance & Accounting*, *51*(3–4), 717–755. https://doi.org/10.1111/jbfa.12796

Bredlich, J. M., Kruse, S., & Brettel, M. (2021). Selecting for Competence or for the Ego: How Narcissistic CEOs Structure Their Top Management Teams. *Academy of Management Proceedings*, *2021*(1), 11782. https://doi.org/10.5465/ambpp.2021.47

Buchholz, F., Jaeschke, R., Lopatta, K., & Maas, K. (2018). The Use of Optimistic Tone by Narcissistic CEOs. *Accounting, Auditing & Accountability Journal*, *31*(2), 531–562. https://doi.org/10.1108/aaaj-11-2015-2292

Buchholz, F., Lopatta, K., & Maas, K. (2019). The Deliberate Engagement of Narcissistic CEOs in Earnings Management. *Journal of Business Ethics*, *167*(4), 663–686. https://doi.org/10.1007/s10551-019-04176-x

Buyl, T., Boone, C., & Wade, J. B. (2017). CEO Narcissism, Risk-Taking, and Resilience: An Empirical Analysis in U.S. Commercial Banks. *Journal of Management*, *45*(4), 1372–1400. https://doi.org/10.1177/0149206317699521

Campbell, W. K., & Foster, J. D. (2007). The Narcissistic Self: Background, an Extended Agency Model, and Ongoing Controversies. In C. Sedikides & S. J. Spencer (Eds.), *The Self* (pp. 115–138). Psychology Press.

Campbell, W. K., Goodie, A. S., & Foster, J. D. (2004). Narcissism, Confidence, and Risk Attitude. *Journal of Behavioral Decision Making*, *17*(4), 297–311. https://doi.org/10.1002/bdm.475

Cao, Z., & Xu, K. (2022). CEO Narcissism, Brand Acquisition and Disposal, and Stock Returns. *Journal of the Academy of Marketing Science*, *50*(4), 777–799. https://doi.org/10.1007/s11747-022-00850-1

Capalbo, F., Frino, A., Lim, M. Y., Mollica, V., & Palumbo, R. (2017). The Impact of CEO Narcissism on Earnings Management. *Abacus*, *54*(2), 210–226. https://doi.org/10.1111/abac.12116

Chatterjee, A., & Hambrick, D. C. (2007). It's All about Me: Narcissistic Chief Executive Officers and Their Effects on Company Strategy and Performance. *Administrative Science Quarterly*, *52*(3), 351–386. https://doi.org/10.2189/asqu.52.3.351

Chatterjee, A., & Hambrick, D. C. (2011). Executive Personality, Capability Cues, and Risk Taking. *Administrative Science Quarterly*, *56*(2), 202–237. https://doi.org/10.1177/0001839211427534

Chatterjee, A., & Pollock, T. G. (2017). Master of Puppets: How Narcissistic CEOs Construct Their Professional Worlds. *Academy of Management Review*, *42*(4), 703–725. https://doi.org/10.5465/amr.2015.0224

Cragun, O. R., Olsen, K. J., & Wright, P. M. (2019). Making CEO Narcissism Research Great: A Review and Meta-Analysis of CEO Narcissism. *Journal of Management*, *46*(6), 908–936. https://doi.org/10.1177/0149206319892678

Dabbebi, A., Lassoued, N., & Khanchel, I. (2022). Peering Through the Smokescreen: ESG Disclosure and CEO Personality. *Managerial and Decision Economics*, *43*(7), 3147–3164. https://doi.org/10.1002/mde.3587

Diamond, J. (2016). Trump on Looking Presidential: 'How Handsome Am I?' [Transcript] [Retrieved: 20.02.2025]. https://edition.cnn.com/2016/04/25/politics/donald-trump-how-handsome-am-i/index.html

Fromm, E. (1992). *The Art of Being*. Robinson.

García-Meca, E., Ramón-Llorens, M.-C., & Martínez-Ferrero, J. (2021). Are Narcissistic CEOs More Tax Aggressive? The Moderating Role of Internal Audit Committees. *Journal of Business Research*, *129*, 223–235. https://doi.org/10.1016/j.jbusres.2021.02.043

Gerstner, W.-C., König, A., Enders, A., & Hambrick, D. C. (2013). CEO Narcissism, Audience Engagement, and Organizational Adoption of Technological Discontinuities. *Administrative Science Quarterly*, *58*(2), 257–291. https://doi.org/10.1177/0001839213488773

Grenyer, B. F. S. (2013). Historical Overview of Pathological Narcissism. In J. S. Ogrodniczuk (Ed.), *Understanding and Treating Pathological Narcissism* (pp. 15–26). American Psychological Association. https://doi.org/10.1037/14041-001

Grijalva, E., Harms, P. D., Newman, D. A., Gaddis, B. H., & Fraley, R. C. (2014). Narcissism and Leadership: A Meta-Analytic Review of Linear and Nonlinear Relationships. *Personnel Psychology*, *68*(1), 1–47. https://doi.org/10.1111/peps.12072

Ham, C., Lang, M., Seybert, N., & Wang, S. (2017). CFO Narcissism and Financial Reporting Quality. *Journal of Accounting Research*, *55*(5), 1089–1135. https://doi.org/10.1111/1475-679x.12176

Ham, C., Seybert, N., & Wang, S. (2018). Narcissism Is a Bad Sign: CEO Signature Size, Investment, and Performance. *Review of Accounting Studies*, *23*(1), 234–264. https://doi.org/10.1007/s11142-017-9427-x

Hambrick, D. C., & Mason, P. A. (1984). Upper Echelons: The Organization as a Reflection of Its Top Managers. *The Academy of Management Review*, *9*(2), 193. https://doi.org/10.2307/258434

Haynes, K. T., Campbell, J. T., & Hitt, M. A. (2016). When More Is Not Enough. *Journal of Management*, *43*(2), 555–584. https://doi.org/10.1177/0149206314535444

Hou, Z., Fairchild, R., & Perotti, P. (2024). CEO Narcissism and Credit Ratings. *Journal of Business Ethics*, *197*(1), 39–72. https://doi.org/10.1007/s10551-024-05691-2

Judd, J. S., Olsen, K. J., & Stekelberg, J. (2017). How Do Auditors Respond to CEO Narcissism? Evidence from External Audit Fees. *Accounting Horizons*, *31*(4), 33–52. https://doi.org/10.2308/acch-51810

Judge, T. A., LePine, J. A., & Rich, B. L. (2006). Loving Yourself Abundantly: Relationship of the Narcissistic Personality to Self- and Other Perceptions of Workplace Deviance, Leadership, and Task and Contextual Performance. *Journal of Applied Psychology*, *91*(4), 762–776. https://doi.org/10.1037/0021-9010.91.4.762

Junge, S., Graf-Vlachy, L., Hagen, M., & Schlichte, F. (2024). Narcissism at the CEO–TMT Interface: Measuring Executive Narcissism and Testing Its Effects on TMT Composition. *Journal of Management*. https://doi.org/10.1177/01492063241226904

Kashmiri, S., Nicol, C. D., & Arora, S. (2017). Me, Myself, and I: Influence of CEO Narcissism on Firms' Innovation Strategy and the Likelihood of Product-Harm Crises. *Journal of the Academy of Marketing Science*, *45*(5), 633–656. https://doi.org/10.1007/s11747-017-0535-8

Kim, B., Lee, S., & Kang, K. H. (2018). The Moderating Role of CEO Narcissism on the Relationship Between Uncertainty Avoidance and CSR. *Tourism Management*, *67*, 203–213. https://doi.org/10.1016/j.tourman.2018.01.018

Kim, J. H. J., & Anderson, R. (2024). CEO Narcissism and the Agency Cost of Debt. *Journal of Empirical Finance*, *77*, 101477. https://doi.org/10.1016/j.jempfin.2024.101477

Kliger, D., & Tsur, I. (2011). Prospect Theory and Risk-Seeking Behavior by Troubled Firms. *Journal of Behavioral Finance*, *12*(1), 29–40. https://doi.org/10.1080/15427560.2011.555028

Lassoued, N., & Khanchel, I. (2022). Voluntary CSR Disclosure and CEO Narcissism: The Moderating Role of CEO Duality and Board Gender Diversity. *Review of Managerial Science*, *17*(3), 1075–1123. https://doi.org/10.1007/s11846-022-00555-3

Li, X., Wang, Y., Yang, M. M., & Tang, Y. (2022). Does Owner CEO Narcissism Promote Exporting SMEs' Market Spreading Strategy? Joint Effects of Asset-Specific Investments and Firm Exporting Experience. *Journal of International Management*, *28*(3), 100923. https://doi.org/10.1016/j.intman.2021.100923

Majors, T. M. (2015). The Interaction of Communicating Measurement Uncertainty and the Dark Triad on Managers' Reporting Decisions. *The Accounting Review*, *91*(3), 973–992. https://doi.org/10.2308/accr-51276

Malhotra, S., Reus, T. H., Zhu, P., & Roelofsen, E. M. (2017). The Acquisitive Nature of Extraverted CEOs. *Administrative Science Quarterly*, *63*(2), 370–408. https://doi.org/10.1177/0001839217712240

Malmendier, U., & Tate, G. (2008). Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction. *Journal of Financial Economics*, *89*(1), 20–43. https://doi.org/10.1016/j.jfineco.2007.07.002

Marquez-Illescas, G., Zebedee, A. A., & Zhou, L. (2018). Hear Me Write: Does CEO Narcissism Affect Disclosure? *Journal of Business Ethics*, *159*(2), 401–417. https://doi.org/10.1007/s10551-018-3796-3

Marquez-Illescas, G., & Zhou, L. (2023). CEO Narcissism and Properties of Analysts' Forecasts. *Accounting Horizons*, *37*(3), 221–241. https://doi.org/10.2308/horizons-2021-010

Martínez-Ferrero, J., García-Meca, E., & Ramón-Llorens, M. C. (2023). What if My Boss Is a Narcissist? The Effects of Chief Executive Officer Narcissism on Female Proportion in Top Management Teams. *Business Ethics, the Environment & Responsibility*, *32*(4), 1201–1216. https://doi.org/10.1111/beer.12559

Miller, J. D., Back, M. D., Lynam, D. R., & Wright, A. G. C. (2021). Narcissism Today: What We Know and What We Need to Learn. *Current Directions in Psychological Science*, *30*(6), 519–525. https://doi.org/10.1177/09637214211044109

Miller, J. D., & Campbell, W. K. (2010). The Case for Using Research on Trait Narcissism as a Building Block for Understanding Narcissistic Personality Disorder. *Personality Disorders: Theory, Research, and Treatment*, *1*(3), 180–191. https://doi.org/10.1037/a0018229

Morf, C. C., & Rhodewalt, F. (2001). Unraveling the Paradoxes of Narcissism: A Dynamic Self-Regulatory Processing Model. *Psychological Inquiry*, *12*(4), 177–196. https://doi.org/10.1207/s15327965pli1204_1

Musk, E. (2018). If I Am a Narcissist (Which Might Be True), at Least I Am a Useful One [Post]. https://x.com/elonmusk/status/1016005381686943744

Mutschmann, M., Hasso, T., & Pelster, M. (2022). Dark Triad Managerial Personality and Financial Reporting Manipulation. *Journal of Business Ethics*, *181*(3), 763–788. https://doi.org/10.1007/s10551-021-04959-1

Nevicka, B., De Hoogh, A. H. B., Van Vianen, A. E. M., & Ten Velden, F. S. (2013). Uncertainty Enhances the Preference for Narcissistic Leaders. *European Journal of Social Psychology*, *43*(5), 370–380. https://doi.org/10.1002/ejsp.1943

Oesterle, M.-J., Elosge, C., & Elosge, L. (2016). Me, Myself and I: The Role of CEO Narcissism in Internationalization Decisions. *International Business Review*, *25*(5), 1114–1123. https://doi.org/10.1016/j.ibusrev.2016.02.001

Olsen, K. J., Dworkis, K. K., & Young, S. M. (2013). CEO Narcissism and Accounting: A Picture of Profits. *Journal of Management Accounting Research*, *26*(2), 243–267. https://doi.org/10.2308/jmar-50638

Olsen, K. J., & Stekelberg, J. (2016). CEO Narcissism and Corporate Tax Sheltering. *The Journal of the American Taxation Association*, *38*(1), 1–22. https://doi.org/10.2308/atax-51251

O'Reilly, C. A., III, Doerr, B., & Chatman, J. A. (2018). "See You in Court": How CEO Narcissism Increases Firms' Vulnerability to Lawsuits. *The Leadership Quarterly*, *29*(3), 365–378. https://doi.org/10.1016/j.leaqua.2017.08.001

Ovid. (2004). *Metamorphoses* (R. J. Tarrant, Ed.). Oxford University Press.

Patel, P. C., & Cooper, D. (2013). The Harder They Fall, the Faster They Rise: Approach and Avoidance Focus in Narcissistic CEOs. *Strategic Management Journal*, *35*(10), 1528–1540. https://doi.org/10.1002/smj.2162

Peterson, S. J., Galvin, B. M., & Lange, D. (2012). CEO Servant Leadership: Exploring Executive Characteristics and Firm Performance. *Personnel Psychology*, *65*(3), 565–596. https://doi.org/10.1111/j.1744-6570.2012.01253.x

Petrenko, O. V., Aime, F., Ridge, J., & Hill, A. (2015). Corporate Social Responsibility or CEO Narcissism? CSR Motivations and Organizational Performance. *Strategic Management Journal*, *37*(2), 262–279. https://doi.org/10.1002/smj.2348

Raskin, R., & Hall, C. S. (1981). The Narcissistic Personality Inventory: Alternate Form Reliability and Further Evidence of Construct Validity. *Journal of Personality Assessment*, *45*(2), 159–162. https://doi.org/10.1207/s15327752jpa4502_10

Raskin, R., Novacek, J., & Hogan, R. (1991). Narcissism, Self-Esteem, and Defensive Self-Enhancement. *Journal of Personality*, *59*(1), 19–38. https://doi.org/10.1111/j.1467-6494.1991.tb00766.x

Raskin, R., & Terry, H. (1988). A Principal-Components Analysis of the Narcissistic Personality Inventory and Further Evidence of Its Construct Validity. *Journal of Personality and Social Psychology*, *54*(5), 890–902. https://doi.org/10.1037/0022-3514.54.5.890

Raskin, R. N., & Hall, C. S. (1979). A Narcissistic Personality Inventory. *Psychological Reports*, *45*(2), 590. https://doi.org/10.2466/pr0.1979.45.2.590

Reina, C. S., Zhang, Z., & Peterson, S. J. (2014). CEO Grandiose Narcissism and Firm Performance: The Role of Organizational Identification. *The Leadership Quarterly*, *25*(5), 958–971. https://doi.org/10.1016/j.leaqua.2014.06.004

Rijsenbilt, A., & Commandeur, H. (2012). Narcissus Enters the Courtroom: CEO Narcissism and Fraud. *Journal of Business Ethics*, *117*(2), 413–429. https://doi.org/10.1007/s10551-012-1528-7

Rosenthal, S. A., & Pittinsky, T. L. (2006). Narcissistic Leadership. *The Leadership Quarterly*, *17*(6), 617–633. https://doi.org/10.1016/j.leaqua.2006.10.005

Schrand, C. M., & Zechman, S. L. C. (2012). Executive Overconfidence and the Slippery Slope to Financial Misreporting. *Journal of Accounting and Economics*, *53*(1–2), 311–329. https://doi.org/10.1016/j.jacceco.2011.09.001

Silove, D., Manicavasagar, V., & Pini, S. (2016). Can Separation Anxiety Disorder Escape Its Attachment to Childhood? *World Psychiatry*, *15*(2), 113–115. https://doi.org/10.1002/wps.20336

Tang, Y., Mack, D. Z., & Chen, G. (2018). The Differential Effects of CEO Narcissism and Hubris on Corporate Social Responsibility. *Strategic Management Journal*, *39*(5), 1370–1387. https://doi.org/10.1002/smj.2761

Turner, K., Adigüzel, F., & Sidhu, J. S. (2024). Chief Executive Officer Narcissism, Corporate Inertia, and Securities Analysts' Stock Recommendations. *Strategic Organization*. https://doi.org/10.1177/14761270231218096

Van Scotter, J. R., II. (2019). Narcissism in CEO Research: A Review and Replication of the Archival Approach. *Management Review Quarterly*, *70*(4), 629–674. https://doi.org/10.1007/s11301-019-00178-1

Wales, W. J., Patel, P. C., & Lumpkin, G. T. (2013). In Pursuit of Greatness: CEO Narcissism, Entrepreneurial Orientation, and Firm Performance Variance. *Journal of Management Studies*, *50*(6), 1041–1069. https://doi.org/10.1111/joms.12034

Yakeley, J. (2018). Current Understanding of Narcissism and Narcissistic Personality Disorder. *BJPsych Advances*, *24*(5), 305–315. https://doi.org/10.1192/bja.2018.20

Yook, K.-H., & Lee, S.-Y. (2020). Chief Executive Officer Narcissism and Firm Value: The Mediating Role of Corporate Social Responsibility in the South Korean Context. *Corporate Social Responsibility and Environmental Management*, *27*(4), 1709–1718. https://doi.org/10.1002/csr.1918

You, S., Li, Z., Jia, L., & Cai, Y. (2022). CEO Narcissism and Innovation Ambidexterity: The Moderating Roles of CEO Power and Firm Reputation. *Journal of Product Innovation Management*, *40*(2), 175–194. https://doi.org/10.1111/jpim.12653

Young, S. M., Du, F., Dworkis, K. K., & Olsen, K. J. (2016). It's All about All of Us: The Rise of Narcissism and Its Implications for Management Control System Research. *Journal of Management Accounting Research*, *28*(1), 39–55. https://doi.org/10.2308/jmar-51024

Zhang, H., Ou, A. Y., Tsui, A. S., & Wang, H. (2017). CEO Humility, Narcissism and Firm Innovation: A Paradox Perspective on CEO Traits. *The Leadership Quarterly*, *28*(5), 585–604. https://doi.org/10.1016/j.leaqua.2017.01.003

Zhu, D. H., & Chen, G. (2014a). CEO Narcissism and the Impact of Prior Board Experience on Corporate Strategy. *Administrative Science Quarterly*, *60*(1), 31–65. https://doi.org/10.1177/0001839214554989

Zhu, D. H., & Chen, G. (2014b). Narcissism, Director Selection, and Risk-Taking Spending. *Strategic Management Journal*, *36*(13), 2075–2098. https://doi.org/10.1002/smj.2322

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Economic Policy Uncertainty and Investment: Cross-Country Evidence

Simon E. Liedtke

Humboldt-Universität zu Berlin

Abstract

When investment is partially irreversible, uncertainty raises the option value of waiting and can delay investment. Yet evidence on the effects of economic policy uncertainty (EPU) remains predominantly US-based. This paper examines whether the negative EPU-investment relationship generalizes across countries. Using firm-level data for 25 countries from 1991 to 2024, the analysis estimates fixed-effects regressions of subsequent capital expenditures on the EPU index. A doubling in EPU is associated with a 10.8% decline in quarterly investment and a 14.9% decline in annual investment, on average. The negative association is widespread across countries and is particularly strong among firms with more irreversible investment, consistent with real options theory. However, the relationship varies substantially in magnitude across national settings with a minority of countries even exhibiting significant positive associations. These findings suggest that policy uncertainty is a broadly relevant determinant of corporate investment, while indicating that firms' responses depend on country-specific conditions.

Keywords: economic policy uncertainty; corporate investment; real options theory; cross-country evidence

1 Introduction

Investment decisions present a fundamental paradox: should companies act quickly to seize opportunities, or wait for better information? In a world of perfect foresight, the answer would be straightforward as every project with a positive net present value would be undertaken immediately, with early movers capturing returns before competitors eliminate the opportunity. However, reality presents a completely different landscape. Investment decisions are invariably made under uncertainty, where the probabilities of future outcomes cannot be anticipated or quantified with precision. This is the very distinction between true uncertainty and risk, as defined by Knight (1921). Uncertainty creates a critical tension between early investment, which allows companies to capture potential gains sooner, and the irreversible nature of most investments, meaning that acting quickly can lock companies into costly mistakes. When investments cannot easily be undone or resources recovered, the option value of waiting for more information before committing resources increases (Bernanke, 1983; Dixit & Pindyck, 1994). Real options theory formalizes this intuition by treating investment opportunities like financial options, in which case timing becomes crucial.

This theoretical debate has attracted a large body of empirical research. Most of this research verifies the negative relationship between uncertainty and investment as predicted by real options theory (e.g., Leahy and Whited, 1995; Guiso and Parigi, 1999; Bloom, 2009). In the decade following the financial crises, investigations into policy uncertainty as a specific type of uncertainty have increased, validating the negative association by using elections as real-world experiments (Jens, 2017; Julio & Yook, 2012).

Of the many forms of macroeconomic uncertainty, (economic) policy uncertainty deserves particular attention. Unlike market-driven uncertainties such as commodity price volatility or technological disruption, policy uncertainty originates directly from government actions and institutional processes. This distinction is important because policy uncertainty affects the fundamental economic rules, through tax rates, regulatory frameworks, monetary policy, etc., which influence every aspect of corporate operations. Unlike market-driven uncertainties, policy uncertainty is determined by government actions, meaning its investment-dampening effects are a form of self-imposed economic friction. Baker et al. (2016) developed a widely used index of economic policy uncertainty (hereafter EPU), which exhibits sharp spikes around major policy events.

While previous studies have predominantly focused on individual countries, primarily documenting negative relationships between the EPU index and investment (Kang et al., 2014; Wang et al., 2014; Baker et al., 2016; Gulen and Ion, 2016, 2018; Amore and Minichilli, 2018; Cui, Wang, et al., 2021; Fujitani et al., 2023), recent investigations by Azimli

I would like to express my sincere gratitude to Prof. Ulf Brüggemann and Christian Bernard for their excellent supervision, insightful guidance, and continued support throughout the development of this thesis. Their feedback was invaluable in sharpening the research question and refining the empirical approach. I am also deeply grateful to Prof. Gabriel Ahlfeldt, whose chair not only provided a strong foundation in econometrics, but also fostered the motivation to apply these methods directly in my own research.

DOI: https://doi.org/10.5282/jums/v11i3pp532-550

© The Author(s) 2026. Published by Junior Management Science.
This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.

(2022) and Wu et al. (2020) have found insignificant and even positive associations in other contexts. These diverging findings suggest that the EPU-investment relationship may be contingent on institutional, political, and economic structures that vary significantly across countries, rather than being universal. The dominance of the US and China in existing literature is particularly limiting, given their exceptional positions in the global economy and their unique institutional frameworks, which may not be applicable to other national contexts.

This thesis closes the existing gap by conducting a structured empirical investigation of the relationship between EPU and investment in more than twenty countries. Based directly on the regression framework introduced by Gulen and Ion (2016), the study uses panel regressions to estimate the impact of EPU on leading firm-level capital expenditures (CAPEX), while controlling for standard firm- and country-level variables. Two data panels are constructed to capture both short-term and structural dynamics: a quarterly panel A (1999–2024), which focuses on high-frequency reactions, and an annual panel B (1991–2023), which extends coverage to countries with limited reporting frequencies. Macroeconomic and firm-level controls are included, with careful adjustments made to accommodate international data availability.

The analysis confirms the core patterns observed in previous studies: higher levels of economic policy uncertainty are consistently associated with lower investment levels in most countries. This effect is evident in both the short-run quarterly panel and in the broader annual dataset. While the magnitude of the effect is similar in industrial and emerging markets, there is significant variation between countries. In some economies, such as Spain and Canada, the effect is particularly strong, whereas in others, unclear or even positive associations emerge. This heterogeneity highlights the importance of institutional, structural, and policy-related differences in shaping investment behavior. Furthermore, the irreversibility channel remains central as firms with higher capital rigidity exhibit significantly greater sensitivity to policy uncertainty.

This study contributes to the literature by providing the first comprehensive cross-country mapping of the EPU-investment relationship, encompassing over twenty countries. This gives researchers both empirical insights and a roadmap showing where future investigations into the institutional determinants of EPU's heterogeneous investment effects are needed. The findings also provide policymakers with concrete evidence of how governmental policy uncertainty manifests itself in corporate investment postponement. This emphasizes the vital role of transparent and credible policy communication in fostering investment confidence.

The thesis proceeds as follows. The next section provides a literature review and details the development of the hypothesis. Section three describes the data, variables and methodology, while section four presents the results. Section five discusses these results and section six concludes.

2 Literature Review and Hypothesis Development

The interplay between uncertainty and corporate investment has long intrigued economists, resulting in a wealth of theoretical and empirical research. Wang et al. (2014) point out that already Knight (1921) was among the first to suggest that uncertainty lies at the heart of entrepreneurial activity and therefore of investing activity. Early theoretical models, such as those by Oi (1961), Hartman (1972; 1976) and Abel (1983), also propose a positive relationship under specific conditions. This is sometimes referred to as growth theory. These models argue that, in perfectly competitive markets with convex adjustment costs, heightened uncertainty increases the expected profitability of capital, thereby encouraging investment. Stiglitz and Weiss (1981) suggest that uncertainty, coupled with asymmetric information, makes risky projects more attractive to firms. Bar-Ilan and Strange (1996) contend that investment lags under uncertainty can incentivize investment. Caballero (1991) argues that imperfect competition can bring the relationship "closer to the results of Hartman and of Abel" but decreasing returns to scale make investment deterrence in response to uncertainty more likely. He emphasizes the lack of robustness in the sign of the uncertainty-investment relationship. Real options theory posits that integrating investment irreversibility in the model, a value of waiting under uncertainty to gain additional information and avoid sunk costs gets created (Bernanke, 1983; Dixit and Pindyck, 1994, among others). Rodrik (1991) amplifies the suggestions made for a negative sign. Leahy and Whited (1995) and Guiso and Parigi (1999) provide early empirical support for this irreversibility channel for US and Italian manufacturing firms, respectively. They find that uncertainty dampens investment, particularly for firms with less reversible capital. Bloom (2009) finds that firms pause investments in times of higher uncertainty proxied by stock market volatility. More recent empirical evidence strongly supports this perspective in the case of policy uncertainty. Julio and Yook (2012) and Jens (2017) document reduced investment during election years, with even bigger influence in close elections.

Of the various measures of uncertainty to test these theories, EPU has emerged as particularly influential. Baker et al. (2016) provide an economic policy uncertainty measure through the EPU index. They track the monthly frequency of newspaper articles that mention a combination of "uncertainty" terms (e.g. "uncertain," "ambiguity"), "policy" terms (e.g. "regulation," "fiscal"), and economic references (e.g. "economy," "market"). Counts from leading national newspapers are normalized by total article volume and averaged across sources, then standardized to create a comparable time series. As demonstrated in their original work, the index reliably spikes during major episodes of policy turmoil, such as the oil price shocks of the 1980s, the global financial crisis, and the Covid-19 pandemic. The global "GEPU" index, developed by S. Davis (2016), most recently reached record highs in April 2025, calling for immediate investigation after trade disputes and geopolitical tensions. Thanks to the dedication

of many scholars, the EPU index is now published in over 30 countries and publicly available on the EPU index website.[1] In some countries like the Netherlands or China, multiple indices are available. The specific indices used for this analysis are specified in the data section. The methodology slightly differs across countries, but all indices rely mainly on the count of newspaper articles. Detailed construction methods are specified on the website and in the respective papers, which are all mentioned in the references. Throughout the text Baker et al. (2016) are consistently referenced given their pioneering contribution to EPU measurement methodology, which has formed the basis for subsequent scholarly developments in this field.

The EPU index quickly drew a lot of attention to its specific impact on investment. Kang et al. (2014) and Gulen and Ion (2016) found a consistent negative relationship between EPU and investment in the US. Wang et al. (2014) and An et al. (2016) replicated those findings for China. Baker et al. (2016) extended the research onto major economies of Canada, France, Germany, India, Italy, Japan, Spain and the UK. However, they do not account for the high share of US data in their analysis, and their findings could potentially just replicate the US findings. Amore and Minichilli (2018) and Fujitani et al. (2023) reinforced prior findings in Italy and Japan, respectively. Gulen and Ion (2016) demonstrate that the relationship disappears after two to three years and is followed by a rebound period with increased investment. **Gulen.2018<empty citation>** demonstrate that higher EPU is associated with reduced merger and acquisition activity, which is an extreme form of investment. Based on this theoretical and empirical foundation, the following hypothesis is formulated:

> **Hypothesis 1 (H1)**: *EPU has a negative effect on corporate investment in the entire sample.*

The effect of the EPU in emerging markets is more nuanced, warranting further investigation. Rottig (2016) provides an in-depth discussion of the differences in institutional features in emerging markets and their implications for corporate behavior, which could suggest different investment behavior in developing countries. Azimli (2022) found no significant link between EPU and investment in the Czech Republic, Hungary, Poland, Russia, and Turkey hypothesizing that the effects predicted by real options theory and growth theory cancel each other out in emerging markets. Meanwhile, Chen et al. (2020) replicate the findings of Gulen and Ion (2016) in Australia, but Wu et al. (2020) identify a positive relationship between EPU and investment in the same country and year. Azqueta-Gavaldón et al. (2023) find that the effect of EPU on investment often depends on the specific type of uncertainty, which differs across major EU nations. Exploring the heterogeneity of countries and whether emerging markets exhibit distinct responses to EPU compared to developed economies is thus a critical step forward. The following two hypotheses are therefore proposed:

> **Hypothesis 2 (H2)**: *The effect of EPU on corporate investment is weaker in the emerging markets.*

> **Hypothesis 3 (H3):** *The relationship between EPU and investment varies across countries.*

A key factor shaping the relationship between EPU and investment seems to be the irreversible nature, which is a central tenet of the real options framework. Bernanke (1983) and Dixit and Pindyck (1994), among others, theorize that when investments involve high sunk costs, uncertainty intensifies the incentive to postpone, since firms cannot easily reverse their decisions. Leahy and Whited (1995) and Guiso and Parigi (1999) were among the first to propose investment irreversibility as the primary channel through which uncertainty and investment are linked. Wang et al. (2014) and Gulen and Ion (2016) show that firms with fewer redeployable assets (e.g., those with higher PPE ratios) are more sensitive to EPU. Similarly, Julio and Yook (2012), Baker et al. (2016), Jens (2017), and Dibiasi et al. (2018) find that irreversibility exacerbates the negative impact of uncertainty, and Dibiasi et al. (2018) specify the positive effect of reversibility on the other hand. This study uses the PPE ratio as a proxy for irreversibility to test whether this mechanism applies to a global sample, providing a broader perspective on its moderating role. Accordingly, the following hypothesis is proposed:

> **Hypothesis 4 (H4)**: *Higher investment irreversibility strengthens the negative effect of EPU on investment.*

These hypotheses provide an analytical framework for examining the global impact of EPU on investment, paying particular attention to variations in market type, cross-country heterogeneity, and the role of investment irreversibility. To systematically test these propositions, this study employs a consistent econometric approach in a quarterly and annual panel, utilizing multiple sample splits to rigorously investigate hypotheses H1 through H4 based on both theoretical predictions and empirical evidence from the literature.

3 Data and Methodology

3.1 Data

This study combines firm-level investment data with measures of macroeconomic uncertainty and economic prosperity. It utilizes the EPU index from Baker et al. (2016), LSEG's Worldscope fundamentals database, and publicly available OECD data. Following Gulen and Ion (2016), firm-quarter observations are employed to maximize the detection of fluctuations in the EPU index and to capture short-term variations. However, quarterly financial statement data is not applicable in every country, as major economies such as the UK, Australia, and India do not mandate quarterly reporting, leading to uninformative findings for these jurisdictions.

[1] Economic Policy Uncertainty Index

To address this issue, a complementary analysis based on annual data is conducted. To the author's knowledge, this dual-frequency approach is a novel contribution to the existing literature.

The EPU index covers 31 countries. The United States and Canada have the longest time series, dating back to 1985, while Nigeria has the shortest, beginning in 2016. While most countries' indices are up to date, updates ceased for the Netherlands and Denmark in 2021, Sweden in 2022, and Mexico, Nigeria, and New Zealand in 2023. Data for Colombia and Turkey have not been updated for 2025. The analysis uses a final EPU dataset derived from the "All Country-Level Data" compilation. Missing countries were manually incorporated due to incomplete integration in the global database. Where multiple indices exist for a country, specific selections were made. In China, the index based on two mainland newspapers is used, as Hong Kong is covered by its own index. For the Netherlands, the EBO-NL index, which filters more strictly for economic policy uncertainty than general policy uncertainty, is adopted. In Denmark, the extended index covers more policy-related terms. For Colombia and Mexico, the indices starting in 1990 and 1996 respectively are taken to analyze a longer time horizon. Table 1 summarizes the resulting sample for each country. Since the EPU index is reported monthly while firm-level data are quarterly or annual, the former is aggregated to the latter's frequency using the arithmetic mean and is then log-transformed via its natural logarithm to address extreme outliers.

Data on firm-level investment and control data are sourced exclusively from LSEG's Worldscope database, accessed via Wharton Research Data Services. The quarterly dataset begins in 1990, although observations for each quarter prior to 1998 are scarce. The annual dataset ranges back to 1985, but there are few observations prior to 1990. Consequently, panel A (quarterly) covers the period from 1999 to 2024, and panel B (annual) covers the period from 1991 to 2023. The integration of the growth variables that span over a year causes the drop of the first year in both panels and the applied lead structure causes the drop of 2024 in panel B since annual financial statements for 2025 have not yet been published. The number of observations per quarter and year is visualized in Figure 1, showing the growing sample size in time in both panels. The drop in firm-quarter observations between 2011 and 2013 in panel A is caused by reduced availability of depreciation, amortization and depletion in Japan.[2]

Identical filtering procedures are applied to both panels. Firms are filtered by country codes corresponding to their primary listing. In line with previous investment literature, financial firms with Industry Group values between 4300 and 4400 are excluded (Azimli, 2022; Fujitani et al., 2023; Gulen & Ion, 2016; Jens, 2017; Julio & Yook, 2012; Kang et al., 2014; Wang et al., 2014). Cross-listed companies with mismatched country codes, which refer to primary listing and nation names, which refer to headquarters, are disregarded due to ambiguous country classification. To prevent bias from very small firms, the bottom 1st percentile of total assets (in $) is filtered out, with thresholds of $5,674 in panel A and $9,842 in panel B. All variables must be non-missing for each observation and firms must have at least twelve consecutive quarters in panel A or three consecutive years in panel B of information (Gulen & Ion, 2016). Firms with varying fiscal year-ends may experience misalignment in measuring economic policy uncertainty and lead structure with the annual panel facing greater misalignment due to fiscal years spanning different 12-month periods, potentially shifting variable timing by several months. However, this variation is minimal and neglected, as over 90% of firms' fiscal quarters align with calendar quarters, ensuring consistent timing for analyses. Table 2 summarizes the panel structures after each filtering step, where the last columns present the final panels.

Capital expenditures (CAPEX) serve as a proxy for firm-level investment (e.g. Julio and Yook, 2012; Wang et al., 2014; Baker et al., 2016; Gulen and Ion, 2016; Jens, 2017; Azimli, 2022). It is calculated as the period-on-period change in property, plant, and equipment (PPE) over a given period, plus depreciation, amortization, and depletion, scaled by the total assets at the beginning of the period. However, this measure is an approximation, as the inclusion of amortization accounts for impairments of intangible assets that do not correspond to PPE. Although LSEG provides a direct CAPEX item, the limited availability of data across many countries on a quarterly basis would substantially reduce the size of the panel, undermining the cross-country design. Thus, despite its approximate nature, an indirect method leveraging basic balance sheet and income statement items, instead of cash flow statements and appendix information, is adopted. Coverage on an annual basis is much better. However, the same indirect calculation method is used in panel B to ensure comparability between the two panels. The correlation between the indirect CAPEX measure and LSEG's direct CAPEX is greater than 0.66 (Pearson) and 0.78 (Spearman) in panel A, and greater than 0.83 (Pearson and Spearman) in panel B. It is systematically higher because of the addition of amortization. This validates the approximation despite the loss of observations and comparability across panels.

Standard control variables for firm-level investment opportunities, as defined by Gulen and Ion (2016), have been incorporated. Operating cash flows (OCF) are computed via the indirect method: net income plus depreciation, amortization, and depletion, minus the change in working capital over the period, scaled by total assets at the beginning of the period. The choice of LSEG's directly provided item from the statement of cash flows mirrors the rationale for the CAPEX calculation. Working capital is defined as the difference between current assets and current liabilities. Tobin's Q (TobinQ) is determined as the sum of market capitalization and total liabilities divided by total assets. Annual sales growth (dSales) is calculated manually as a decimal. The PPE ratio, defined as PPE divided by total assets, serves as a proxy for investment irreversibility, consistent with prior studies (e.g. Gulen and Ion, 2016; Azimli, 2022).

[2] Regression results remain robust without Japan in the samples.

Table 1: Country Overview

Country	N - A	N - B	Mean - A	Mean - B	HDI	Classification
Australia	0	26,900	N/A	0.100	0.915	Industrial
Belgium	726	1,812	0.019	0.067	0.907	Industrial
Brazil	16,333	5,276	0.016	0.069	0.721	Emerging
Canada	42,295	15,767	0.024	0.117	0.908	Industrial
Chile	9,520	2,592	0.017	0.069	0.803	Industrial
China	2,755	58,224	0.014	0.064	0.662	Emerging
Colombia	790	402	0.018	0.074	0.714	Emerging
Denmark	3,326	2,888	0.017	0.070	0.912	Industrial
France	1,184	16,664	0.015	0.066	0.873	Industrial
Germany	18,803	14,601	0.016	0.071	0.916	Industrial
Greece	5,948	4,825	0.010	0.064	0.850	Industrial
India	788	27,508	0.013	0.048	0.560	Emerging
Ireland	352	1,104	0.014	0.077	0.886	Industrial
Italy	6,825	5,252	0.015	0.069	0.866	Industrial
Japan	198,680	95,556	0.010	0.040	0.898	Industrial
Korea (South)	92,011	35,722	0.013	0.063	0.861	Industrial
Mexico	6,607	1,410	0.019	0.074	0.735	Emerging
Netherlands	1,657	1,492	0.015	0.064	0.914	Industrial
New Zealand	0	2,017	N/A	0.084	0.903	Industrial
Poland	23,236	6,814	0.016	0.071	0.826	Industrial
Spain	2,684	2,864	0.015	0.073	0.855	Industrial
Sweden	21,557	7,210	0.018	0.079	0.911	Industrial
Turkey	15,867	3,811	0.024	0.138	0.726	Emerging
United Kingdom	1,317	24,124	0.011	0.073	0.893	Industrial
United States	342,439	109,537	0.017	0.080	0.909	Industrial
Total	815,700	474,372	0.015	0.068		

Notes: This table reports the number of firm-quarter (A) observations and firm-year (B) observations for each country in the final samples derived from LSEG's Worldscope database via WRDS. The sample includes all countries with available EPU data from Baker et al. (2016) after all filtering steps applied. The means refer to the average CAPEX in both panels and are computed as the arithmetic means. HDI is sourced from United Nations data, with countries classified as emerging (HDI < 0.8) or industrialized (HDI ≥ 0.8) per United Nations Development Programme (2020) for their average HDI from 1990 to 2023.

Table 2: Sample Construction Summary

Panel A: Quarterly Data						
	Original sample	Financials	Cross-listed	Bottom 1% assets	Missing values	<12 quarters
Firm-Quarters	3,054,840	2,747,227	2,545,569	1,864,112	882,296	815,700
Firms	64,997	58,825	53,808	50,225	26,289	17,674
Quarters	144	144	144	112	104	104
Panel B: Annual Data						
	Original sample	Financials	Cross-listed	Bottom 1% assets	Missing values	<3 years
Firm-Years	1,375,192	1,213,598	1,106,734	828,014	486,191	474,372
Firms	85,586	75,935	68,987	63,096	45,905	38,345
Years	37	37	37	36	33	33

Notes: The original sample is defined through LSEG's quarterly/annual fundamentals database filtered for countries with the EPU index available from Baker et al. (2016) through the 'Nation Code'. The quarterly database is available since 1990 and the annual since 1985. All steps in the annual and quarterly data cleaning process are equal. Following the literature financial companies are filtered out defined by 'Industry Group' values between 4,300 and 4,400. Cross-listed observations identified through deviating country codes tracking primary listing and Nation names tracking headquarters are also removed. The bottom 1% of total assets in $ are filtered out. The threshold for the quarterly panel is $5,674 and for the annual $9,842. All variables except for the leads 2,3 and 4 in CAPEX need to be available. In the last step firms are required to have at least twelve consecutive quarters or three consecutive years in the annual panel of these variables available resulting in the final panels A and B in the last column.

Additional controls for economic investment opportunities include real GDP growth (dGDP_Q/dGDP_Y) and the Consumer Confidence Index (CCI), as outlined by Gulen and Ion (2016). Data for both variables were retrieved from the OECD Data Explorer, with GDP growth provided in decimal format at quarterly and annual frequencies, and monthly CCI data aggregated to quarterly and annual levels via its arithmetic mean. Canadian CCI data is manually sourced from the Federal Reserve Economic Data (FRED). Unlike Gulen and Ion (2016), who used forecasted GDP growth, this study uses realized GDP growth due to the unavailability of historic forecasts for non-OECD countries in the database. Croatia, Hong

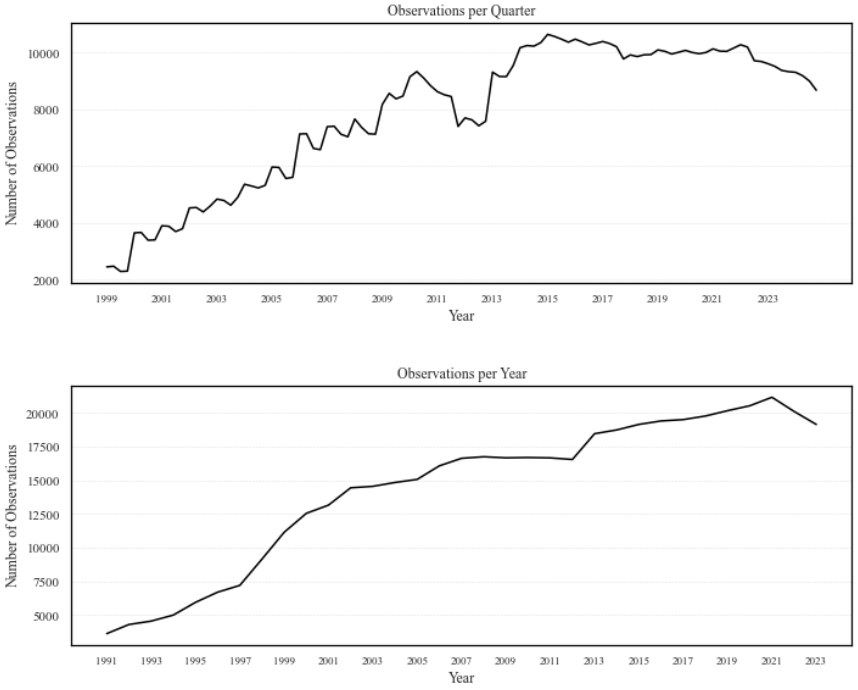


Figure 1: Observations over Time

Notes: Observations per quarter and year in the final panel A (above) and B (below) over time. Note that the first year in both panels is not included due to the integration of growth variables. The years correspond to the independent variables and end in 2023 in the observations per year because of missing annual financial statements for 2025.

Kong, Nigeria, Pakistan, and Singapore are not reported in either database and Russia lacks CCI data. Gulen and Ion (2016) show that these control variables have significant influence on investment. Therefore, the countries are not further considered in the analysis. This leaves 23 and 25 countries in panels A and B, respectively, as Australia and New Zealand fall out of panel A.

Following Gulen and Ion (2016), controls for overall economic uncertainty include profit margin growth volatility (profit_sd) and stock return volatility (stock_sd). These are calculated cross-sectionally as the standard deviation of the growth in the respective measure from one period to the next. To exclude the influence of newly listed firms, only firms with over 80 quarters of data (20 years in panel B) are included in these calculations (Bloom, 2009; Gulen & Ion, 2016). As both variables exhibit significant outliers and are strictly positive, they are log-transformed using their natural logarithm. While this deviates from the approach adopted by Gulen and Ion (2016), the log transformation addresses the considerable right-skewness in these variables and mitigates the influence of extreme outliers in the regression. This appears to be a greater concern in the global sample than in the US. All variables are winsorized at the 1st and 99th percentiles and standardized by their sample standard deviation. Table 3 presents descriptive statistics of all variables in both panels.

For the comparison between emerging and industrial markets, the United Nations' Human Development Index (HDI) is employed, with values above 0.8 classified as very highly developed countries by United Nations Development Programme (2020) here defined as industrial markets. Given the study's time horizon over two decades, the HDI in 2023 alone does not reflect historical development trends; thus, the mean HDI per country from 1990 to 2023 is used for classification, with results reported in Table 1, alongside the number of observations and the average investment rate per country in each panel. The average CAPEX differs significantly across countries ranging from only 1% in Greece to 2.4% in Canada and Turkey in panel A and from 4.8% in India to 13.8% in Turkey in panel B. Panel A has more total observations despite including fewer years and countries than panel B. It also allows for shorter investigations of investment behavior in response to increasing EPU and replicates the design of Gulen and Ion (2016). Nevertheless, panel B offers a more balanced sample. The share of observations from the US and other major economies, such as Japan and South Korea is reduced from 42%, 24% and 11% in panel A to 23%, 20% and 8% in panel B, respectively. Furthermore, the higher coverage of firms in panel B for countries without mandatory quarterly reporting extends the analysis to several countries, such as Australia and the UK. As shown in Table 2, it also includes more firms, which could indicate better representation within countries.

Table 3: Descriptive Statistics

Panel A: Quarterly Data								
	Count	Mean	Std	Min	25%	50%	75%	Max
CAPEX_lead1	815,700	0.015	0.029	-0.070	0.003	0.009	0.019	0.179
EPU_log	815,700	4.838	0.398	4.019	4.572	4.817	5.093	5.996
TobinQ	815,700	1.490	2.028	0.097	0.512	0.882	1.593	14.296
OCF	815,700	-0.002	0.115	-0.762	-0.004	0.012	0.031	0.248
dSales	815,700	0.161	0.658	-0.848	-0.061	0.051	0.196	4.732
dGDP_Q	815,700	0.005	0.012	-0.062	0.001	0.006	0.009	0.053
CCI	815,700	99.724	1.465	95.831	98.746	99.901	100.759	102.797
stock_sd_log	815,700	-0.529	1.537	-2.199	-1.583	-0.999	-0.085	5.716
profit_sd_log	815,700	1.976	2.892	-2.269	-0.829	2.673	4.040	9.211
Panel B: Annual Data								
	Count	Mean	Std	Min	25%	50%	75%	Max
CAPEX_lead1	474,372	0.068	0.123	-0.234	0.013	0.039	0.084	0.768
EPU_log	474,372	4.787	0.456	3.868	4.472	4.755	5.016	5.967
TobinQ	474,372	1.434	1.952	0.104	0.548	0.838	1.481	13.869
OCF	474,372	-0.043	0.456	-3.178	-0.018	0.047	0.108	0.512
dSales	474,372	0.208	0.863	-0.89	-0.047	0.059	0.207	6.686
dGDP_Y	474,372	0.029	0.031	-0.059	0.014	0.027	0.044	0.114
CCI	474,372	99.934	1.733	94.743	98.867	100.05	100.99	103.32
stock_sd_log	474,372	0.489	1.577	-1.501	-0.593	0.039	1.240	6.262
profit_sd_log	474,372	1.900	3.077	-3.022	-0.864	1.821	4.341	11.114

Notes: Observations are firm-quarters in panel A and firm-years in panel B. All firm-level variables are obtained from LSEG's Worldscope quarterly and annual fundamentals databases accessed through WRDS on 10.07.2025. The log-transformed EPU index from Baker et al. (2016) is from the "all-country data", with countries not integrated but individually available, added manually and all values aggregated to the quarter or year. Quarterly and annual real GDP growth and monthly CCI data are from OECD databases. Quarterly and annual CCI data is aggregated by the mean. The uncertainty proxies represent the respective period-on-period log-transformed cross-sectional standard deviations in profit margin growth and stock returns in each panel. The selection of control variables follows Gulen and Ion (2016). Panel A includes 23 countries from 1999 to 2024 and panel B 25 countries from 1991 to 2023. CAPEX_lead1 leads all other variables by one period (quarter or year). All variables are winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions.

3.2 Methodology

This study builds on the regression framework of Gulen and Ion (2016) assessing its applicability within a global sample of quarterly and annual financial statements using a dual panel approach. Having presented their baseline model, the authors introduce controls for macroeconomic investment opportunities and uncertainty to address an omitted variable bias. The cross-country design is constrained by the limited availability of certain variables on a global scale, necessitating a selective inclusion of those controls. Nonetheless, all firm-level controls and the most statistically significant macroeconomic controls are retained. The results in the US closely align with their results from the US, demonstrating the inclusion of the most important variables.

Both panels apply consistent filters: removing financial firms, cross-listed entities with unaligned country codes to nation names, and firms lacking at least twelve consecutive quarters (three years) of observations. Panel A spans 1999-2024, emphasizing short-run reactions, while panel B covers 1991–2023, broadening country coverage and capturing longer-term adjustments. The model is specified as:

$$\text{Investment}_{i,t+l} = \alpha_{i,t} + \beta_1 \,\text{EPU}_{c,t} + \beta_2 X_{i,t} + \beta_3 Z_{c,t} + \beta_4 Q_{c,t} + \varepsilon_{i,t}$$

In this equation, (i) and (c) index the firm and country, respectively, (t) denotes the period (year or year-quarter), and ($l = [1,2,3,4]$) represents the lead of the dependent variable, CAPEX, which serves as the proxy for investment. The use of leads for the dependent variable is a standard practice in empirical investment literature to mitigate mechanical correlations between dependent and independent variables (Wang et al., 2014; Gulen and Ion, 2016; Fujitani et al., 2023, among others). Up to four leads are analyzed to replicate the design from Gulen and Ion (2016) and analyze the short-term horizon up to one year. Analyzing leads up to four years ensures insights into the duration of the relationship and the rebound period suggested by Gulen and Ion (2016) after two to three years. However, the lead structure restricts the 2024 analysis in the annual panel, as CAPEX data for 2025 is not yet available. The primary independent variable is the log-transformed Economic Policy Uncertainty (EPU) index from Baker et al. (2016). The log-transformation reduces the influence of extreme outliers and facilitates coefficient interpretation in relative increases. Vectors X and Z encompass firm-level and macroeconomic control variables, respectively, designed to account for investment opportunities and overall economic uncertainty. X includes operating cash flow (OCF) as net income minus the change in working capital plus depreciation, amortization and depletion scaled by beginning of the period total assets, Tobin's Q (TobinQ) as the sum of market capitalization and total liabilities divided by total assets and year-on-year sales growth (dSales) expressed as a decimal. Z includes real period-on-period GDP

growth (dGDP_Q/dGDP_Y), the Consumer Confidence Index (CCI) as investment opportunity proxies and the standard deviations of period-on-period profit margin growth (profit_sd) and stock returns (stock_sd). That is only a selection of the macro controls used by Gulen and Ion (2016) and the used variables slightly deviate from theirs to fit the global sample better. First, real GDP growth is used instead of forecasted GDP growth, because of data unavailability in many countries. That leads to a more backward-looking variable than an opportunity variable. Second, stock returns are calculated by period-on-period changes in market capitalization from the Worldscope database instead of daily stock returns. This measure captures only part of stock market dynamics but is used due to global data constraints. Especially in panel B, that variable needs to be examined under caution as it uses year-on-year stock returns. Lastly, both volatility measures are log-transformed to handle outliers. For the quarterly panel, quarter dummies (Q) are included to address seasonality, while standard errors (ε) are clustered at both the firm and time levels to account for potential cross-sectional and serial correlation (Gulen & Ion, 2016). Firm fixed effects (α) are incorporated to control for unobserved heterogeneity, a practice consistent with prior studies (Fujitani et al., 2023; Gulen & Ion, 2016; Jens, 2017; Julio & Yook, 2012; Wang et al., 2014). Numerous studies examine the relationship between EPU, investment and firm characteristics in detail (e.g. Kang et al., 2014; Wang et al., 2014; Gulen and Ion, 2016, Amore and Minichilli, 2018). This study aims to provide an overview of countries where further investigation is needed, which is why only the investment irreversibility channel is explored. Given the absence of further firm-level conditioning, the integration of firm fixed effects is especially important. Following the rationale of Gulen and Ion (2016), time fixed effects are excluded as they would absorb the variation of the EPU index.

This regression model is applied uniformly across all analyses, with variations limited to the composition of the samples. To address H1, the relationship between EPU and corporate investment is examined initially in the full panel. Due to the disproportionate influence of the US, particularly in panel A, the analysis is repeated without the US. A subsequent analysis partitions the panels into industrial and emerging markets to examine H2. The classification is done by the HDI index and summarized in Table 1. The central analysis involves country-specific regressions addressing H3. To explore the irreversibility channel in H4, both panels are split by the sample median of the PPE ratio and analyzed separately for each country. A median split is preferred over an interaction term to avoid potential mechanical correlations between CAPEX, which depends heavily on PPE and total assets, and the PPE to total assets ratio itself.

4 Results

4.1 Panel A

This subsection evaluates the empirical findings from quarterly (1999–2024, 23 countries) panel regressions, testing the hypotheses outlined in Section 2 regarding the impact of Economic Policy Uncertainty (EPU) on firm-level capital expenditures (CAPEX). The analysis considers time trends, correlations, and regression outcomes, emphasizing cross-country heterogeneity and investment irreversibility. Figure 2 demonstrates a clear upward trend in the average EPU since 1999, reflecting increasing policy uncertainty or the heightened media reporting of it. In contrast, CAPEX follows a cyclical trajectory, peaking at 3.29% in Q1 2019. While descriptive, these divergent patterns are consistent with a possible inverse relationship, providing preliminary support for H1. Nevertheless, trend-based inferences must be interpreted with caution. Table 4 presents the correlation matrix of the independent variables, showing Pearson correlations in the lower triangle and Spearman correlations in the upper triangle. EPU_log is negatively associated with all proxies for investment opportunities at both the firm and macro levels, and positively correlated with profit margin volatility, though it shows a slightly negative correlation with stock market volatility. The non-parametric Spearman results confirm the robustness of these patterns, except for stock_sd, which is positively correlated under Spearman. These results support the expectation that higher policy uncertainty is linked to weaker investment conditions. It also suggests low connection to overall economic uncertainty. The low collinearity across independent variables enhances the interpretability of regression coefficients and confirms the model's suitability. The cross-country design also benefits from variation in the EPU across countries because it provides greater variation in the main explanatory variable. Correlations of the country-specific EPU indices (detailed in Table 5) confirm that, despite the co-movement through substantial global shocks, the cross-national panel displays enough diversity. Notably, many indices are highly correlated with the US index.

Table 6 reports the findings of the full panel regression and a robustness test omitting US data, which comprises 42% of the sample. Across all four quarterly leads, EPU_log shows a robust and statistically significant negative association with CAPEX. The non-US sample highlights a slightly greater effect, especially in the fourth quarter. That confirms that results are not driven solely by US dynamics, an important improvement over past studies heavily reliant on US data. A one-standard-deviation increase in the EPU index is associated with a 0.032 standard deviation decline in CAPEX in the full sample and 0.034 decline in the non-US sample within the subsequent quarter. This implies that a doubling of EPU is associated with a 16.2 basis point reduction in CAPEX in the full sample and a 17.2 basis point reduction in the non-US sample, corresponding to 10.8% and 12.29% of the respective average investment rates. Crucially, historical EPU spikes underscore the real-world relevance of these effects: the index tripled during the global financial crisis and doubled during the COVID-19 pandemic. These magnitudes imply that a meaningful share of the observed decline in firm-level investment (Figure 2) during such episodes can be attributed to elevated economic policy uncertainty.

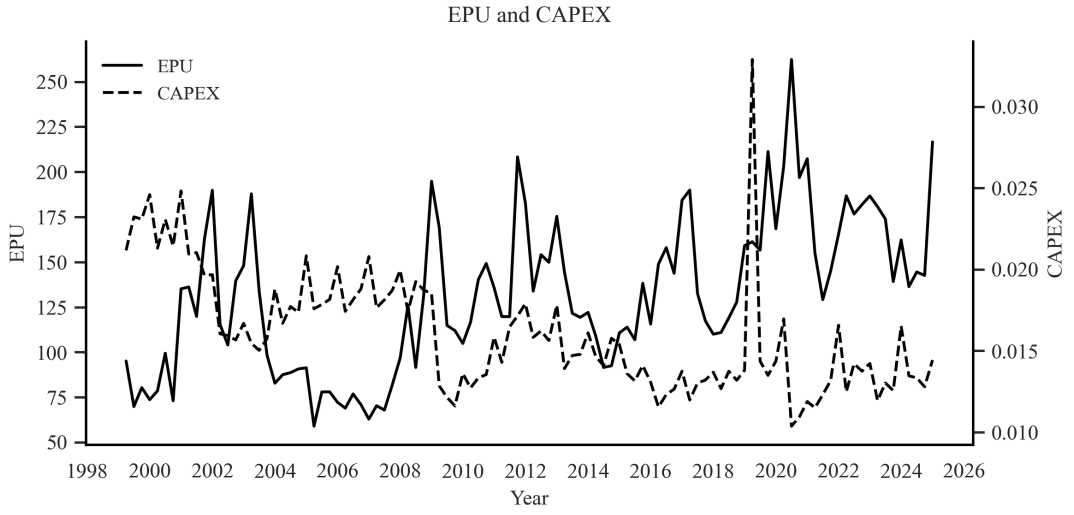


Figure 2: Time Series of Quarterly average EPU and CAPEX

Notes: This figure displays the average EPU index from Baker et al. (2016) and the average capital expenditures by lagged total assets (dotted) per quarter in the entire panel A. Time series graph from panel B is available in Appendix B.

Table 4: Correlation Matrix: Variables – Panel A

	EPU_log	TobinQ	OCF	dSales	dGD_Q	CCI	stock_sd_log	profit_sd_log	PPE_ratio
EPU_log	1	0.008	-0.009	-0.035	-0.075	-0.290	0.014	0.093	0.014
TobinQ	-0.007	1	-0.026	0.113	0.070	0.112	0.550	0.581	-0.114
OCF	-0.007	-0.313	1	0.054	0.003	0.018	-0.047	-0.025	0.168
dSales	-0.005	0.128	-0.073	1	0.039	0.045	0.011	0.057	-0.040
dGDP_Q	-0.117	0.036	-0.004	0.032	1	0.213	0.145	0.101	0.008
CCI	-0.288	0.061	-0.009	0.024	0.245	1	0.147	-0.009	0.013
stock_sd_log	-0.092	0.28	-0.081	0.038	0.078	0.155	1	0.605	-0.045
profit_sd_log	0.049	0.345	-0.091	0.071	0.089	0.032	0.528	1	-0.057
PPE_ratio	0.018	-0.131	0.112	-0.036	0.004	0.016	-0.016	0.006	1

Notes: Pairwise Pearson (lower triangle) and Spearman (upper triangle) correlations between all independent variables in panel A. All variables are winsorized at the 1st and 99th percentiles. See variable definitions in Appendix A and panel B's correlation matrix in Appendix B.

The regression coefficients of the control variables have the expected signs. All firm-level proxies for investment opportunities are positively associated with CAPEX in the full and without US samples for all leads. Only at the three-quarter horizon does OCF not show a significant coefficient. Macro controls, for both economic investment opportunities and uncertainty generally exert little influence, though CCI can be a positive predictor of CAPEX two and three quarters ahead (full sample), while profit volatility is negatively linked to investment in the non-US sample. Real GDP growth exerts little influence on future corporate investment. These results affirm that firm-level factors primarily drive investment, yet that EPU stands out among macro determinants. Worth noting, the model's R-squared is consistently higher when US data is included, suggesting greater explanatory power for US firms than for those from other countries. Together, these findings provide robust empirical support for H1, confirming a persistent negative link between economic policy uncertainty and global capital investment.

To address H2, the regression framework is applied separately to industrial and emerging markets, as specified in Table 7. These are split by the arithmetic mean HDI of 0.8 as defined by United Nations Development Programme (2020) from 1990 to 2023 (overview in Table 1). However, emerging markets account for only 5.3% of firm-quarter observations, which necessarily limits statistical power and robustness in that group. The results in industrial markets mirror those of the full sample almost exactly, with the EPU_log effect remaining negative and statistically significant, and the firm-level determinants positive across all projection horizons, except for OCF. Differently, CCI is now strictly positively associated with future investment in industrial markets. In the emerging markets the EPU_log coefficient is statistically weaker, but economically stronger than for industrial markets, especially at one and four quarters ahead: a doubling in EPU corresponds to a 45.9 bp drop (from a mean CAPEX rate of 1.9%, i.e., a 23.9% decrease) versus 13.1 bp (8.85% reduction of a 1.48% average investment rate) in industrial

Table 5: Correlation Matrix: EPU across Countries – Panel A

	BEL	BRA	CAN	CHL	CHN	COL	DNK	FRA	DEU	GRC	IND	IRE	ITA	JPN	KOR	MEX	NLD	POL	ESP	SWE	TUR	GBR	USA
BEL	1	0.32	0.63	0.55	0.35	0.52	0.59	0.57	0.65	0.02	0.38	0.70	0.52	0.59	0.61	0.00	0.58	0.59	0.74	0.42	0.15	0.34	0.64
BRA	0.28	1	0.54	0.63	0.21	0.67	0.64	0.62	0.65	0.09	-0.13	0.57	0.26	0.35	0.47	-0.03	0.01	0.60	0.47	0.56	-0.25	0.57	0.48
CAN	0.64	0.54	1	0.62	0.66	0.62	0.83	0.89	0.88	0.72	-0.17	0.83	0.62	0.70	0.68	-0.29	0.42	0.59	0.79	0.79	0.08	0.90	0.78
CHL	0.58	0.56	0.61	1	0.55	0.82	0.69	0.72	0.81	-0.10	0.06	0.73	0.40	0.37	0.67	0.14	0.30	0.76	0.69	0.75	-0.05	0.53	0.65
CHN	0.41	0.20	0.61	0.59	1	0.67	0.54	0.45	0.65	-0.53	-0.13	0.61	0.07	0.26	0.66	0.67	-0.04	0.77	0.34	0.25	-0.07	0.39	0.70
COL	0.52	0.67	0.65	0.80	0.68	1	0.7	0.63	0.85	-0.27	0.05	0.76	0.08	0.29	0.72	0.41	-0.02	0.82	0.49	0.59	-0.22	0.41	0.58
DNK	0.66	0.57	0.82	0.72	0.62	0.72	1	0.81	0.84	0.24	0.03	0.77	0.61	0.69	0.64	-0.04	0.31	0.68	0.78	0.70	-0.05	0.75	0.75
FRA	0.58	0.58	0.89	0.68	0.42	0.64	0.74	1	0.85	0.19	0.17	0.78	0.59	0.56	0.74	-0.06	0.27	0.71	0.79	0.74	-0.09	0.81	0.74
DEU	0.66	0.58	0.88	0.83	0.65	0.84	0.82	0.82	1	0	0.07	0.85	0.45	0.53	0.82	-0.01	0.35	0.83	0.80	0.78	-0.02	0.74	0.77
GRC	-0.02	0.10	0.70	-0.19	-0.55	-0.23	0.10	0.25	-0.03	1	0.21	-0.08	0.29	0.32	0.08	-0.09	0.20	-0.23	0.04	0.18	0.05	0.23	0.04
IND	0.38	-0.21	-0.09	0.02	-0.19	-0.09	0.02	0.17	0.08	0.23	1	0.09	0.24	0.14	0.16	-0.04	0.63	-0.08	0.38	0.07	0.28	-0.17	0.01
IRE	0.74	0.51	0.83	0.75	0.61	0.75	0.76	0.78	0.86	-0.06	0.13	1	0.37	0.43	0.70	-0.13	0.32	0.75	0.75	0.63	-0.05	0.70	0.68
ITA	0.54	0.22	0.64	0.37	0.02	0.13	0.63	0.60	0.43	0.31	0.26	0.40	1	0.61	0.52	0.12	0.52	0.42	0.59	0.49	0.02	0.49	0.60
JPN	0.62	0.35	0.74	0.36	0.17	0.30	0.66	0.62	0.53	0.32	0.13	0.52	0.59	1	0.63	0.11	0.49	0.45	0.51	0.54	0.02	0.56	0.67
KOR	0.56	0.43	0.67	0.64	0.65	0.69	0.61	0.69	0.78	0.07	0.11	0.71	0.49	0.61	1	0.25	0.38	0.76	0.68	0.57	0.07	0.65	0.81
MEX	0.06	-0.06	-0.13	0.16	0.62	0.42	0.04	-0.08	-0.01	-0.03	-0.03	-0.09	0.18	0.12	0.25	1	0.02	0.27	-0.14	-0.14	0.11	-0.17	0.25
NLD	0.62	-0.01	0.43	0.33	-0.01	-0.01	0.38	0.28	0.38	0.16	0.73	0.38	0.57	0.50	0.31	0.13	1	0.24	0.42	0.37	0.28	0.21	0.45
POL	0.63	0.55	0.61	0.81	0.81	0.81	0.78	0.66	0.84	-0.28	-0.09	0.75	0.43	0.44	0.73	0.29	0.31	1	0.67	0.56	-0.09	0.56	0.73
ESP	0.75	0.42	0.79	0.67	0.25	0.49	0.79	0.80	0.77	0.10	0.38	0.76	0.63	0.61	0.65	-0.14	0.46	0.66	1	0.66	0.06	0.65	0.70
SWE	0.45	0.49	0.78	0.69	0.22	0.65	0.66	0.73	0.75	0.17	0.06	0.63	0.51	0.61	0.57	-0.03	0.39	0.55	0.68	1	-0.02	0.73	0.69
TUR	0.08	-0.30	-0.04	-0.04	-0.02	-0.23	-0.15	-0.13	-0.03	0.00	0.25	-0.09	-0.01	-0.06	0.07	0.18	0.26	-0.09	-0.01	-0.01	1	-0.18	0.00
GBR	0.46	0.55	0.92	0.54	0.39	0.54	0.69	0.87	0.72	0.25	-0.17	0.75	0.53	0.65	0.67	-0.16	0.26	0.56	0.74	0.74	-0.21	1	0.70
USA	0.70	0.45	0.79	0.64	0.66	0.59	0.83	0.76	0.73	0.02	0.04	0.72	0.63	0.72	0.74	0.23	0.50	0.73	0.75	0.70	-0.09	0.73	1

Notes: Pairwise Pearson (lower triangle) and Spearman (upper triangle) correlations of log transformed EPU indices (Baker et al., 2016), computed on quarterly averages over the panel A. All values are winsorized at the 1st and 99th percentiles. See Appendix B for correlations in panel B.

economies at the four-quarter lead. Yet, no significant relationships are found in the second and third quarters ahead. Notably, CCI has a negative effect in emerging markets, unlike in industrial countries. In the full sample the positive direction in the latter and the negative in the former seem to cancel each other out in part. Future work could dig deeper into the different mechanisms for CCI in the two markets. Firm-level controls have less significant impact in the emerging markets, with sales growth standing out as the strongest determinant over all four quarter leads. Tobin's Q as the primary investment opportunity variable does not seem to predict future investment in the emerging markets. Increasing macroeconomic uncertainty is not associated with decreasing future investment, as in industrial markets. Indeed, stock return volatility is associated with an increase in investment four quarters later in the emerging markets. Interestingly, the R-squared in emerging markets is significantly higher than in industrial markets. Moreover, it stays relatively stable across the leads. That could suggest a better model fit in emerging markets and a longer effect of the independent variables on investment. Overall, the results do not suggest systematic differences between the developed and developing countries. Furthermore, the small emerging market sample precludes confident inference here, and so coefficients need to be interpreted with caution. H2 is not supported by the quarterly evidence. However, this will be re-examined using the more balanced annual panel.

The test of H3, based on country-specific regressions (full sample in Table 8), reveals considerable heterogeneity across countries. While the overall and US findings are not consistently replicated, no country exhibits a statistically significant positive relationship. Significant negative coefficients are found for the United States, Japan, Canada, Poland, and Spain. Notably, four of these five countries rank among those with the largest number of observations, suggesting that insignificant results in other countries may partly reflect limited sample size. However, this cannot be the sole explanation: countries such as South Korea, Sweden, and Germany have many observations, yet show no statistically significant relationship, while Spain does, with only 2,684 observations. In fact, the absolute impact in Spain on investment is over three times that of the US (-0.132 vs.-0.032). That translates into a decline of 66.7 bp versus 16.2 bp. Interpreted as relative reductions in average national investment rates, this shows a reduction of 44.5% in Spain and 9.5% in the US, which is more than four times stronger. Notably, Gulen and Ion (2016) observe a reduction of 12.2 bp in the regression results addressing omitted variable bias, which is closest to this model. This corresponds to a relative reduction of 8.71% from their average CAPEX of 1.4% in 2016. The results presented here are very close to their findings. The relative comparison of coefficients between countries has its limitation, as average CAPEX rates differ substantially across samples. Canada's is 2.4% while the US's average CAPEX is 1.7% in panel A. That results in a discrepancy between 144% higher absolute drop in Canada but only a 73% higher relative reduction in average investment levels. However, the absolute impact on the average investment level is directly comparable. The results discussed above establish heterogeneity (supporting H3), though for countries with smaller samples, confidence intervals and robustness suffer accordingly.

On H4, the irreversibility channel is probed by dividing the sample at the median PPE ratio, as a proxy for investment irreversibility (Table 8). Interestingly, no significant negative coefficient is detected in the low-irreversibility group. Indeed, Germany and Italy yield significant positive coefficients. Conversely, all countries with significant negative re-

Table 6: EPU and Investment Regression – Panel A

Full Sample					*Excluding US*			
	CAPEX_lead1	CAPEX_lead2	CAPEX_lead3	CAPEX_lead4	CAPEX_lead1	CAPEX_lead2	CAPEX_lead3	CAPEX_lead4
EPU_log	-0.032***	-0.028***	-0.030***	-0.033***	-0.034***	-0.033***	-0.035***	-0.044***
	(0.010)	(0.008)	(0.008)	(0.007)	(0.009)	(0.008)	(0.006)	(0.009)
TobinQ	0.117***	0.111***	0.098***	0.085***	0.150***	0.140***	0.122***	0.100***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.016)	(0.017)	(0.016)	(0.015)
OCF	0.018***	0.010***	0.002	0.006**	0.025***	0.013***	0.003	0.020***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.004)
dSales	0.044***	0.041***	0.033***	0.030***	0.041***	0.041***	0.036***	0.033***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)
dGDP_Q	-0.008	0.005	0.007*	0.007	-0.004	0.009	0.008	0.007
	(0.006)	(0.007)	(0.004)	(0.006)	(0.007)	(0.007)	(0.005)	(0.005)
CCI	0.023	0.022**	0.023**	0.018	-0.005	0.005	0.009	-0.004
	(0.015)	(0.010)	(0.012)	(0.014)	(0.012)	(0.008)	(0.008)	(0.011)
stock_sd_log	-0.021*	0.020	-0.005	-0.018	-0.001	-0.012	0.011	0.019
	(0.012)	(0.022)	(0.006)	(0.014)	(0.013)	(0.011)	(0.017)	(0.015)
profit_sd_log	-0.005	-0.028**	-0.017	0.020	-0.040**	-0.022*	0.012	0.005
	(0.025)	(0.013)	(0.015)	(0.026)	(0.018)	(0.011)	(0.013)	(0.017)
N	815,700	784,165	764,557	747,493	473,261	453,051	440,505	429,482
R^2	0.011	0.010	0.008	0.007	0.006	0.006	0.005	0.005
Firm FE	yes	yes	yes	yes	yes	yes	yes	yes
Quarter dummy	yes	yes	yes	yes	yes	yes	yes	yes
Cluster firm	yes	yes	yes	yes	yes	yes	yes	yes
Cluster quarter	yes	yes	yes	yes	yes	yes	yes	yes

Notes: Regression results for panel A (23 countries, 1999–2024) using firm-quarter observations, reported for the full sample and excluding the United States. The dependent variable, CAPEX_lead{t+l}, is capital expenditures scaled by lagged total assets, measured over one to four quarter leads. The main explanatory variable is the log of the Economic Policy Uncertainty index (EPU_log) from Baker et al. (2016). Firm-level controls, per Gulen and Ion (2016), include Tobin's Q (TobinQ, market capitalization plus liabilities divided by total assets), operating cash flow (OCF, net income plus depreciation minus working capital change, scaled by lagged assets), and year-on-year sales growth (dSales). Macroeconomic controls, also per Gulen and Ion (2016), include real quarterly GDP growth (dGDP_Q) and the Consumer Confidence Index (CCI) from the OECD Data Explorer, capturing investment opportunities, and cross-sectional stock return volatility (stock_sd_log) and profit margin growth volatility (profit_sd_log), capturing uncertainty. Accounting data is from LSEG's Worldscope via WRDS. Specifications include firm fixed effects and quarter dummies for heterogeneity and seasonality. Variables are winsorized at the 1st and 99th percentiles and standardized by their respective sample standard deviation. Standard errors are clustered at the firm and quarter levels and shown in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

sults in the full sample (and Sweden and Germany) retain negative coefficients in the high-irreversibility group, validating the prediction from real options theory that the effect of uncertainty intensifies where investment is less reversible.

Again, relative comparisons of coefficients are not feasible in this context due to the lack of average CAPEX by irreversibility split. The economically higher absolute results could be a consequence of higher impact, but also higher average investment rates. The average investment rates in the high irreversibility sample could be higher, if companies which historically invested proportionally more in PPE continue to do so. For countries with few observations, H4 can hardly be tested, since the sample sizes are even smaller in the two subsamples. However, the uniformity and significance of coefficient signs strongly corroborate H4, even without further considering the economic interpretation and insufficient data in many countries. However, this will be addressed in the following section.

4.2 Panel B

The analysis of panel B both expands and complements the quarterly results. It allows for greater weighting of countries with weaker reporting frequencies in panel A and ensures a more even representation of industrialized and emerging markets. Robustness checks are supported by the inclusion of correlations and trend graphs, which are presented in Appendix B. However, the central results and patterns are detailed below with a focus on differences from panel A results.

The regression estimates for the full and non-US annual samples (see Table 9) confirm a significant and negative link between EPU and CAPEX, which is consistent with H1 and the findings from the quarterly analysis. However, the effect dissipates for time horizons beyond two to three years, which is the same recovery period noted by Gulen and Ion (2016), although no rebound period is observed after that recovery with positive coefficients. In the full sample, a doubling in EPU is associated with a slightly higher reduction in investment of 14.90% one year later (from the 6.78% average investment rate) compared to the sample without the US with a 12.79% reduction. Those are slightly higher, but very similar values compared to panel A. Tobin's Q shows a consistent positive association with investment across all horizons up to four years. In contrast, OCF is negatively correlated throughout, likely influenced by extreme left-tail values in its distribution. Annual GDP growth is positively linked to

Table 7: Industrial and Emerging Markets Comparison – Panel A

Industrial Markets					*Emerging Markets*			
	CAPEX_lead1	CAPEX_lead2	CAPEX_lead3	CAPEX_lead4	CAPEX_lead1	CAPEX_lead2	CAPEX_lead3	CAPEX_lead4
EPU_log	-0.028***	-0.026***	-0.027***	-0.026***	-0.046*	-0.035	-0.044	-0.091**
	(0.010)	(0.008)	(0.009)	(0.008)	(0.025)	(0.023)	(0.033)	(0.038)
TobinQ	0.118***	0.113***	0.100***	0.087***	-0.008	-0.038	-0.033	-0.023
	(0.005)	(0.005)	(0.005)	(0.005)	(0.043)	(0.053)	(0.048)	(0.041)
OCF	0.018***	0.010***	0.002	0.004	0.039***	0.015	0.006	0.067***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.012)	(0.012)	(0.013)	(0.024)
dSales	0.041***	0.038***	0.028***	0.025***	0.079***	0.084***	0.092***	0.082***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.018)	(0.017)	(0.020)	(0.025)
dGDP_Q	-0.008*	-0.001	0.006**	0.003	-0.006	0.051*	0.017	0.031**
	(0.005)	(0.005)	(0.003)	(0.006)	(0.022)	(0.031)	(0.019)	(0.013)
CCI	0.035**	0.031***	0.033***	0.033**	-0.049	-0.034*	-0.035**	-0.065**
	(0.015)	(0.010)	(0.013)	(0.016)	(0.032)	(0.020)	(0.016)	(0.031)
stock_sd_log	-0.025**	0.018	-0.006	-0.022	0.046	0.023	0.019	0.088**
	(0.012)	(0.022)	(0.006)	(0.014)	(0.045)	(0.043)	(0.032)	(0.038)
profit_sd_log	-0.002	-0.028**	-0.024	0.018	-0.005	-0.012	0.030	0.038
	(0.027)	(0.013)	(0.015)	(0.028)	(0.037)	(0.034)	(0.050)	(0.053)
N	772,560	742,627	723,969	707,744	43,140	41,538	40,588	39,749
R^2	0.012	0.011	0.009	0.007	0.029	0.031	0.029	0.041
Firm FE	yes	yes	yes	yes	yes	yes	yes	yes
Quarter dummy	yes	yes	yes	yes	yes	yes	yes	yes
Cluster firm	yes	yes	yes	yes	yes	yes	yes	yes
Cluster quarter	yes	yes	yes	yes	yes	yes	yes	yes

Notes: Regression results for panel A (23 countries, 1999–2024) using firm-quarter observations, split by emerging (HDI ≤ 0.8) and industrial (HDI > 0.8) markets, as classified in Table 1 per United Nations Development Programme (2020). The dependent variable is CAPEX_lead{t+l}, capital expenditures scaled by lagged total assets, measured over one to four quarter leads. The main explanatory variable is EPU_log, the log of the Economic Policy Uncertainty index from Baker et al. (2016). Firm-level controls include Tobin's Q, operating cash flow, and sales growth, while macroeconomic controls include real quarterly GDP growth and changes in the Consumer Confidence Index, both from the OECD Data Explorer. Uncertainty controls include cross-sectional stock return volatility and profit margin growth volatility, per Gulen and Ion (2016). Accounting data is from LSEG's Worldscope via WRDS. Specifications include firm fixed effects and quarter dummies for heterogeneity and seasonality. Variables are winsorized at the 1st and 99th percentiles and standardized by their respective sample standard deviation. Standard errors are clustered at the firm and quarter levels and shown in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. See Appendix B for annual regression results.

CAPEX at two-year leads, whereas CCI yields no positive coefficients. Despite differences in control variables to panel A, the findings support the acceptance of H1 on EPU.

Panel B offers a more balanced sample, where emerging markets make up 25% of firm-year observations, ensuring a stronger test of market-type effects. Table 10 shows that in both industrial and emerging markets, EPU_log remains negatively associated with subsequent CAPEX, with reductions of 82.3 basis points (an 11.93% decline from a 6.9% investment rate) in industrial economies and 67.3 basis points (10.68% off 6.3%) in emerging markets. Notably, Panel B reverses the quarterly pattern: both average investment rates and the strength of the EPU effect are lower in emerging markets than in industrial ones. CCI continues to be a more reliable predictor in industrial economies. Tobin's Q is positively related to investment in emerging markets, unlike in panel A. The effect fades after two years, in contrast to its persistence over four years in industrial countries. Moreover, profit margin growth volatility shows a significant negative association with investment in industrial economies, but the opposite pattern appears in emerging markets. Higher R-squared values in industrial countries further suggest better model fit. Overall, these findings contradict the quarterly results rather than reveal systematic differences in the EPU-investment relationship, reinforcing the rejection of H2.

A broader set of country-specific regression results (Table 11) becomes available owing to increased observation numbers. Australia and New Zealand are included in panel B and countries previously suffering from sample size constraints yield more meaningful results. Statistically significant negative EPU-CAPEX relationships are established in 11 countries. In addition to the ones specified for panel A, those are: China, South Korea, France, Germany, Denmark, Netherlands and Ireland. Meanwhile the significant effects in Japan and Poland are absent in panel B. Contrasts are notable as positive results appear in India, UK, Sweden, and New Zealand. In many countries, such as Australia or Turkey, no significant results are found at all. That reinforces mixed results in related literature in those countries (Azimli, 2022; Chen et al., 2020; Wu et al., 2020). Coefficients range from -0.158 in Ireland (295 bp or 38.31% reduction) to +0.211 in Sweden (394 bp or 49.87% increase), presenting completely opposing effects from EPU on investment in different countries. The presence of positive coefficients is especially intriguing as they deviate from the norm in the literature and findings in most countries in this study. Uncertainty could here be perceived more as a growth option than as potential sunk costs. A deeper analysis of these countries could im-

Table 8: Country Comparison and Irreversibility – Panel A

	Full Sample			*Low Irreversibility*			*High Irreversibility*		
	EPU_log	N	R^2	EPU_log	N	R^2	EPU_log	N	R^2
United States	-0.032**	342,439	0.019	0.017	201,196	0.018	-0.085***	141,243	0.033
Japan	-0.024***	198,680	0.005	-0.003	89,133	0.002	-0.040***	109,547	0.011
Korea (South)	0.002	92,011	0.003	0.009	33,875	0.002	-0.003	58,136	0.007
Canada	-0.078***	42,295	0.029	0.001	15,390	0.005	-0.109***	26,905	0.054
Poland	-0.040***	23,236	0.009	0.005	10,827	0.006	-0.065***	12,409	0.018
Sweden	-0.034	21,557	0.013	-0.001	16,526	0.019	-0.148*	5,031	0.015
Germany	-0.003	18,803	0.011	0.023*	11,089	0.014	-0.026*	7,714	0.021
Brazil	-0.029	16,333	0.012	0.008	7,420	0.01	-0.043	8,913	0.016
Turkey	-0.057	15,867	0.092	-0.017	6,445	0.074	-0.083	9,422	0.107
Chile	0.049	9,520	0.02	-0.011	1,717	0.008	0.061	7,803	0.027
Italy	0.071	6,825	0.013	0.102*	3,857	0.024	0.065	2,968	0.027
Mexico	-0.055	6,607	0.028	0.115	1,459	0.056	-0.074	5,148	0.03
Greece	-0.063	5,948	0.016	-0.047	1,545	0.014	-0.064	4,403	0.023
Denmark	0.064	3,326	0.018	0.110	1,370	0.04	0.023	1,956	0.023
China	-0.015	2,755	0.017	0.040	1,322	0.019	-0.019	1,433	0.022
Spain	-0.132**	2,684	0.041	-0.034	1,127	0.036	-0.129**	1,557	0.049
Netherlands	-0.054	1,657	0.01	-0.036	1,050	0.012	-0.103	607	0.013
United Kingdom	0.049	1,317	0.035	0.030	512	0.045	0.092	805	0.076
France	-0.018	1,184	0.084	0.021	738	0.216	-0.001	446	0.077
Colombia	0.277	790	0.124	0.385	227	0.252	0.217	563	0.091
India	-0.009	788	0.033	-0.016	549	0.025	0.064	239	0.074
Belgium	-0.013	726	0.022	-0.067	237	0.031	0.011	489	0.044
Ireland	-0.021	352	0.051	-0.004	239	0.042	-0.049	113	0.236

Notes: Regression results for panel A (23 countries, 1999–2024) using firm-quarter observations per country, split by low (below median) and high (above median) PPE ratio, defined as Property, Plant, and Equipment divided by total assets, as a proxy for investment irreversibility per Gulen and Ion (2016). The table is sorted by full sample observations. The dependent variable is CAPEX_lead{t+l}, capital expenditures scaled by lagged total assets, measured over one to four quarter leads. The main explanatory variable is EPU_log, the log of the Economic Policy Uncertainty index from Baker et al. (2016). Firm-level controls include Tobin's Q, operating cash flow, and sales growth, while macroeconomic controls include real quarterly GDP growth and changes in the Consumer Confidence Index, both from the OECD Data Explorer. Uncertainty controls include cross-sectional stock return volatility and profit margin growth volatility, per Gulen and Ion (2016). Accounting data is from LSEG's Worldscope via WRDS. Specifications include firm fixed effects and quarter dummies for heterogeneity and seasonality. Variables are winsorized at the 1st and 99th percentiles and standardized by their respective sample standard deviation. Standard errors are clustered at the firm and quarter levels. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

prove the understanding of how uncertainty influences corporate investment. The heterogeneity strongly supports H3 and merits further inquiry into the roots of national differences.

Analysis of H4 through irreversibility splits (Table 11) reiterates the quarterly finding: high-irreversibility samples generally amplify negative EPU effects, whereas low-irreversibility samples show weaker evidence of the relationship, China and Spain being the only notable exceptions with a persistent but less significant coefficient. In India and New Zealand, the positive relationship from the full sample is statistically strengthened in the high irreversibility sample and lowered in the low irreversibility sample. This suggests that different mechanisms are at play in these countries than those suggested by real options theory and the findings in all other countries. In Sweden and the UK, the positive association is driven by firms facing lower irreversibility. Due to the lack of mean CAPEX data per irreversibility group, the relative interpretation is not extended, but in terms of significance and direction, H4 is again strongly affirmed, if countries with overall positive effects from EPU on investment are excluded as anomalies.

Taken together, these results demonstrate that the negative correlation between economic policy uncertainty and corporate investment is robust and prevalent across nations. This relationship is evident even when the US is excluded, across industrial and emerging markets, and particularly in countries with high statistical power. Investment irreversibility emerges as a key amplifying channel, supporting predictions from real options theory. Meanwhile, the apparent lack of systematic differences between industrial and emerging markets and the clear evidence of country-specific heterogeneity add nuance to findings from panel A. In summary, the available data supports hypotheses H1, H3, and H4, while H2 cannot be confirmed based on the available data. The positive associations observed in India, the United Kingdom, Sweden, and New Zealand are particularly noteworthy. In India and New Zealand, the irreversibility channel is not evident either. These findings suggest that distinct mechanisms may be operating in these countries and warrant further investigation.

5 Discussion

This thesis presents clear empirical findings on the hypotheses. However, they must be interpreted while considering several methodological and conceptual limitations.

Table 9: EPU and Investment Regression – Panel B

Full Sample					*Excluding US*			
	CAPEX_lead1	CAPEX_lead2	CAPEX_lead3	CAPEX_lead4	CAPEX_lead1	CAPEX_lead2	CAPEX_lead3	CAPEX_lead4
EPU_log	-0.054***	-0.031**	-0.027	-0.022	-0.049***	-0.026**	-0.023	-0.016
	(0.010)	(0.014)	(0.017)	(0.015)	(0.009)	(0.013)	(0.017)	(0.014)
TobinQ	0.201***	0.103***	0.049***	0.027***	0.201***	0.101***	0.051***	0.030***
	(0.010)	(0.008)	(0.007)	(0.006)	(0.011)	(0.014)	(0.010)	(0.008)
OCF	-0.023***	-0.019***	-0.011**	-0.010**	-0.030***	-0.026***	-0.019***	-0.023***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
dSales	0.031***	-0.003	0.000	-0.010*	0.030***	-0.001	0.002	-0.009*
	(0.005)	(0.004)	(0.004)	(0.005)	(0.006)	(0.005)	(0.004)	(0.005)
dGDP_Y	0.029*	0.028**	0.007	0.018	0.026*	0.028**	0.010	0.020
	(0.016)	(0.014)	(0.013)	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)
CCI	0.009	-0.003	-0.006	-0.007	0.004	-0.006	-0.009	-0.013
	(0.010)	(0.008)	(0.009)	(0.011)	(0.010)	(0.008)	(0.008)	(0.011)
stock_sd_log	-0.015	-0.011	-0.018	-0.011	0.013	0.019	0.000	0.004
	(0.011)	(0.012)	(0.018)	(0.018)	(0.012)	(0.012)	(0.015)	(0.014)
profit_sd_log	-0.040**	-0.013	0.024	0.039**	-0.026	-0.001	0.030	0.040**
	(0.017)	(0.017)	(0.023)	(0.019)	(0.019)	(0.016)	(0.022)	(0.020)
N	474,372	413,184	349,558	288,064	364,835	319,883	272,825	225,389
R^2	0.029	0.007	0.002	0.002	0.021	0.006	0.002	0.002
Firm FE	yes	yes	yes	yes	yes	yes	yes	yes
Cluster firm	yes	yes	yes	yes	yes	yes	yes	yes
Cluster year	yes	yes	yes	yes	yes	yes	yes	yes

Notes: Regression results for panel B (25 countries, 1991–2023) using firm-year observations, reported for the full sample and excluding the United States. The dependent variable, CAPEX_lead{t+l}, is capital expenditures scaled by lagged total assets, measured over one to four year leads. The main explanatory variable is the log of the Economic Policy Uncertainty index (EPU_log) from Baker et al. (2016). Firm-level controls, per Gulen and Ion (2016), include Tobin's Q (TobinQ, market capitalization plus liabilities divided by total assets), operating cash flow (OCF, net income plus depreciation minus working capital change, scaled by lagged assets), and year-on-year sales growth (dSales). Macroeconomic controls, also per Gulen and Ion (2016), include real annual GDP growth (dGDP_Y) and the Consumer Confidence Index (CCI) from the OECD Data Explorer, capturing investment opportunities, and cross-sectional stock return volatility (stock_sd_log) and profit margin growth volatility (profit_sd_log), capturing uncertainty. Accounting data is from LSEG's Worldscope via WRDS. Specifications include firm fixed effects and heterogeneity. Variables are winsorized at the 1st and 99th percentiles and standardized by their respective sample standard deviation. Standard errors are clustered at the firm and year levels and shown in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

While these limitations are inherent to cross-country empirical research of this scope, they are presented transparently to contextualize and strengthen the credibility of the results.

A fundamental challenge lies in approximating real economic behavior using published financial statements. While these statements provide structured, standardized, and internationally comparable data, they inherently convert continuous, strategic business processes into discrete, backward-looking periods. Furthermore, firms may engage in earnings management to meet expectations, manage volatility, or influence perceptions. This can result in a discrepancy between a firm's actual financial condition and its reported financial position. Cui, Yao, et al. (2021) demonstrate that earnings management intensifies during periods of high economic policy uncertainty, amplifying this limitation in the empirical setting of this study.

In line with previous studies, this thesis employs capital expenditure (CAPEX) as the primary indicator of investment (Azimli, 2022; Gulen & Ion, 2016; Jens, 2017; Julio & Yook, 2012; Wang et al., 2014). However, two major limitations must be acknowledged. First, CAPEX only captures investment in tangible assets, specifically additions to property, plant, and equipment (PPE). It excludes investments in intangible assets such as R&D, software, or organizational capital, and other non-current investments, such as financial subsidiaries. Consequently, it provides only a partial view of firms' total investment behavior. Secondly, this study uses a self-constructed CAPEX measure, defined as the change in PPE plus depreciation, amortization, and depletion, to maximize data availability and ensure comparability across countries and time. While depreciation and depletion relate to physical capital, amortization is associated with intangible assets. By adding amortization to PPE flows, the measure incorporates impairments of intangible assets, while still excluding actual intangible investment. This constitutes a conceptual inconsistency. Impairments of goodwill can be particularly significant during periods of economic stress, which EPU seems to be correlated with (Table 4). Nevertheless, the constructed CAPEX proxy correlates highly with the directly reported Worldscope CAPEX item in periods and countries where both are available, supporting its validity. Still, it must be emphasized that this study focuses on one dimension of investment. Closely related yet distinct research examines how EPU affects investments or outcomes in innovation. For instance, Atanassov et al. (2024) document a positive association between EPU and R&D, while Cui, Wang, et al. (2021) and Mbanyele and Wang (2022) find negative effects of EPU on patent activity in China and across countries, respectively.

Table 10: Industrial and Emerging Markets – Panel B

Industrial Markets					*Emerging Markets*			
	CAPEX_lead1	CAPEX_lead2	CAPEX_lead3	CAPEX_lead4	CAPEX_lead1	CAPEX_lead2	CAPEX_lead3	CAPEX_lead4
EPU_log	-0.044***	-0.023	-0.016	-0.021	-0.036**	-0.006	-0.020	-0.008
	(0.015)	(0.021)	(0.024)	(0.017)	(0.017)	(0.021)	(0.020)	(0.031)
TobinQ	0.206***	0.104***	0.049***	0.025***	0.077***	0.068***	0.035	0.039
	(0.010)	(0.008)	(0.007)	(0.007)	(0.022)	(0.024)	(0.024)	(0.029)
OCF	-0.027***	-0.020***	-0.008	-0.011**	0.009	-0.005	-0.027***	0.000
	(0.004)	(0.006)	(0.006)	(0.005)	(0.012)	(0.008)	(0.009)	(0.010)
dSales	0.025***	-0.007*	-0.001	-0.013**	0.057***	0.018	0.005	0.000
	(0.005)	(0.004)	(0.004)	(0.006)	(0.013)	(0.011)	(0.007)	(0.008)
dGDP_Y	0.019	0.004	0.007	0.023*	0.025*	0.048	-0.010	0.014
	(0.014)	(0.010)	(0.011)	(0.013)	(0.013)	(0.030)	(0.016)	(0.034)
CCI	0.029**	0.018	0.013	0.002	-0.003	-0.011	-0.012	-0.021
	(0.014)	(0.015)	(0.019)	(0.015)	(0.014)	(0.014)	(0.013)	(0.018)
stock_sd_log	-0.018	-0.016	-0.021	-0.018	-0.010	0.022	0.015	0.035
	(0.012)	(0.012)	(0.017)	(0.019)	(0.014)	(0.043)	(0.047)	(0.040)
profit_sd_log	-0.089***	-0.069***	-0.016	0.036	0.073*	0.079***	0.097**	0.048
	(0.016)	(0.019)	(0.027)	(0.026)	(0.038)	(0.030)	(0.040)	(0.031)
N	377,741	329,514	279,068	232,644	96,631	83,670	70,490	55,420
R^2	0.035	0.009	0.002	0.001	0.014	0.012	0.008	0.004
Firm FE	yes	yes	yes	yes	Yes	yes	yes	yes
Cluster firm	yes	yes	yes	yes	Yes	yes	yes	yes
Cluster year	yes	yes	yes	yes	Yes	yes	yes	yes

Notes: Regression results for panel B (25 countries, 1991–2023) using firm-year observations, split by emerging (HDI $\leq$ 0.8) and industrial (HDI $>$ 0.8) markets, as classified in Table 1 per United Nations Development Programme (2020). The dependent variable is CAPEX_lead{t+l}, capital expenditures scaled by lagged total assets, measured over one to four year leads. The main explanatory variable is EPU_log, the log of the Economic Policy Uncertainty index from Baker et al. (2016). Firm-level controls include Tobin's Q, operating cash flow, and sales growth, while macroeconomic controls include real annual GDP growth and changes in the Consumer Confidence Index, both from the OECD Data Explorer. Uncertainty controls include cross-sectional stock return volatility and profit margin growth volatility, per Gulen and Ion (2016). Accounting data is from LSEG's Worldscope via WRDS. Specifications include firm fixed heterogeneity. Variables are winsorized at the 1st and 99th percentiles and standardized by their respective sample standard deviation. Standard errors are clustered at the firm and year levels and shown in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

This underlines that investment responses to uncertainty are multi-faceted, and CAPEX alone does not capture the full behavioral spectrum. Meinen and Röhe (2017) find similar results across EU countries using macroeconomic investment aggregates. Kolev and Randall (2024) reinforce it in the EU with survey data.

Another important limitation is the composition of countries in the dataset. While the sample formally spans over 20 countries, it is heavily concentrated in a small group of large economies. In the quarterly panel, the top five countries account for over 86% of firm-quarters; in the annual panel, the figure is 73%. Conversely, the bottom 50% of countries contribute only 3% and 7% of observations, respectively. Even when the US is excluded, the panels are still dominated by industrial economies, which limits the ability to draw meaningful conclusions about emerging markets or smaller jurisdictions. These imbalances significantly constrain the interpretability of the regression results not split by country. Furthermore, in many countries, the number of observations is simply too low to allow for robust inference.

An additional sampling bias arises from the selection of firms. The study includes only large, listed firms, following standard data quality filters such as exclusion of private and cross-listed companies. This selection bias means that the results reflect the behavior of a specific corporate subset. Many sectors, especially in emerging markets, remain dominated by small or unlisted firms whose responses to policy uncertainty may differ substantially. Indeed, a great body of literature investigates how firm characteristics influence the relationship between EPU and investment. For example, Gulen and Ion (2016) demonstrate that firms with high government contract exposure are more sensitive to EPU. Wang et al. (2014) found that non-state-owned firms, firms with low internal finance, and firms with lower returns on invested capital exhibit greater responsiveness. Kang et al. (2014) highlight that firms facing high internal uncertainty react more strongly as well. Amore and Minichilli (2018) demonstrate that family ownership mitigates the negative effect of EPU on investment in Italy. These findings illustrate the critical role of firm-level heterogeneity in shaping investment behavior under uncertainty. The present study does not directly investigate these dimensions, but addresses unobserved heterogeneity using firm fixed effects, which partially absorb time-invariant structural characteristics. Moreover, the analysis includes investment irreversibility, a key dimension of firm-level heterogeneity, operationalized through a median split on the PPE-to-assets ratio. This reflects its central role in uncertainty-investment models as demonstrated in Leahy and Whited (1995), Wang et al. (2014), Gulen and Ion (2016), Baker et al. (2016), Jens (2017), and Dibiasi et al. (2018).

Another challenge lies in the construction and interpreta-

Table 11: Country Comparison and Irreversibility – Panel B

	Full Sample			*Low Irreversibility*			*High Irreversibility*		
	EPU_log	N	R^2	EPU_log	N	R^2	EPU_log	N	R^2
United States	-0.073**	109,537	0.057	-0.009	64,314	0.065	-0.152***	45,223	0.073
Japan	-0.001	95,556	0.009	0.009	41,714	0.004	-0.005	53,842	0.019
China	-0.073***	58,224	0.019	-0.032**	28,770	0.019	-0.109***	29,454	0.027
Korea (South)	-0.088**	35,722	0.02	-0.046	13,408	0.006	-0.114**	22,314	0.036
India	0.051*	27,508	0.007	-0.026	12,132	0.006	0.087**	15,376	0.011
Australia	-0.054	26,900	0.046	0.017	12,667	0.022	-0.109	14,233	0.072
United Kingdom	0.035*	24,124	0.039	0.049***	14,868	0.035	0.011	9,256	0.068
France	-0.052**	16,664	0.03	-0.032	11,608	0.025	-0.107***	5,056	0.066
Canada	-0.172***	15,767	0.059	-0.065	5,647	0.04	-0.256***	10,120	0.092
Germany	-0.061***	14,601	0.03	-0.028	8,229	0.031	-0.102***	6,372	0.043
Sweden	0.211*	7,210	0.015	0.279**	5,368	0.025	0.091	1,842	0.014
Poland	-0.078	6,814	0.031	-0.036	3,356	0.026	-0.091	3,458	0.052
Brazil	-0.070	5,276	0.024	-0.005	2,150	0.021	-0.116	3,126	0.033
Italy	-0.008	5,252	0.035	0.044	3,154	0.03	-0.030	2,098	0.063
Greece	0.094	4,825	0.086	-0.023	1,653	0.022	0.180	3,172	0.104
Turkey	-0.182	3,811	0.235	-0.022	1,582	0.158	-0.249**	2,229	0.284
Denmark	-0.151**	2,888	0.033	-0.026	1,159	0.06	-0.209**	1,729	0.034
Spain	-0.134**	2,864	0.062	-0.086*	1,280	0.039	-0.164**	1,584	0.087
Chile	-0.019	2,592	0.023	-0.025	509	0.039	-0.012	2,083	0.027
New Zealand	0.122**	2,017	0.035	0.137*	855	0.059	0.112**	1,162	0.048
Belgium	-0.017	1,812	0.035	0.030	894	0.047	-0.025	918	0.046
Netherlands	-0.141*	1,492	0.063	-0.082	929	0.067	-0.288**	563	0.073
Mexico	0.036	1,410	0.023	0.059	317	0.131	0.037	1,093	0.037
Ireland	-0.158***	1,104	0.074	-0.043	521	0.062	-0.142*	583	0.123
Colombia	-0.001	402	0.032	0.325	102	0.097	-0.188	300	0.032

Notes: Regression results for panel B (25 countries, 1991–2023) using firm-year observations per country, split by low (below median) and high (above median) PPE ratio, defined as Property, Plant, and Equipment divided by total assets, as a proxy for low and high investment irreversibility. The table is sorted by full sample observations. The dependent variable is CAPEX_lead{t+l}, capital expenditures scaled by lagged total assets, measured over one to four year leads. The main explanatory variable is EPU_log, the log of the Economic Policy Uncertainty index from Baker et al. (2016). Firm-level controls include Tobin's Q, operating cash flow, and sales growth, while macroeconomic controls include real annual GDP growth and changes in the Consumer Confidence Index, both from the OECD Data Explorer. Uncertainty controls include cross-sectional stock return volatility and profit margin growth volatility, per Gulen and Ion (2016). Accounting data is from LSEG's Worldscope via WRDS. Specifications include firm fixed effects and annual dummies for heterogeneity and seasonality. Variables are winsorized at the 1st and 99th percentiles and standardized by their respective sample standard deviation. Standard errors are clustered at the firm and year levels. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

tion of the EPU index itself. The index is derived from newspaper coverage of uncertainty-related keywords, allowing for systematic tracking of policy-related uncertainty over time and across countries (Baker et al., 2016). However, the number and nature of newspaper sources vary significantly: some indices are based on just one newspaper (e.g., Brazil, Colombia or Ireland), and major European countries such as Germany, France, Italy, and Spain rely on only two. This limited coverage increases the risk of editorial bias, and differences in journalistic style or digitization can affect the volume and tone of published uncertainty. Moreover, the keyword-based methodology may omit articles that discuss uncertainty in broader or indirect terms, potentially underestimating true uncertainty exposure. A particularly important issue is the long-term upward trend in EPU across most countries. This could reflect a genuine increase in economic policy uncertainty, but it may also stem from structural shifts in media ecosystems. The distinction between public and published uncertainty complicates interpretation. The present study uses only the newspaper-based variant of the index, even in countries like the US where broader composite indices exist. This choice follows cross-country comparability and the empirical strategy of Gulen and Ion (2016), who show that the news-based EPU index best captures the uncertainty-investment relationship in the US. The construction of the index is discussed at length in Baker et al. (2016). Azqueta-Gavaldón et al. (2023) investigate the sources of Economic Policy Uncertainty (EPU) in the euro area and their varying impacts on investment across countries. Their findings indicate that while most types of uncertainty affect investment differently from one country to another, political uncertainty and domestic regulatory uncertainty consistently reduce investment throughout the region.

As with all studies of this kind, the risk of omitted variable bias remains a key concern. Investment decisions are shaped by an extensive array of economic, strategic, and institutional factors, many of which are difficult to measure or unavailable in a consistent cross-country format. Although standard controls are included, the relatively low R-squared values across all model specifications underscore the substantial share of unexplained investment variation. This is not a sign of model failure, but a structural reality of this research context.

The included variables also create challenges. The dis-

tribution of operating cash flow is notably skewed, particularly in the annual panel, where extreme negative values may disproportionately influence the negative relationship with CAPEX. The volatility measure based on market capitalization changes may also understate true stock market volatility as it is calculated by stock returns at the frequency of the respective panel instead of daily stock returns. The decision to apply a binary irreversibility measure rather than a continuous interaction reflects concerns over mechanical collinearity between CAPEX and the PPE ratio. The use of a global median to split firms by investment irreversibility does not account for structural differences across countries, where fixed asset intensities and capital market constraints may vary substantially.

Moreover, the temporal structure of the model introduces its own limitations. All explanatory variables are lagged relative to the investment outcome to avoid mechanical correlation with financial statement-based controls. This strategy is particularly important given the overlapping components between those variables and CAPEX. However, such lag structures also imply a specific causal direction. In the quarterly panel, where the first lead represents one quarter, this assumption is more plausible. In the annual panel, by contrast, the first lead corresponds to one full year, raising the risk of a simultaneity bias.

Despite these limitations, the results of this study reveal a robust and consistent negative correlation between economic policy uncertainty and corporate investment across a broad set of industrialized and emerging economies. However, several countries deviate from this global pattern. For instance, in the annual panel, statistically significant positive relationships are found in India, the UK, New Zealand and Sweden. South Korea stands out as the only country with many observations but no observed relationship in the quarterly panel, while results for Australia remain inconclusive. These findings underscore the need for deeper national studies.

6 Conclusion

Although economic policy uncertainty and its consequences for corporate investment have attracted considerable scholarly attention, existing evidence remains largely confined to very few countries. This thesis explores this relationship on an unprecedented global scale, by assembling a dual-frequency panel encompassing quarterly data from 23 countries and annual data from 25 countries over more than two decades. In doing so, it addresses a critical gap in the literature by examining the generalizability of the uncertainty-investment relationship across diverse institutional, regulatory, and market environments.

Building on the empirical framework of Gulen and Ion (2016), this study regressed firm-level capital expenditures on the log-transformed index of economic policy uncertainty from Baker et al. (2016), alongside standard firm- and macro-level controls. The design retains their use of firm fixed effects, lead structures, and double-clustered standard errors, while extending the analysis across a much broader set of economies and over both quarterly and annual horizons. This methodology allowed for direct comparison with seminal US- and China-focused results.

Across both frequencies, higher EPU consistently preceded lower investment. That relationship seems to vanish after two to three years, as suggested by Gulen and Ion (2016) for the US, but no observable rebound period was established. Although the negative effect appeared somewhat weaker in emerging markets at the quarterly frequency, the annual panel revealed comparable sensitivity in both industrial and emerging subsamples. A detailed country-by-country examination uncovered substantial heterogeneity. Countries like Spain or Canada displayed particularly strong negative responses, while others showed no discernible effect, and India, the UK, New Zealand and Sweden even exhibited positive associations in the annual analysis. Crucially, firms with greater capital irreversibility proved to be more responsive to uncertainty, except for India and New Zealand. This corroborates real-options theory and underscores the role of sunk-cost intensity in shaping investment behavior.

This thesis makes two primary contributions by documenting these patterns across a wide array of national contexts. Firstly, it demonstrates the robustness of the policy uncertainty effect on corporate investment in many countries other than the US and China, showing that it is relevant in both mature and developing economies. Second, it highlights the importance of country-specific factors that modulate this effect, calling for further research to investigate the sources of cross-country variation. Deeper firm-level investigation through alternative investment measures, including R&D, intangible assets, and firm characteristics could deepen our understanding of the factors through which economic policy uncertainty affects investment and how it is stimulated. This study should particularly call for interest in countries where positive relationships were found.

This work provides policymakers with valuable insights on how companies react to their decisions and communication. It also provides stakeholders involved in investment decisions around the world with knowledge of how others react to policy uncertainty. Most importantly, it offers a structured country overview that highlights where future country-specific research is most needed.

References

Abel, A. B. (1983). Optimal Investment Under Uncertainty. *The American Economic Review*, *73*(1), 228–233. http://www.jstor.org/stable/1803942

Algaba, A., Borms, S., Boudt, K., & van Pelt, J. (2020). The Economic Policy Uncertainty Index for Flanders, Wallonia and Belgium. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3580000

Ali, S., Badshah, I., Demirer, R., & Hegde, P. (2022). Economic Policy Uncertainty and Institutional Investment Returns: The Case of New Zealand. *Pacific-Basin Finance Journal*, *74*, 101797. https://doi.org/10.1016/j.pacfin.2022.101797

Amore, M. D., & Minichilli, A. (2018). Local Political Uncertainty, Family Control, and Investment Behavior. *Journal of Financial and Quantitative Analysis*, *53*(4), 1781–1804. https://doi.org/10.1017/S002210901800025X

An, H., Chen, Y., Luo, D., & Zhang, T. (2016). Political Uncertainty and Corporate Investment: Evidence From China. *Journal of Corporate Finance*, *36*, 174–189. https://doi.org/10.1016/j.jcorpfin.2015.11.003

Arbatli, E., Davis, S., Ito, A., & Miake, N. (2017). Policy Uncertainty in Japan. *National Bureau of Economic Research*. https://doi.org/10.3386/w23411

Armelius, H., Hull, I., & Stenbacka Köhler, H. (2017). The Timing of Uncertainty Shocks in a Small Open Economy. *Economics Letters*, *155*, 31–34. https://doi.org/10.1016/j.econlet.2017.03.016

Atanassov, J., Julio, B., & Leng, T. (2024). The Bright Side of Political Uncertainty: The Case of R&D. *The Review of Financial Studies*, *37*(10), 2937–2970. https://doi.org/10.1093/rfs/hhae023

Azimli, A. (2022). The Impact of Policy, Political and Economic Uncertainty on Corporate Capital Investment in the Emerging Markets of Eastern Europe and Turkey. *Economic Systems*, *46*(2), 100974. https://doi.org/10.1016/j.ecosys.2022.100974

Azqueta-Gavaldón, A., Hirschbühl, D., Onorante, L., & Saiz, L. (2023). Sources of Economic Policy Uncertainty in the Euro Area. *European Economic Review*, *152*, 104373. https://doi.org/10.1016/j.euroecorev.2023.104373

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Bar-Ilan, A., & Strange, W. C. (1996). Investment Lags. *The American Economic Review*, *86*(3), 610–622. https://www.jstor.org/stable/2118214

Bergman, U. M., & Worm, C. H. (2020). *Economic Policy Uncertainty and Consumer Perceptions: The Danish Case*. https://www.researchgate.net/profile/u-bergman/publication/343832963_economic_policy_uncertainty_and_consumer_perceptions_the_danish_case

Bernanke, B. S. (1983). Irreversibility, Uncertainty, and Cyclical Investment. *The Quarterly Journal of Economics*, *98*(1), 85. https://doi.org/10.2307/1885568

Bialkowski, J. P., Klepka, R., & Slawik, A. (2025). Local and External Economic Policy Uncertainty and Their Impact on Exchange Rates in Emerging Markets: Evidence From Poland. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5104696

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica*, *77*(3), 623–685. https://doi.org/10.3982/ECTA6248

Bonaime, A., Gulen, H., & Ion, M. (2018). Does Policy Uncertainty Affect Mergers and Acquisitions? *Journal of Financial Economics*, *129*(3), 531–558. https://doi.org/10.1016/j.jfineco.2018.05.007

Caballero, R. J. (1991). On the Sign of the Investment-Uncertainty Relationship. *The American Economic Review*, *81*(1), 279–288. https://www.jstor.org/stable/2006800

Cerda, R., Silva, A., & Valente, J. T. (2016). *Economic Policy Uncertainty Indices for Chile* [Doctoral dissertation, working paper]. https://policyuncertainty.com/media/EPU_Chile.pdf

Chen, X., Le, C. H. A., Shan, Y., & Taylor, S. (2020). Australian Policy Uncertainty and Corporate Investment. *Pacific-Basin Finance Journal*, *61*, 101341. https://doi.org/10.1016/j.pacfin.2020.101341

Cui, X., Wang, C., Liao, J., Fang, Z., & Cheng, F. (2021). Economic Policy Uncertainty Exposure and Corporate Innovation Investment: Evidence From China. *Pacific-Basin Finance Journal*, *67*, 101533. https://doi.org/10.1016/j.pacfin.2021.101533

Cui, X., Yao, S., Fang, Z., & Wang, H. (2021). Economic Policy Uncertainty Exposure and Earnings Management: Evidence From China. *Accounting & Finance*, *61*(3), 3937–3976. https://doi.org/10.1111/acfi.12722

Davis, S. J., Liu, D., & Sheng, X. S. (2019). Economic Policy Uncertainty in China Since 1949: The View From Mainland Newspapers. https://stevenjdavis.com/s/epu-in-china-view-from-mainland-newspapers-august-2019.pdf

Davis, S. (2016). An Index of Global Economic Policy Uncertainty. *National Bureau of Economic Research*. https://doi.org/10.3386/w22740

Dibiasi, A., Abberger, K., Siegenthaler, M., & Sturm, J.-E. (2018). The Effects of Policy Uncertainty on Investment: Evidence From the Unexpected Acceptance of a Far-Reaching Referendum in Switzerland. *European Economic Review*, *104*, 38–67. https://doi.org/10.1016/j.euroecorev.2018.01.002

Dixit, A. K., & Pindyck, R. S. (1994). *Investment Under Uncertainty*. Princeton University Press.

Fedesarrollo. (n.d.). Índice de Incertidumbre de la Política Económica en Colombia (IPEC) [Retrieved July 17, 2025]. https://www.repository.fedesarrollo.org.co/handle/11445/4626/discover

Fujitani, R., Hattori, M., & Yasuda, Y. (2023). Domestic and International Effects of Economic Policy Uncertainty on Corporate Investment and Strategic Cash Holdings: Evidence From Japan. *Journal of the Japanese and International Economies*, *69*, 101272. https://doi.org/10.1016/j.jjie.2023.101272

Ghirelli, C., Pérez, J. J., & Urtasun, A. (2019). A New Economic Policy Uncertainty Index for Spain. *Economics Letters*, *182*, 64–67. https://doi.org/10.1016/j.econlet.2019.05.021

Guiso, L., & Parigi, G. (1999). Investment and Demand Uncertainty. *The Quarterly Journal of Economics*, *114*(1), 185–227. https://doi.org/10.1162/003355399555981

Gulen, H., & Ion, M. (2016). Policy Uncertainty and Corporate Investment. *Review of Financial Studies*, hhv050. https://doi.org/10.1093/rfs/hhv050

Hardouvelis, G. A., Karalas, G., Karanastasis, D., & Samartzis, P. (2018). Economic Policy Uncertainty, Political Uncertainty and the Greek Economic Crisis. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3155172

Hartman, R. (1972). The Effects of Price and Cost Uncertainty on Investment. *Journal of Economic Theory*, *5*(2), 258–266. https://doi.org/10.1016/0022-0531(72)90105-6

Hartman, R. (1976). Factor Demand With Output Price Uncertainty. *The American Economic Review*, *66*(4), 675–681. https://www.jstor.org/stable/1806710

Jens, C. E. (2017). Political Uncertainty and Investment: Causal Evidence From U.S. Gubernatorial Elections. *Journal of Financial Economics*, *124*(3), 563–579. https://doi.org/10.1016/j.jfineco.2016.01.034

Julio, B., & Yook, Y. (2012). Political Uncertainty and Corporate Investment Cycles. *The Journal of Finance*, *67*(1), 45–83. https://doi.org/10.1111/j.1540-6261.2011.01707.x

Kang, W., Lee, K., & Ratti, R. A. (2014). Economic Policy Uncertainty and Firm-Level Investment. *Journal of Macroeconomics*, *39*, 42–53. https://doi.org/10.1016/j.jmacro.2013.10.006

Kilic, I., & Balli, F. (2024). Measuring Economic Country-Specific Uncertainty in Türkiye. *Empirical Economics*, *67*(4), 1649–1689. https://doi.org/10.1007/s00181-024-02594-z

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Riverside Press Cambridge.

Kolev, A., & Randall, T. (2024). *The Effect of Uncertainty on Investment: Evidence From EU Survey Data*. European Investment Bank. https://data.europa.eu/doi/10.2867/723130

Kroese, L., Kok, S., & Parlevliet, J. (2015). Beleidsonzekerheid in Nederland. *Economisch Statistische Berichten*, (4715), 464–467.

Leahy, J., & Whited, T. (1995). The Effect of Uncertainty on Investment: Some Stylized Facts. *National Bureau of Economic Research*. https://doi.org/10.3386/w4986

Mbanyele, W., & Wang, F. (2022). Economic Policy Uncertainty and Industry Innovation: Cross Country Evidence. *The Quarterly Review of*

Economics and Finance, *84*, 208–228. https://doi.org/10.1016/j.qref.2022.01.012

Meinen, P., & Röhe, O. (2017). On Measuring Uncertainty and Its Impact on Investment: Cross-Country Evidence From the Euro Area. *European Economic Review*, *92*, 161–179. https://doi.org/10.1016/j.euroecorev.2016.12.002

Meng, X., Guo, H., & Li, J. (2023). Economic Policy Uncertainty and Corporate Investment: Evidence From China. *Economic Change and Restructuring*, *56*(6), 4491–4529. https://doi.org/10.1007/s10644-023-09564-y

Oi, W. Y. (1961). The Desirability of Price Instability Under Perfect Competition. *Econometrica*, *29*(1), 58. https://doi.org/10.2307/1907687

Rodrik, D. (1991). Policy Uncertainty and Private Investment in Developing Countries. *Journal of Development Economics*, *36*(2), 229–242. https://doi.org/10.1016/0304-3878(91)90034-S

Rottig, D. (2016). Institutions and Emerging Markets: Effects and Implications for Multinational Corporations. *International Journal of Emerging Markets*, *11*(1), 2–17. https://doi.org/10.1108/IJoEM-12-2015-0248

Soric, P., & Lolic, I. (2017). Economic Uncertainty and Its Impact on the Croatian Economy. *Public Sector Economics*, *41*(4), 443–477. https://doi.org/10.3326/pse.41.4.3

Stiglitz, J. E., & Weiss, A. (1981). Credit Rationing in Markets With Imperfect Information. *The American Economic Review*, *71*(3), 393–410. https://www.jstor.org/stable/1802787

United Nations Development Programme. (2020). Data and Statistics Readers Guide. Retrieved July 17, 2025, from https://hdr.undp.org/reports-and-publications/2020-human-development-report/data-readers-guide.

Wang, Y., Chen, C. R., & Huang, Y. S. (2014). Economic Policy Uncertainty and Corporate Investment: Evidence From China. *Pacific-Basin Finance Journal*, *26*, 227–243. https://doi.org/10.1016/j.pacfin.2013.12.008

Wu, J., Zhang, J., Zhang, S., & Zou, L. (2020). The Economic Policy Uncertainty and Firm Investment in Australia. *Applied Economics*, *52*(31), 3354–3378. https://doi.org/10.1080/00036846.2019.1710454

Zalla, R. (2017). Economic Policy Uncertainty in Ireland. *Atlantic Economic Journal*, *45*(2), 269–271. https://doi.org/10.1007/s11293-017-9536-8

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Breaking the Glass Ceiling: The Effects on Perceived Equality of Women Joining the Management Board

Jan Patrick Tollkühn

Bocconi University

Abstract

Recent years have seen women being appointed to board positions at the highest rate ever, supported by awareness of gender inequality and increased regulatory pressure. While external effects of diversity have been explored, changes in internal perceptions remain unexplored. Using a sample of 76,240 ratings on the employer review site Kununu, this study explores the impact of women joining the management board of DAX40 firms for the first time in Germany. To test the effect, a staggered event study with time and firm fixed effects is utilised. The results show that there is an increase in gender equality perception in the quarters following the appointment, with the increase being most significant in the second and third quarters. The identification of non-significant results in the fourth quarter suggests a normalisation effect. The findings outline how changes in leadership shape internal perceptions of equality.

Keywords: board gender diversity; employee reviews; employee perceptions; corporate governance; event study

1 Introduction

Within the Environmental, Social, and Governance (ESG) framework, the "Social" dimension is an often-underexplored component. The undefined nature of the dimension makes it difficult for firms to move from the do no significant harm principle towards substantial contributions across the value chain (Becchetti et al., 2022).

Diversity, as one of the core elements of the "Social" dimension of ESG, has gathered increased interest from stakeholders, regulators, and shareholders (Directive 2022/2381, 2022). Momentum regarding ESG principles varies considerably across regions; nonetheless, they remain relevant to corporate governance discussions globally[1].

Within the "Social" pillar of ESG, gender diversity within firms and especially corporate leadership has emerged as a significant area of research as academics and policymakers increasingly examine how leadership composition affects firm outcomes (Chytis et al., 2024; Yahya, 2023). While a significant amount of literature has been devoted to exploring the effects of board diversity on financial performance and other firm-level outcomes, little is known about the effects of changes in board-level diversity on employee perceptions within organisations. Yet internal perceptions of the management board and Chief Executive Officer (CEO) are significant in shaping organisational culture, employee engagement, and ultimately affecting firm outcomes (Lin et al., 2024; Mölders et al., 2019; Wang et al., 2017).

This research takes place during a rise of digital transparency. Digital platforms reduce information asymmetry by allowing individuals to share their experiences and giving light to previously unobserved areas of institutions. This digitalisation has enabled crowdsourced insights into firms and their practices.

One such digital platform is Robinhood, which has set out to "democratise finance" (Robinhood Markets, Inc., 2025a). Robinhood, as a retail investor trading platform, offers "commission-free trading" on various financial products (Robinhood Markets, Inc., 2025b). Through cost-free access they achieve "democratisation," as investing is accessible to the public.

While this transition is happening within financial markets, similar trends of digitalisation and information accessibility are seen in other areas. One such area is the emergence of employer review sites, which allow current and former employees to inform future applicants about the employer. While this information is relevant to possible future applicants, it has also presented a novel, unfiltered source of information regarding the corporate culture of firms and the effects these have on employees.

Examples of employer review sites include Glassdoor and

I also want to express my sincere gratitude to my thesis supervisor, Professor Roberto Vincenzi, for granting me the freedom to explore my topic and for his guidance throughout my research. Further, I want to thank Saskia Spang for her valuable feedback during my studies.

[1] Despite the momentum of ESG-principles, it is noteworthy that Diversity, Equity and Inclusion initiatives have experienced pushback in some jurisdictions, notably in the United States. This pushback follows political developments in the United States challenging ESG agendas (Iacurci, 2025; Kishan & Brogger, 2025; Sorkin et al., 2025; Yu, 2025).

DOI: https://doi.org/10.5282/jums/v11i3pp551-560

© The Author(s) 2026. Published by Junior Management Science.
This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.

Kununu. Glassdoor operates globally, having accumulated over 212 million reviews for 2.5 million employers (Glassdoor LLC., 2025). Kununu is a similar platform to Glassdoor; however, focusing on firms in Germany, Austria, and Switzerland. It hosts 13 million reviews for just 10,000 employers (Kununu, 2025).

Using data from the employer review site Kununu, this thesis investigates the effect on perceived gender equality of the first woman joining the management board within DAX40 firms in Germany. The DAX40 is the leading stock index in Germany, tracking the performance of the largest listed 40 German companies (Stoxx Ltd., 2025).

Germany presents a particularly good setting for this investigation due to their dual board structure consisting of a management board (Vorstand) and supervisory board (Aufsichtsrat). The management board is responsible for managing day-to-day activities, while the supervisory board consists of shareholder and employee representatives and is obligated to monitor the management board on an ongoing basis (Strenger et al., 2023). The dual structure allows an examination of the effects of management board members who are directly involved in operational decision-making.

This study opens with reviewing existing literature focusing on the information potential of employee reviews and the effects of board diversity on firms. From this, it will describe the constructed sample, leading over into the empirical investigation of the research question. A discussion of the empirical results and a conclusion, outlining the implications of the results and future avenues of research, will finish the paper.

2 Literature Review and Hypothesis Development

2.1 Wisdom of Crowds and Employee Reviews

Wisdom of crowds refers to the notion that group judgements "can be valuable guides in reaching business and investment decisions" (Mennis, 2006, p. 63). The notion was first presented by Surowiecki (2004), who identified the phenomenon that averaging independent judgements on a topic can yield better results than the decision of an individual expert.

In recent years, the concept of "wisdom of crowds" has been extended to new areas of organisational research, such as finance and accounting. This was furthered by an increased information base about individual companies by insiders, such as employees, through electronic word of mouth (Gimpl, 2025). Online employer review platforms such as Kununu and Glassdoor played pivotal roles in allowing analyses of insider information.

Symitsi et al. (2018) use Glassdoor reviews to explore the relationship between US public firm stock returns and employee satisfaction. The paper finds a positive relationship between employee satisfaction and firm performance. Notably, the research also finds abnormal equity returns. These are returns exceeding asset pricing models. The abnormal returns suggest that information about employee satisfaction has not been fully priced in by market participants.

Taking a European perspective, Becker et al. (2022) examine how employee satisfaction measured via Glassdoor reviews relates to stock returns during COVID-19. Becker et al. (2022) focus on firms within the Russell 3000 index. Findings by them suggest that firms that receive higher employee review scores during COVID-19 experience higher stock returns.

Analysing 1.5 million Glassdoor reviews, Brown et al. (2025) use the reviews as proxies of corporate culture within firms and relate them to the accuracy of management forecasts. To measure the information environment within firms, a textual analysis is conducted. Brown et al. (2025) conclude that management forecasts are more accurate in corporate culture settings when employees perceive a stronger organisational culture and better communication channels.

2.2 Board Diversity and Effects on the Firm

2.2.1 Financial Performance

The financial performance of firms is amongst the most extensively studied effects of board diversity, albeit yielding mixed results.

Carter et al. (2003) are amongst the earliest to examine the relationship between board diversity and firm value for Fortune 1000 firms. In their research, they find that diversity, measured by the presence of women or minorities, has a positive association with Tobin's Q, which serves as a proxy for firm value.

Subsequent research by Carter et al. (2010) finds a result contrary to the previous study in 2003 while investigating firms in the S&P 500. They find that there is no statistically significant positive relationship between the financial performance of the firms as measured by Tobin's Q or Return on Assets (ROA) and board membership of women or minorities. While this paper highlights the absence of a positive relationship, it also notes that there is no negative relationship to firm performance.

Research by Campbell and Mínguez-Vera (2008) explores the same relationship in the context of Spain. Their findings report positive effects on firm performance, measured by an approximation for Tobin's Q, for increased board diversity measured by the Blau and Shannon indices. Campbell and Mínguez-Vera (2008) argue that their results suggest that "increased gender diversity can be achieved without destroying shareholder value" (Campbell & Mínguez-Vera, 2008, p. 447). They also predict that positive discrimination to boost female presence in boardrooms would likely continue.

Adams and Ferreira (2009) explore the impact of women within boardrooms while laying a focus on the role of corporate governance mechanisms within firms. They find that gender-diverse boards are associated with stronger monitoring efforts. Adams and Ferreira (2009) reach two conclusions in their paper. Firstly, gender diversity can enhance firm performance when increased monitoring is needed and corporate governance is weak. On the other hand, it may destroy firm value in the setting of good corporate governance through excessive oversight.

Post and Byron (2015) conduct a meta-analysis of 140 studies, finding mixed results. On the one hand, they find no statistically significant relationship between market-based metrics, such as Tobin's Q and female board membership, while also finding that countries with higher gender parity have a positive relationship and in countries with less gender parity the relationship may turn negative. Post and Byron (2015) conclude this to be attributed to differences in reactions by investors.

Taken together, existing literature suggests that increased gender diversity within boardrooms does not necessarily increase firm financial performance. However, given certain conditions such as weak governance frameworks, gender diversity on boards may contribute positively to firm value.

2.2.2 Innovation

Research related to board diversity has extended beyond financial performance and encapsulates other areas such as innovation. Innovation is critical to ensure the long-term viability and competitiveness of firms within markets (Porter, 1992).

J. Chen et al. (2018) find that the presence of female board members is associated with greater innovation, measured using patent count and patent citations. Their empirical research also finds that these positive effects are associated with improved firm performance in innovation-intensive industries.

Expanding upon this, I.-J. Chen et al. (2023) separate board diversity into inherent, acquired, and aggregated board diversity. I.-J. Chen et al. (2023) define inherent diversity as characteristics based on demographic features, while acquired diversity relates to education, age, and expertise. Exploring this, I.-J. Chen et al. (2023) find that there were significant associations of all three types of diversity with innovation output measured by citations and patents.

The findings by I.-J. Chen et al. (2023) also point out that higher acquired board diversity is attributed to higher research and development expenses, which was not found to be the case for inherent board diversity.

This distinction suggests that inherent diversity may facilitate the development of existing innovation capabilities, while acquired diversity may directly enhance innovation investment and resource allocation.

2.2.3 Risk-Taking

Empirical studies strongly suggest that board diversity can significantly influence risk levels within firms. Psychological studies have also illustrated that women, on average, exhibit lower levels of risk appetite (Byrnes et al., 1999; Dawson, 2023). Taking this into account, diversity on corporate boards can therefore level risk-taking by corporates.

Karavitis et al. (2021) showed this concept implicitly through exploring changes in loan spreads attributed to female directors on the board of directors. Analysing bank loan data, a reduction in loan spread by 20.8 basis points was identified for every one percent increase in the proportion of female board directors. Since loan spreads reflect a lender's risk assessment of the firm, this implies that markets perceive firms with greater gender diversity as less risky.

From the reduction in loan spread, it can be concluded that the presence of female directors reduces risk-taking activities by firms. Bernile et al. (2018) explore the effects of board diversity on firm risk, as measured by volatility of stock returns, using a sample of 21,572 firm-year observations. Bernile et al. (2018) explore the concepts of diversity further, focusing not exclusively on gender diversity, but constructing a multidimensional index to capture diversity. In the empirical results, it is found that higher degrees of board diversity cause less risky financial policy adoptions by boards. Evidence of this behaviour is further reflected through higher dividend payout ratios and lower leverage ratios, which are typically interpreted as indicators of reduced financial risk exposure (Bernile et al., 2018).

Faccio et al. (2016) empirically examine the role of gender at the CEO position in relation to firm leverage. Their research finds a highly significant negative association, suggesting that firms led by a female CEO adopt less aggressive financial structures. Looking at board composition, Adams and Ferreira (2009) find that women have higher board meeting attendance rates than their male counterparts. Within their research, they also find that more gender-diverse boards have a higher likelihood of CEO turnover when stock performance is bad, pointing to them being more risk averse and displaying a more disciplined approach to CEO oversight.

Disclosure literature also indicates positive effects of board diversity, including gender diversity. Seebeck and Vetter (2022) explore the impact of board gender diversity on risk disclosure related to Brexit in the United Kingdom. Their research finds that gender board diversity has a positive effect on corporate risk disclosures. Albeit a critical mass of 20 to 40% female directors must be reached before positive results take effect.

Relating to financial disclosures, Abbott et al. (2012) finds a significant relationship between having at least one woman as a board member and the likelihood of financial restatements. Abbott et al. (2012) conclude that the presence of female board members likely reduces the pace of decision-making and allows for more viewpoints to be communicated and considered. From this, it can be concluded that this will reduce risk to the firm and shareholders.

On the other hand, Adams and Funk (2012) outline findings contrary to the narrative that female board directors are more risk averse. Studying gender differences in Swedish boardrooms, Adams and Funk (2012) find that female board members are less concerned with power but exhibit greater risk tolerance. Therefore, they conclude that having female board members may even lead to increased risk-taking.

Overall, conflicting results are identified by various researchers in different settings. Therefore, it remains unclear whether there is an association between firm risk-taking and diversity at the board level.

2.3 Hypothesis Development

While prior research has focused on the relationship between gender board diversity and various aspects such as firm financial performance, innovation, and risk-taking, the impact on employee perceptions has not been fully explored, especially how such board-level diversity affects employee perception of gender equality. In this study, perceived gender equality refers to an employee's subjective belief that men and women have equal opportunities, are fairly treated, and are included in decision-making processes (Ely & Thomas, 2001; Nishii, 2013).

Perceptions of gender equality positively affect employee commitment to an organisation by increasing retention and their engagement. In addition, firms that are perceived as more inclusive are able to attract better talent (Dauth et al., 2023; Dixon-Fyle et al., 2020).

According to signalling theory, changes to the management board communicate not only internal priorities but also changes in values. For example, a woman being appointed for the first time could illustrate that an emphasis is being placed on better inclusion and gender equality (Connelly et al., 2011).

Symbolic management theory further suggests that organisational actions serve a symbolic function with the aim of sending messages to internal and external stakeholders. Even if the signals are not immediately connected to operational modifications, they may result in changes to how an organisation is perceived (Westphal & Park, 2020). As employees are internal stakeholders closely connected to the business, they are likely to be influenced by a change in leadership, particularly if such a significant change like a woman joining the management board for the first time is taking place. Considering this, employee review platforms such as Kununu allow researchers to derive meaningful insights into changes in employee perceptions of leadership transitions. Based on this rationale, the following hypothesis is developed.

> H1. The first appointment of a woman to a firm's management board is associated with a significant increase in employee-perceived gender equality.

3 Data and Sample

3.1 Sample Construction

The sample for this study consists of exchange-listed German firms that were members of the leading German stock index. For the time period between the first quarter of 2012 and the first quarter of 2024, 53 firms were identified that were members of the DAX30 and later the DAX40.

Changes in the composition of the index are not relevant due to the inclusion of all companies that were members of the stock index during the sample period.

3.2 Employee Review Data

To capture perceived equality, data from Kununu.com was used. Kununu is an employee review site where employees are encouraged to publish reviews about their current and former employers. The employee reviews focus on companies in Germany, Austria, and Switzerland.

Users of Kununu can provide numerical ratings and publish text reviews on a variety of topics regarding their working environment, including work-life balance, the image of the employer, and many other topics. Among these different categories of rating, this study will focus on the rating regarding "Gleichberechtigung" (gender equality).

To collect all the relevant data from Kununu, a web-scraping Python program was developed that extracted all the reviews for the relevant companies. Accessing these companies, the program captured a total of 76,240 individual employee reviews between the first quarter of 2012 and the first quarter of 2024.

The variable that is investigated within this thesis reflects employee perceptions on gender equality within their respective company. Kununu allows the employee to use a five-point Likert scale, where 5 illustrates a high degree of equality and 1 corresponds to a complete lack of equality. The study will use the same five-point Likert scale to investigate the effects of the first woman joining the management board.

The Kununu data is aggregated to reflect quarterly average ratings. If there are no reviews for a quarter, the missing datapoint is filled using a backwards-looking moving average from the past four quarters. If this is not possible using four quarters, the same is repeated with three, with two, and finally the previous quarter. The data point is left blank if this method cannot be applied.

The backwards-filling methodology is used especially in earlier time periods where employee reviews are limited due to a lack of awareness about employer review sites.

Furthermore, the scores about communication and total employee satisfaction are also extracted from Kununu. The communication score reflects employee perceptions of clarity, openness and effectiveness of internal communication within the company. The total score is an aggregate measure calculated by Kununu summarising overall employee satisfaction across multiple areas of workplace culture, including evaluations of how interesting tasks are and career development opportunities.

3.3 Firm Board Data

Previous research uses data on board diversity and composition from BoardEx, ExecuComp, or other databases (Bernile et al., 2018; Doan & Iskandar-Datta, 2020; Ye et al., 2019). Historical coverage of German firms in these databases proved to be limited and inaccurate, especially with regard to executive board size and composition.

To ensure data reliability, this analysis reviewed all annual reports of the selected companies within the sample and retrieved data on management board size and composition according to gender. Information regarding the gender of

the CEO was also retrieved. Gender was tracked using photos and the name of the management board member available in the annual report.

Management board composition was tracked on a monthly basis but later aggregated to a quarterly frequency to match the structure of the employee review data.

Clear definitions of the variables explored in this analysis are given in Table 1.

4 Methodology and Empirical Results

4.1 Methodology

From my hypotheses in 2.3, this study employs a staggered event study framework to quantify the effects on employee-perceived equality when the first woman joined the management board in a DAX40 company. A staggered framework is needed due to differences regarding the time when the first appointment of a female board member took place.

Event studies are widely used in economics and finance contexts (MacKinlay, 1997). Currie et al. (2020) show a large increase of event study frameworks in National Bureau of Economic Research Working Papers and in the top five research journals after 2010[2].

A balanced panel dataset is constructed for each firm meeting the requirements listed in 3.1 for the sample time period of the first quarter of 2012 to the first quarter of 2024. Firms that have female management board members throughout the entire period are excluded to allow for clearer identification of changes in employee-perceived equality (Gormley & Matsa, 2014).

For each firm-treatment, the quarter in which the first woman joins the management board is designated as event time $k = 0$. Post-treatment effects may not materialise immediately or may vary in their significance. The analysis focuses on the four quarters before and after the treatment period ($k \in [-4, +4]$). Conducting the analysis with this methodology enables the estimation of anticipatory effects and delayed responses in employee-perceived equality. Using these estimators allows for a clear examination of the parallel trend assumption, which is central to event studies and difference-in-differences designs (Borusyak et al., 2024; Schmidheiny & Siegloch, 2023).

The dependent variable in this analysis is the equality score attributed to a firm in a particular quarter. The regression model to capture the effects on employee-perceived equality of a woman joining the management board for the first time is the following:

$$\text{Equality_score}_{it} = \beta_0 + \sum_{\substack{k \in [-4,\ +4] \\ k \neq -1}} \beta_k \times \text{EventTime}_{kit} + \gamma_i + \delta_t + \mathbf{X}_{it} + \varepsilon_{it}$$

where EventTime_{kit} denotes dummy variables for the quarters prior to and after the treatment. The variable k defines the relative distance to the treatment quarter, with $k = 0$ being the quarter where the treatment takes place. The period $k = -1$ is excluded as it serves as the reference period (Huntington-Klein, 2021). Firm fixed effects are controlled with the γ_i, and refer to firm-specific time-invariant effects such as industry. Quarterly time fixed effects are accounted for using the δ_t. Time fixed effects absorb time-varying shocks such as governmental regulation or macroeconomic shocks. The vector $\mathbf{X}_{it}$ controls for other firm-specific metrics, such as board size or other employee-specific metrics, such as overall employee satisfaction and employee perceptions of internal communication. The coefficient β_k is the estimator of the regression tracking the average treatment effects on the $\text{Equality_score}_{it}$.

The model is estimated using Ordinary Least Squares with time and firm fixed effects. Standard errors are clustered at the firm level to address autocorrelation and heteroskedasticity within firms over time.

4.2 Summary Statistics

4.2.1 Descriptive Statistics

Table 2 presents the descriptive statistics for the main variables used in the analysis. There are 364 firm-quarter observations considered in the analysis.

The dependent variable Equality_score has a mean of 3.76, indicating a generally positive opinion about gender equality within firms over time. The standard deviation of 0.70 indicates that there is moderate variability between firms over time.

With regard to board structure, during the observations, management boards of DAX40 companies had on average 5.64 board members (Total_board_count), with a standard deviation of 2.17 indicating variability between firms. For this same period, only 0.6 women were on average management board members (Female_board_count). The mean of the variable Female_ceo indicates that for the observation period only 1% of the time a CEO of the firm was also female. Overall, both variables suggest that female corporate leadership in German DAX40 firms remains rare.

Total_score is an overall indication of employee satisfaction with their employer. For the time observations selected, it had a mean of 3.61 with a standard deviation of 0.59. Reviews concerning internal company communication (Comm_score) are slightly lower at 3.26. The variable Log_review_count is the natural logarithm of the number of reviews per quarter per firm. It exhibits a mean of 2.56 with a standard deviation of 1.06. This indicates that the number of reviews varies significantly between firms and quarters.

4.2.2 Correlation Analysis

Table 3 displays Pearson pairwise correlation between the variables included in the analysis. Equality_score displays a strong and positive correlation with Total_score (0.78) and Comm_score (0.68). This indicates that perceptions of gender equality are closely linked to employee perceptions about

[2] With regard to the "top-five journals", Currie et al. (2020) consider the following: *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*.

company communication and overall satisfaction with the employer.

There is neither a positive nor negative relationship between Equality_score and board characteristics (Total_board _count: -0.02, Female_board_count: -0.01, Female_ceo: -0.03). This shows that there is no linear relationship between board characteristics and Equality_score.

It must also be noted that Total_score and Comm_score are themselves highly correlated to each other (0.88). This strong correlation raises concerns regarding multicollinearity problems in regressions, if both are used as control variables. However, this is not applicable here, as both are used separately.

4.3 Regression Results

This section of the thesis presents the results of the regressions calculated. The regression estimates the effects of the first woman joining the management board on employee perception of gender equality. Table 4 reports the coefficient estimates and standard errors for the quarters surrounding the appointment ($k \in [-4, +4]$, $k \neq -1$). All models include firm and time fixed effects.

4.3.1 Baseline Specifications

Model 1 begins with a regression that has no controls other than firm and time fixed effects. The coefficient estimate following the event is positive and statistically significant following $k = +2$ ($\beta_2 = 0.2317$, $p_2 < 0.1$). At $k = +3$ the effect remains positive and significant ($\beta_3 = 0.2275$, $p_3 < 0.1$), however, at $k = +4$ the effect stops being significant but stays positive ($\beta_4 = 0.2044$).

Model 2 continues with adding Total_board_count as a control variable, thereby controlling for board size. Prior studies, such as those by Bernile et al. (2018), Carter et al. (2003), and Post and Byron (2015), have used board size as a control variable for analyses regarding board composition and firm outcomes. The addition of this control variable strengthens the significance of $k = +2$ ($\beta_2 = 0.2952$, $p_2 < 0.05$) and $k = +3$ ($\beta_3 = 0.2828$, $p_3 < 0.05$). At $k = +4$ the effect is considerably higher than in Model 1 ($\beta_4 = 0.2634$). Although the effect is higher it remains not statistically significant.

Model 3 adds another board characteristic, Female_ceo. The coefficient estimators remain similar to the estimators in Model 2 ($\beta_2 = 0.3042$, $p_2 < 0.05$; $\beta_3 = 0.2908$, $p_3 < 0.05$). Notably, the coefficient from Female_ceo is negative; however, it is not statistically significant.

4.3.2 Including Other Employee Perceptions

Other employee perceptions are added into Models 4 and 5 to account for the possibility that the effect on employee perceptions of equality is driven by a change in other employee perceptions.

Model 4 drops the control variable, Female_ceo, and replaces it with Comm_score. The removal is due to the low variation of this variable within the sample period, which would limit its interpretability within the model. The effects at $k = +2$ and $k = +3$ lose their significance but remain positive ($\beta_2 = 0.1114$; $\beta_3 = 0.1684$). The effect during the other post quarters is also positive. Comm_score is highly significant, illustrating a positive effect on the Equality_score ($\beta_{\text{Comm_score}} = 0.5556$, $p_{\text{Comm_score}} < 0.01$).

Model 5 removes the Comm_score and replaces it with Total_score to capture possible relations with overall employee satisfaction. Time effects are still positive but remain not statistically significant. At time $k = +2$ and $k = +3$, the effects are smaller than in the previous models ($\beta_2 = 0.0760$; $\beta_3 = 0.0571$). The effect of Total_score is larger than the effect of Comm_score, suggesting that employee perceptions of equality are closely linked to their total satisfaction ($\beta_{\text{Total_score}} = 0.8864$, $p_{\text{Total_score}} < 0.01$).

Caution must be exercised when interpreting the results of Models 4 and 5. High correlations are observed between the variables Comm_score and Total_score, as seen in Table 3. There are also high correlations with Equality_score. Together this raises multicollinearity concerns. While Comm_score and Total_score are modelled separately to avoid multicollinearity, their thematic similarity limits the ability to isolate their individual effects.

Model 6 removes employee perceptions but instead controls for the number of reviews for each firm-quarter. The variable Log_review_count is not statistically significant, yet is positive, indicating a positive effect on gender equality perceptions ($\beta_{\text{Log_review_count}} = 0.0705$). The event time coefficients are positive and statistically significant for some of the quarters following the first woman joining the management board. The coefficients are $\beta_2 = 0.2850$ ($p_2 < 0.05$) and $\beta_3 = 0.2675$ ($p_3 < 0.05$).

4.3.3 Constant Term

The constant represents the baseline of the Equality_score that is achieved when all independent variables are equal to zero and only fixed effects are controlled for.

Across Models 1 to 3 the constant is highly statistically significant ($p < 0.001$). The coefficients are 2.7951, 3.5655 and 3.6372 for the Models 1 to 3, respectively.

Models 4 and 5 see a reduction in the size of the constant. Model 4 has a baseline equality of 2.1535 and Model 5 has only 0.6792. The statistical significance of the constant also changes; whilst in Model 4 the constant is still highly significant ($p < 0.001$), the constant in Model 5 loses all statistical significance. Model 6 regains this statistical significance ($p < 0.001$) and a constant value of 3.4467.

4.3.4 Model Fit and Model Error

The bottom of Table 4 includes information about the model fit and the statistical significance of the model.

Adjusted R^2 indicates the proportion of variance explained by the regression while accounting for the number of variables used in the regression. The lowest Adjusted R^2 is achieved by Model 3 with 0.344. Model 5 which includes the control variables Total_board_count and Total_score

achieved the highest Adjusted R^2 with 0.650. When replacing Total_score with Comm_score Adjusted R^2 falls to 0.546. In conclusion, Model 5 shows the best fit due to having the highest Adjusted R^2.

All models are highly statistically significant as presented by the probability of their F-Statistics. Each of the models presents these through p-values below 0.001. From this, the null hypothesis is rejected. The null hypothesis in this case is that all of the coefficients, excluding the intercepts/constant, are equal to zero.

All models are robust through the inclusion of time and firm fixed effects. This helps account for shocks, as well as heterogeneity across firms. The use of these fixed effects strengthens the robustness of the models by mitigating omitted variable bias.

Table 4 also presents Durbin-Watson statistics to ensure no autocorrelation. Durbin-Watson statistics are bound to the range $[0, 4]$, with 2 indicating no autocorrelation of the residuals. Although the models have a slightly negative autocorrelation, the size of these are in the range of 2.148 and 2.359, indicating that there is little to no autocorrelation.

5 Discussion

5.1 Interpretation of Empirical Findings

This thesis investigated whether the appointment of a woman for the first time to the management board of a German DAX40 company results in an increase in perceived equality of employees. To investigate this, the study draws upon a staggered event study methodology with time and firm fixed effects. The results of the regression reveal a modest, statistically significant increase in perceived gender equality.

The positive effect is observed at $k = +2$ and $k = +3$. The results are significant at levels between $p_2 < 0.1$ and $p_3 < 0.05$. Whilst the coefficients at $k = +4$ are still positive, they are not statistically significant.

Throughout most models, the time effects post-treatment ($k = +1$, $k = +2$, $k = +3$ and $k = +4$) are positive. This provides evidence that the effect of appointing a woman to the management board for the first time may result in a positive increase in equality perceptions.

The effect at $k = +1$ is not statistically significant in any of the models. This may reflect a lag in how employees respond to changes in leadership composition. According to symbolic management theory, employees notice the change to the management board. However, they are hesitant to react to these changes before other substantive measures supporting gender equality are taken, initiated by the appointment of a woman (Westphal & Park, 2020).

The reduction in size of the positive coefficient in period $k = +4$ suggests that employees may experience a normalisation effect. Normalisation effects have been observed in the field of psychology. Adaptation-level theory suggests that a person will become insensitive to the effect of constant stimuli (Bowling, 2014; Helson, 1964). It can be argued that this is taking place here, with the effects of the woman being on the management board being the stimuli and employees no longer perceiving this effect.

Board characteristics such as board size and the CEO being female show no significant relationship with equality perceptions. However, post-treatment effects in Model 3 are stronger and more significant than in Model 1. This suggests that the symbolic impact of appointing a woman to the management board may influence employee perceptions of gender equality more than changes in structural board characteristics.

When employee sentiment variables regarding overall employee satisfaction and communication perception are added, the time effects lose their significance. This observation suggests that these characteristics may be driving forces behind the increase in perceived gender equality. However, this is not fully clear and should be interpreted with caution due to the high correlations between Total_score, Comm_score and Equality_score. When the employee sentiment controls are removed but Log_review_count is inserted the statistical significance at $k = +2$ and $k = +3$ persists.

To date, no prior study has explored the relationship between the first appointment of a woman to the management board and employee perceptions of gender equality.

This thesis adds to existing literature on organisational culture and employee perceptions. The results of this analysis show that employee review data facilitates observations of inner firm dynamics. This supports the existence of the "wisdom of crowds" phenomenon, which allows the capture of meaningful firm-level insights with regard to perceived gender equality (Gimpl, 2025; Surowiecki, 2004).

The findings also add to the understanding of the effects of changes in leadership composition on employee perceptions. Previous research has focused almost exclusively on external effects, such as firm value, innovation, or risk-taking (Adams & Ferreira, 2009; Adams & Funk, 2012; I.-J. Chen et al., 2023).

5.2 Limitations of the Study

Despite this thesis offering novel insights, limitations to its methodology must be acknowledged.

Firstly, although the event study uses firm and time fixed effects in an effort to isolate the effect of a woman joining a management board for the first time, the findings remain correlational. Unobserved factors may have influenced the regression, increasing employee perceptions of equality. Reverse causality is also a concern (Godfrey et al., 2024). Reverse causality would see that firms with higher perceived equality perceptions are more likely to appoint women to their management board. However, this appears less likely, as no anticipatory increase in gender equality perceptions is observed within the models.

Another limitation is caused by the study relying on employee reviews. Employee reviews on Kununu are subjective

and therefore can easily introduce bias (Gimpl, 2025). Employees may also find it difficult to quantify equality perceptions, therefore having difficulties making accurate inputs. Especially if employees perceive their employer as not satisfactory, they may be inclined to select a low rating for all categories, not analysing each on their merits.

The sample selected for this thesis is also limited in its scope (Eisenhardt, 1989). With regard to geography, it is restricted to German firms included within the DAX40 index, thereby excluding smaller and non-listed firms. With regard to timeline, this thesis is quite expansive, as it focuses on a time when most firms first appointed a female board member. Using earlier data from Kununu may have resulted in a not sufficiently large information base.

Finally, the use of quarters within the sample also introduces limitations. These lack granularity, potentially missing anticipatory effects or post-treatment effects.

5.3 Future Research Avenues

The methodology of this paper could be adapted to utilise the most modern approaches of panel datasets. A difference-in-differences model could be used to explore this research question again, potentially yielding results with higher statistical significance, as the effects of no woman joining could be modelled with higher precision.

Utilising the Kununu dataset, the research question can be expanded to explore the effect of a woman joining the management board, regardless of her being the first. A follow-up study of this could lead to an evaluation of differences in size effects of a second woman joining compared to the first.

While the use of Kununu data is insightful, the dataset underlying the study could be expanded by using review data from Glassdoor instead of Kununu. This would allow for additional analysis of the effect of women joining the management board with a broader geographic scope.

Furthermore, one future research avenue is to use the text reviews on Kununu to derive a measurement of perceived equality. Brown et al. (2025) conducted a ChatGPT-based textual analysis. As their research focused on forecast precision, their ChatGPT model developed “a measure of a firm’s internal information environment” (Brown et al., 2025, p. 13). For the context of perceived equality, a measure of the firm’s gender equality environment could be developed.

Another possibility is to explore the impact of perceived equality on firm outcomes. For example, literature regarding the positive effect of diversity on firm value by Carter et al. (2003) can be reevaluated using perceived equality instead of measurable diversity.

6 Conclusion

This thesis examined the impact of the first appointment of a woman to the management board on employee perception on equality within DAX40 companies. Utilising a staggered event study framework with time and firm fixed effects and drawing on electronic word-of-mouth information from Kununu, this study identifies an increase in perceived equality, with the effect being strongest in the second and third quarters following the appointment.

While the effect is no longer significant in the fourth quarter, the results are robust across the models, allowing for the conclusion that women joining the management board do create measurable impacts on employee perceptions of gender equality. The effect is not explained by traditional board characteristics such as size and whether the CEO is female. This suggests that while appointments may be of a symbolic nature, they also have meaningful effects on employees.

These findings contribute to literature regarding the effects of board-level diversity. However, it explores internal effects, not external firm effects such as firm value or others. It also contributes to literature by illustrating how symbolic leadership changes affect internal perceptions.

For companies, these results yield practical implications. The results underscore the importance of gender-diverse boards and their effect on internal perceptions of fairness and equality. Firms should accompany the appointment of the first woman to the management board with continuous policy and procedural changes that are needed to have a lasting positive impact on gender equality perceptions within firms.

Future research should use textual analysis and text reviews to examine external effects of employee-perceived equality on firm performance, innovation, and risk-taking.

References

Abbott, L. J., Parker, S., & Presley, T. J. (2012). Female Board Presence and the Likelihood of Financial Restatement. *Accounting Horizons*, *26*(4), 607–629. https://doi.org/10.2308/acch-50249

Adams, R. B., & Ferreira, D. (2009). Women in the Boardroom and Their Impact on Governance and Performance. *Journal of Financial Economics*, *94*(2), 291–309. https://doi.org/10.1016/j.jfineco.2008.10.007

Adams, R. B., & Funk, P. (2012). Beyond the Glass Ceiling: Does Gender Matter? *Management Science*, *58*(2), 219–235. Retrieved May 6, 2026, from https://www.jstor.org/stable/41406385

Becchetti, L., Bobbio, E., Prizia, F., & Semplici, L. (2022). Going Deeper into the S of ESG: A Relational Approach to the Definition of Social Responsibility. *Sustainability*, *14*(15), 9668. https://doi.org/10.3390/su14159668

Becker, M., Cardazzi, A., & McGurk, Z. (2022). Employee Satisfaction and Stock Returns During the COVID-19 Pandemic. *Journal of Behavioral and Experimental Finance*, *33*, 100603. https://doi.org/10.1016/j.jbef.2021.100603

Bernile, G., Bhagwat, V., & Yonker, S. (2018). Board Diversity, Firm Risk, and Corporate Policies. *Journal of Financial Economics*, *127*(3), 588–612. https://doi.org/10.1016/j.jfineco.2017.12.009

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting Event-Study Designs: Robust and Efficient Estimation. *Review of Economic Studies*, *91*(6), 3253–3285. https://doi.org/10.1093/restud/rdae007

Bowling, N. (2014). Adaptation-Level Theory. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 28–29). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_25

Brown, A. B., Wang, V. X., & Zhou, A. (2025). Employee Perceptions of Corporate Culture and Management Forecast Accuracy: Evidence from Glassdoor and ChatGPT. *Accounting Horizons*, 1–20. https://doi.org/10.2308/HORIZONS-2023-142

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender Differences in Risk Taking: A Meta-Analysis. *Psychological Bulletin*, *125*(3), 367–383. https://doi.org/10.1037/0033-2909.125.3.367

Campbell, K., & Mínguez-Vera, A. (2008). Gender Diversity in the Boardroom and Firm Financial Performance. *Journal of Business Ethics*, *83*(3), 435–451. https://doi.org/10.1007/s10551-007-9630-y

Carter, D. A., D'Souza, F., Simkins, B. J., & Simpson, W. G. (2010). The Gender and Ethnic Diversity of US Boards and Board Committees and Firm Financial Performance. *Corporate Governance: An International Review*, *18*(5), 396–414. https://doi.org/10.1111/j.1467-8683.2010.00809.x

Carter, D. A., Simkins, B. J., & Simpson, W. G. (2003). Corporate Governance, Board Diversity, and Firm Value. *Financial Review*, *38*(1), 33–53. https://doi.org/10.1111/1540-6288.00034

Chen, I.-J., Lin, W. C., Lo, H.-C., & Chen, S.-S. (2023). Board Diversity and Corporate Innovation. *Review of Quantitative Finance and Accounting*, *61*(1), 63–123. https://doi.org/10.1007/s11156-023-01145-4

Chen, J., Leung, W. S., & Evans, K. P. (2018). Female Board Representation, Corporate Innovation and Firm Performance. *Journal of Empirical Finance*, *48*, 236–254. https://doi.org/10.1016/j.jempfin.2018.07.003

Chytis, E., Eriotis, N., & Mitroulia, M. (2024). ESG in Business Research: A Bibliometric Analysis. *Journal of Risk and Financial Management*, *17*(10), 460. https://doi.org/10.3390/jrfm17100460

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling Theory: A Review and Assessment. *Journal of Management*, *37*(1), 39–67. https://doi.org/10.1177/0149206310388419

Currie, J., Kleven, H., & Zwiers, E. (2020). Technology and Big Data Are Changing Economics: Mining Text to Track Methods. *AEA Papers and Proceedings*, *110*, 42–48. https://doi.org/10.1257/pandp.20201058

Dauth, T., Schmid, S., Baldermann, S., & Orban, F. (2023). Attracting Talent Through Diversity at the Top: The Impact of TMT Diversity and Firms' Efforts to Promote Diversity on Employer Attractiveness. *European Management Journal*, *41*(1), 9–20. https://doi.org/10.1016/j.emj.2021.10.007

Dawson, C. (2023). Gender Differences in Optimism, Loss Aversion and Attitudes Towards Risk. *British Journal of Psychology*, *114*(4), 928–944. https://doi.org/10.1111/bjop.12668

Directive 2022/2381. (2022). Directive (EU) 2022/2381 of the European Parliament and of the Council of 23 November 2022 on Improving the Gender Balance Among Directors of Listed Companies and Related Measures (Text with EEA Relevance). Retrieved May 6, 2026, from http://data.europa.eu/eli/dir/2022/2381/oj

Dixon-Fyle, S., Dolan, K., Hunt, D. V., & Prince, S. (2020). *Diversity Wins: How Inclusion Matters* (tech. rep.). McKinsey & Company. Retrieved July 24, 2025, from https://www.mckinsey.com/featured-insights/diversity-and-inclusion/diversity-wins-how-inclusion-matters%5C#/

Doan, T., & Iskandar-Datta, M. (2020). Are Female Top Executives More Risk-Averse or More Ethical? Evidence from Corporate Cash Holdings Policy. *Journal of Empirical Finance*, *55*, 161–176. https://doi.org/10.1016/j.jempfin.2019.11.005

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, *14*(4), 532. https://doi.org/10.2307/258557

Ely, R. J., & Thomas, D. A. (2001). Cultural Diversity at Work: The Effects of Diversity Perspectives on Work Group Processes and Outcomes. *Administrative Science Quarterly*, *46*(2), 229–273. https://doi.org/10.2307/2667087

Faccio, M., Marchica, M.-T., & Mura, R. (2016). CEO Gender, Corporate Risk-Taking, and the Efficiency of Capital Allocation. *Journal of Corporate Finance*, *39*, 193–209. https://doi.org/10.1016/j.jcorpfin.2016.02.008

Gimpl, N. (2025). The Wisdom of Electronic Employee Crowds—Employee Reviews as a Data Source in Finance, Accounting, Economics, and Management Research: A Systematic Literature Review. *Intelligent Systems in Accounting, Finance and Management*, *32*(2), e70007. https://doi.org/10.1002/isaf.70007

Glassdoor LLC. (2025). About Us | Glassdoor. Retrieved July 15, 2025, from https://about-us.glassdoor.com/site-us/

Godfrey, C., Hoepner, A. G. F., Lin, M.-T., & Poon, S.-H. (2024). Women on Boards and Corporate Social Irresponsibility: Evidence from a Granger Style Reverse Causality Minimisation Procedure. *The European Journal of Finance*, *30*(1), 1–27. https://doi.org/10.1080/1351847X.2020.1841664

Gormley, T. A., & Matsa, D. A. (2014). Common Errors: How to (and Not to) Control for Unobserved Heterogeneity. *Review of Financial Studies*, *27*(2), 617–661. https://doi.org/10.1093/rfs/hht047

Helson, H. (1964). *Adaptation-Level Theory*.

Huntington-Klein, N. (2021). Chapter 17—Event Studies. In *The Effect: An Introduction to Research Design and Causality* (First, pp. 407–434). Chapman and Hall/CRC. Retrieved July 15, 2025, from https://theeffectbook.net/ch-EventStudies.html

Iacurci, G. (2025). 'game over' for esg investing due to trump backlash? analysts say no [Section: Personal Finance]. Retrieved July 14, 2025, from https://www.cnbc.com/2025/03/31/trumps-backlash-isnt-game-over-for-esg-investing.html

Karavitis, P., Kokas, S., & Tsoukas, S. (2021). Gender Board Diversity and the Cost of Bank Loans. *Journal of Corporate Finance*, *71*, 101804. https://doi.org/10.1016/j.jcorpfin.2020.101804

Kishan, S., & Brogger, T. H. (2025). Why ESG Faces Backlash and Its Future Under Trump 2.0. *Bloomberg.com*. Retrieved July 14, 2025, from https://www.bloomberg.com/news/articles/2025-03-03/why-esg-faces-backlash-under-trump-2-0

Kununu. (2025). Arbeitgeber bewerten Finde deinen besten Arbeitgeber. Retrieved July 15, 2025, from https://www.kununu.com/

Lin, S., Zhang, Y., Wang, Y., & Geng, H. (2024). The Impact of CEO Reputation on ESG Performance in Varying Managerial Discretion Contexts. *Finance Research Letters*, *70*, 106368. https://doi.org/10.1016/j.frl.2024.106368

MacKinlay, A. C. (1997). Event Studies in Economics and Finance. *Journal of Economic Literature*, *35*(1), 13–39. Retrieved July 15, 2025, from https://www.jstor.org/stable/2729691

Mennis, E. A. (2006). Review of The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations. *Business Economics*,

41(4), 63–64. Retrieved July 12, 2025, from https://www.jstor.org/stable/23490286

Mölders, S., Brosi, P., Spörrle, M., & Welpe, I. M. (2019). The Effect of Top Management Trustworthiness on Turnover Intentions via Negative Emotions: The Moderating Role of Gender. *Journal of Business Ethics*, *156*(4), 957–969. https://doi.org/10.1007/s10551-017-3600-9

Nishii, L. H. (2013). The Benefits of Climate for Inclusion for Gender-Diverse Groups. *Academy of Management Journal*, *56*(6), 1754–1774. https://doi.org/10.5465/amj.2009.0823

Porter, M. E. (1992). Capital Disadvantage: America's Failing Capital Investment System. *Harvard Business Review*, *70*(5), 65–82.

Post, C., & Byron, K. (2015). Women on Boards and Firm Financial Performance: A Meta-Analysis. *The Academy of Management Journal*, *58*(5), 1546–1571. Retrieved July 11, 2025, from https://www.jstor.org/stable/24758233

Robinhood Markets, Inc. (2025a). About Us. Retrieved July 14, 2025, from https://robinhood.com/us/en/about/legal/

Robinhood Markets, Inc. (2025b). Explore US Stock and ETF Tokens & Crypto | Robinhood EU. Retrieved July 15, 2025, from https://robinhood.com/eu/de/

Schmidheiny, K., & Siegloch, S. (2023). On Event Studies and Distributed-Lags in Two-Way Fixed Effects Models: Identification, Equivalence, and Generalization. *Journal of Applied Econometrics*, *38*(5), 695–713. https://doi.org/10.1002/jae.2971

Seebeck, A., & Vetter, J. (2022). Not Just a Gender Numbers Game: How Board Gender Diversity Affects Corporate Risk Disclosure. *Journal of Business Ethics*, *177*(2), 395–420. https://doi.org/10.1007/s10551-020-04690-3

Sorkin, A. R., Mattu, R., Warner, B., Kessler, S., Merced, M. J. d. l., & Hirsch, L. (2025). D.E.I. on the Ropes Ahead of the Next Trump Era. *The New York Times*. Retrieved July 14, 2025, from https://www.nytimes.com/2025/01/17/business/dei-trump-meta-zuckerberg.html

Stoxx Ltd. (2025). DAX. Retrieved July 15, 2025, from https://stoxx.com/index/dax/

Strenger, C., Redenius-Hövermann, J., & Demirtaş, G. (2023). The Governance System of Germany: Background and Discussion of Its Code. *Frankfurt School: Coroprate Governance Institute Working Paper*. Retrieved July 14, 2025, from https://www.frankfurt-school.de/en/dam/jcr:dfe1ae1d-57ca-40f5-9c6f-397a2ff5d909/CGI_WP_GovernanceSystemOfGermany_20230727.pdf

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*. Doubleday.

Symitsi, E., Stamolampros, P., & Daskalakis, G. (2018). Employees' Online Reviews and Equity Prices. *Economics Letters*, *162*, 53–55. https://doi.org/10.1016/j.econlet.2017.10.027

Wang, D., Feng, T., & Lawton, A. (2017). Linking Ethical Leadership with Firm Performance: A Multi-Dimensional Perspective. *Journal of Business Ethics*, *145*(1), 95–109. https://doi.org/10.1007/s10551-015-2905-9

Westphal, J., & Park, S. H. (2020). Introduction to the Symbolic Management Perspective. In *Symbolic Management* (1st, pp. 1–18). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198792055.003.0001

Yahya, H. (2023). Female Leadership and ESG Performance of Firms: Nordic Evidence. *Corporate Governance*, *25*(1), 109–127. https://doi.org/10.1108/CG-03-2023-0129

Ye, D., Deng, J., Liu, Y., Szewczyk, S. H., & Chen, X. (2019). Does Board Gender Diversity Increase Dividend Payouts? Analysis of Global Evidence. *Journal of Corporate Finance*, *58*, 1–26. https://doi.org/10.1016/j.jcorpfin.2019.04.002

Yu, S. (2025). Trump Administration Pushback Hits US Gender Diversity Efforts. *Financial Times*. Retrieved July 14, 2025, from https://www.ft.com/content/c234079a-f5cb-455e-a3dd-bc44de7235ee

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Toward Reliable Emission Reporting: A Comparative Study of GHGRP Self-Reports and Climate TRACE Satellite Data

Felix Maximilian Kania

Technische Universität München

Abstract

Accurate and transparent greenhouse gas (GHG) emissions data is crucial for effective climate mitigation, yet existing reporting systems remain inconsistent and difficult to verify. Carbon accounting has emerged to give structure and legitimacy to these measurement efforts by mandating rules for affected GHG emitters through programs such as the Greenhouse Gas Reporting Program (GHGRP) by the United States government. Historically, these programs have relied on self-reporting, significantly limiting the verifiability of corporate-reported data. In contrast, emerging non-profit organizations such as Climate TRACE (CT) estimate facility-level GHG emissions using satellite-based remote sensing. This study quantifies facility-level discrepancies between these datasets and identifies their key drivers.

To do so, I systematically matched and compared facilities from both datasets based on reported emissions. I identified key predictors of these discrepancies using machine learning and Bayesian inference. My findings reveal that facility-specific effects drive most observed disparities, while parent-company and geographic influences play a secondary role. Industry-wide effects contribute minimally, with reporting year and total emissions volume having no significant impact. These results suggest that discrepancies stem from isolated inaccuracies rather than systemic errors, underscoring the need for hybrid verification frameworks that integrate self-reported (bottom-up) inventories with independent satellite-based monitoring (top-down) to enhance emissions transparency and accountability.

Keywords: greenhouse gas emissions; carbon accounting; emissions monitoring; satellite remote sensing; emissions verification

1 Introduction

Despite international climate agreements, global GHG emissions continue to rise, reaching a record high of over 50 billion metric tons of carbon dioxide equivalent (CO2e) in 2023, intensifying the risk of exceeding the 1.5°C warming threshold set by the Paris Agreement (UNFCCC, 2015). This escalation accelerates climate disruptions, including extreme weather, biodiversity loss, and economic instability (European Commission, 2024; IEA, 2024; IPCC, 2023). Given the urgency of addressing climate change (Fawzy et al., 2020), effective climate action depends on GHG reductions and the accuracy and transparency of emissions data, as reliable measurements are essential for tracking progress, enforcing regulations, and designing effective mitigation strategies. The growing emphasis on measuring and reporting emissions has raised critical concerns about different reporting systems' accuracy, consistency, and comparability (Huertas et al., 2012). However, discrepancies in carbon accounting - whether from self-reported corporate disclosures, regulatory inventories, or international emissions databases - pose a significant challenge to climate governance (Schaltegger & Csutora, 2012). Without reliable reporting, policymakers cannot design effective mitigation strategies; companies may misrepresent their carbon footprints, and global climate agreements risk being undermined by flawed emissions accounting (Klaaßen & Stoll, 2021). Carbon accounting methods have been developed to measure and verify emissions to address these challenges. While self-reported emissions inventories have been the dominant approach, they are susceptible to methodological inconsistencies and underreporting (Stechemesser & Guenther, 2012). In contrast, emerging satellite-based emissions tracking technologies offer independent, high-resolution verification, potentially transforming emissions accountability and climate policy enforcement (Kiemle et al., 2017).

While previous studies have compared different carbon accounting methods (Kennelly et al., 2019) including analyses of major public emissions databases (Moran & Wood, 2014) and satellite-based estimation techniques (Goetz et al., 2009), few have systematically examined discrepancies between facility-level self-reported emissions and independent satellite-based estimates. Research on Climate TRACE (CT) has primarily focused on sectoral trends or large-scale emissions monitoring, leaving questions about its reliability at the facility level (Gurney et al., 2024). This gap is particularly significant given the growing role of independent verification

DOI: https://doi.org/10.5282/jums/v11i3pp561-597

© The Author(s) 2026. Published by Junior Management Science.
This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.

in climate governance and emissions trading schemes.

This study explores the convergence of two opposing approaches in carbon accounting: traditional self-reported inventories, which rely on standardized methodologies and industry disclosures, and emerging satellite-based monitoring, which offers real-time, observational data. While regulatory programs like the U.S. Environmental Protection Agencies' (EPA) GHGRP (Environmental Protection Agency, 2024e) are based on legally mandated self-reporting, independent satellite-based systems such as CT (Climate TRACE, 2024b) provide an alternative, potentially more objective means of emissions verification (J. He & Huang, 2023). However, discrepancies between these two approaches raise questions about their accuracy, consistency, and potential biases in reporting methods. Understanding these discrepancies is essential for improving emissions transparency and strengthening hybrid verification systems that integrate both approaches.

The primary objective of my study is to investigate discrepancies between self-reported emissions from GHGRP and independent satellite-based estimates from CT. Specifically, I aim to identify the key predictors of these discrepancies at the facility level, assess the extent and direction of reporting biases, and evaluate the broader implications for emissions accounting and policy enforcement. By systematically comparing these two data sources, I contribute to the growing field of hybrid carbon accounting methods that integrate self-reported and observational approaches.

To achieve these objectives, my study follows a central research question:

> What are the key predictors of discrepancies in facility-level emissions data between the self-reported GHGRP and independent satellite-based Climate TRACE system?

To further refine this inquiry, my study examines the following sub-questions:

1. To what extent do facility characteristics, such as size, ownership, and reported emissions volume, influence discrepancies?
2. How do sector classification (subparts) and county-level effects contribute to discrepancies in emissions reporting?
3. How consistent are these findings across different analytical approaches, including variations in matching thresholds and statistical models?

By addressing these questions, I systematically evaluate the drivers of emissions discrepancies, offering a deeper understanding of the accuracy and limitations of self-reported and satellite-based emissions inventories.

With this, I contribute across three key domains. First, my study provides new empirical insights into facility-level variations in emissions reporting, contributing to the growing body of research on hybrid verification systems. By systematically identifying predictors of discrepancies, it advances the understanding of how different carbon accounting methodologies compare in practice. Second, by uncovering the specific conditions under which self-reported and satellite-based emissions data diverge, this research offers evidence-based recommendations for improving regulatory frameworks. The findings emphasize the importance of independent verification mechanisms in compliance programs such as GHGRP and inform policymakers on strategies to enhance emissions transparency and accuracy. Lastly, my study introduces a structured, replicable approach for matching and comparing self-reported emissions data with independent observational sources. By refining methods for facility-level emissions verification, the research informs future applications of satellite-based monitoring in carbon accounting, supporting the development of more reliable hybrid verification systems.

Through the integration of self-reported and satellite-based approaches, my study underscores the potential for hybrid verification systems to strengthen global emissions reporting, ultimately supporting more effective climate policy and accountability mechanisms.

2 Literature review

2.1 Carbon Accounting: An overview

Adopting the Kyoto Protocol in 1997 marked the formal beginning of internationally coordinated efforts to measure, report, and regulate GHG emissions, laying the groundwork for modern carbon accounting practices by introducing legally binding reduction commitments and market-based mechanisms (United Nations, 1998). Following the definition by Tang (2017), carbon accounting is a "system that uses accounting methods to record and analyze climate-change information as well as account for and report carbon-related assets, liabilities, expenses, and income for decision-making purposes" (p. 2). In recent years, more and more studies have been conducted in the field of carbon accounting because it is an essential tool in fighting climate change (Schaltegger & Csutora, 2012), which demands holistic approaches for its management and the mitigation of negative impacts (Schaltegger & Csutora, 2012). This trend is also reflected in the increasing legal codification of carbon accounting, as more laws and regulations have been introduced to standardize and enforce reporting practices (Chapple et al., 2013). Thus far, a handful of nations have led the efforts to reduce emissions, which have been particularly successful in Europe, which grew its GDP and simultaneously reduced emissions between 1995 and 2019 (Schaltegger & Csutora, 2012). Upon further inspection, however, many of these successes can be attributed to shifting emission-heavy production to other countries, particularly those in Asia (Schaltegger & Csutora, 2012), underscoring the need for internationally coordinated efforts in carbon accounting to eventually prevent climate change. Historically, carbon accounting has evolved in four distinct

stages: initially considered a subset of environmental management accounting, it later emerged as a separate field with growing attention from policymakers and businesses. Over time, its scope expanded to include supply chain and product-level footprints, ultimately progressing toward broader climate change accounting incorporating adaptation strategies and indirect impacts (Csutora & Harangozo, 2017).

In its basic form, carbon accounting "provides a mechanism through which carbon emissions can be measured and quantified and helps in knowing the emissions status of companies and making the necessary strategic decisions to achieve mitigation" (Hazaea, et al., 2023, p. 1). Thus, it is a decision-making tool at various institutional levels, including governments, corporations, and regulatory bodies. It enables informed policy development, business strategies, and sustainability initiatives (Schaltegger & Csutora, 2012). This is achieved through two basic approaches: documenting emissions and guiding mitigation efforts, underscoring its importance in measuring emissions and shaping proactive environmental strategies (Schaltegger & Csutora, 2012). Furthermore, carbon accounting is essential for integrating emissions data into policy frameworks and corporate decision-making processes. Establishing a quantitative basis for emissions allows organizations to assess their environmental impact and set reduction targets, offering transparency and accountability in sustainability reporting (Schaltegger & Csutora, 2012). Carbon accounting can be an emancipatory tool beyond its technical role in emissions tracking, challenging conventional accounting paradigms and integrating broader social and environmental concerns into corporate and policy decision-making (Ascui, 2014). The practice is central to climate governance and corporate responsibility by linking emissions data with regulatory compliance, financial performance, and strategic planning. Its ability to inform decision-making ensures that mitigation efforts are data-driven, allowing stakeholders to develop more effective policies and sustainable business models.

Carbon accounting follows the framework established by the GHG Protocol, the most widely used international standard (World Resources Institute, 2024). Emissions are categorized into three scopes: Scope 1 includes direct emissions from owned or controlled sources, such as company vehicles and boilers; Scope 2 covers indirect emissions from purchased electricity and other energy sources; and Scope 3 accounts for all other indirect emissions, including those from supply chains, transportation, and product usage. While Scopes 1 and 2 are relatively straightforward to measure, Scope 3 presents a more significant challenge due to its complexity and the need to include emissions along the entire value chain (Csutora & Harangozo, 2017). Beyond this, carbon accounting can be categorized into two significant methodologies: production-based accounting (PB), which attributes emissions to the geographic location of production, and consumption-based accounting (CB), which allocates emissions based on the end-user of goods and services (Marlowe & Clarke, 2022). The choice between PB and CB methodologies significantly impacts reported emissions, as studies have shown that CB often results in higher emission values due to the inclusion of embedded emissions in imported goods and services, which are otherwise excluded in PB calculations (Marlowe & Clarke, 2022). While traditional carbon accounting relies heavily on field sampling and land cover classifications, direct remote sensing methods offer a more spatially consistent way to monitor carbon stocks (Goetz et al., 2009). Recently, hybrid systems combining top-down (atmospheric measurements) and bottom-up (inventory reporting) emission measurements have been shown to significantly reduce the previously observed discrepancies between purely top-down and purely bottom-up (such as GHGRP) reporting systems (Chan et al., 2024).

In recent years, the transition from voluntary to mandatory carbon reporting has gained momentum, driven by the introduction of international reporting standards such as IFRS S1 and S2, which require companies to integrate carbon disclosures alongside traditional financial statements (Tang, 2024). While this shift enhances transparency and accountability, it presents significant challenges, including aligning carbon reporting boundaries with financial reporting structures and using standardized methodologies to ensure comparability across jurisdictions (Tang, 2024). While mandatory reporting aims to enhance transparency, companies may still have incentives to exaggerate carbon reduction claims, making regulatory oversight essential to mitigate greenwashing risks(Tang, 2024). Regulatory enforcement enhances carbon management effectiveness, which may, in turn, improve investor confidence and long-term financial stability (Sitompul et al., 2023). Several international initiatives, including the International Sustainability Standards Board's IFRS S1 and S2 standards and the EU's Corporate Sustainability Reporting Directive (CSRD), are refining disclosure requirements to enhance the quality, comparability, and consistency of reported emissions data. These regulations aim to bridge the gap between voluntary and mandatory disclosures by ensuring companies provide investors and regulators with standardized, audited, and comparable carbon information (Alfieri et al., 2024). The effectiveness of carbon reduction efforts varies by industry and regulatory environment. Research suggests that organizational capital plays a more decisive role in reducing emissions in carbon-intensive sectors. In jurisdictions with an Emissions Trading System (ETS), companies face more substantial financial and regulatory incentives to cut emissions (Papadopoulos, 2022).

Implementing carbon accounting extends beyond helping to reduce global carbon emissions; it can also enhance firms' financial performance by enhancing competition (Hazaea et al., 2023). A systematic review of more than 200 articles found that 59% of studies confirm a positive relationship between carbon management strategies and corporate financial performance, with 50% demonstrating a significant impact, reinforcing the financial incentives for sustainability efforts (Sitompul et al., 2023). However, some stud-

ies indicate that while carbon management enhances short-term financial performance through efficiency gains, long-term financial benefits depend on a firm's ability to offset initial investment costs and navigate regulatory complexities. The financial benefits of carbon management primarily arise from cost efficiency, enhanced productivity, product differentiation, and improved market positioning, highlighting the strategic advantages of emissions reductions beyond compliance (Sitompul et al., 2023). Nevertheless, the financial impact of carbon performance is not always linear. Evidence suggests a U-shaped relationship, where firms with superior or poor carbon performance benefit financially, while those in the middle face competitive disadvantages (Lewandowski, 2017). Empirical evidence from the EU-ETS suggests that carbon management does not continuously improve financial performance. Companies with environmental certifications face profitability declines when emissions rise, likely due to increased compliance costs or external scrutiny (Tuesta et al., 2020). A recent analysis of over 3800 firm-year observations of US publicly listed companies showed that businesses with higher organizational capital (specialized knowledge, skills, innovative technologies, and efficient logistics) tend to have lower GHG emissions (Provaty et al., 2024).

Additionally, while carbon emission mitigation can positively correlate with profitability, it can simultaneously lead to adverse investor reactions, as reflected in lower stock market valuations, potentially discouraging companies from ambitious emission reduction efforts (Lewandowski, 2017). Another study found that good carbon performance can reduce a firm's profitability and stock market valuation. Specifically, a 1% increase in carbon performance is associated with an average 0.055% decrease in return on assets (ROA), indicating that firms face financial trade-offs when improving their emissions profile (Busch et al., 2022).

Despite its growing importance, carbon accounting faces several limitations that impact its effectiveness and credibility. One key challenge is the lack of standardized methodologies, which creates inconsistencies in reporting and interpreting emissions data across different organizations and regulatory frameworks (Rogelj et al., 2019). One of the persistent challenges in carbon accounting is the significant measurement uncertainty associated with emissions estimates, which arises due to variations in data quality, reporting methodologies, and assumptions underlying emission factors (Marlowe & Clarke, 2022). This variation complicates efforts to compare emissions performance and undermines the reliability of carbon accounting as a decision-making tool (Hazaea et al., 2023). Additionally, the quality of carbon data remains a significant concern. An analysis of corporate disclosures revealed that emission data is often derived from a mix of reported figures and estimates. Only about 15% of data stems directly from company-reported sources, while the rest is derived from indirect indicators such as energy use and sustainability reports (Busch et al., 2022). A further significant limitation is the lack of comparability between reported emissions data across different firms and jurisdictions, making it difficult to assess performance, track progress, or enforce accountability (Marlowe & Clarke, 2022). The lack of harmonized global statistics on carbon emissions leads to data gaps and inconsistencies, as companies often rely on rough estimates. This affects firm-level and macro-level emissions reporting, complicating comparability and standardization across industries and jurisdictions (Alfieri et al., 2024).

Next, variations in how carbon performance is defined contribute to inconsistencies in financial impact assessments. Lewandowski (2017) highlights that some firms report carbon emissions primarily to establish a baseline comparison with industry peers, while others emphasize improvements in carbon performance over time to demonstrate progress. This discrepancy in disclosure strategies complicates the interpretation of carbon data, making it difficult to assess the impact of emission reductions on financial performance.

Determining reporting boundaries for carbon accounting presents another significant challenge, especially compared to traditional financial reporting frameworks (Rogelj et al., 2019). Control methods prescribed by accounting standards generally govern financial statements. In contrast, carbon accounting often allows for alternative boundary definitions, such as the equity method, which is not widely recognized in financial reporting (Tang, 2024). As carbon accounting becomes increasingly integrated into financial disclosures, aligning its reporting boundaries with financial statements is crucial to prevent inconsistencies and ensure clarity for investors and stakeholders (Tang, 2024). However, empirical evidence suggests that participation in emissions trading schemes like the EU-ETS does not necessarily translate into improved carbon performance. Firms within the EU-ETS do not exhibit better emissions management than non-participating firms, indicating that compliance alone does not drive proactive carbon reduction efforts. While organizational capital enhances a firm's ability to implement low-carbon strategies, financial constraints can undermine these efforts. High R&D and technology transition costs can prevent highly organized firms from fully realizing their emissions reduction potential (Provaty et al., 2024). As Schaltegger and Csutora (2012) emphasize, carbon "accounting must support managers to take the right decisions" (p. 1). However, if methodologies differ significantly, decision-makers may struggle to derive meaningful insights from data, potentially weakening emissions reduction efforts. Another concern is that carbon disclosure is often motivated by a desire to enhance corporate legitimacy rather than a genuine commitment to sustainability. Some companies use carbon accounting primarily as a public relations tool to maintain their image rather than to drive meaningful emissions reductions (Schaltegger & Csutora, 2012). This suggests that carbon reporting may sometimes be used defensively to manage external scrutiny rather than as a proactive emission reduction mechanism. Moreover, Scope 3 emissions measurement remains a particularly contested area, even in jurisdictions with mandatory carbon reporting (Tang, 2024). The complexity of supply chain emissions and data availability challenges create further barriers to reliable carbon accounting (Tang, 2024).

Furthermore, investor skepticism about the quality and reliability of carbon disclosures persists due to limited verification mechanisms. While investors increasingly push for transparency in emissions reporting, there is insufficient emphasis on validating the accuracy of reported data. This raises concerns about whether carbon accounting influences corporate decision-making and sustainability strategies or merely functions as a compliance exercise (Hazaea et al., 2023). Companies may selectively disclose favorable carbon data without stringent verification mechanisms while obscuring negative performance, underscoring the growing need for independent carbon assurance frameworks like ISAE 3410 (Tang, 2024). A growing concern in the field is the lack of expertise in carbon accounting, which has resulted in inconsistencies and misinterpretations of data. Many organizations lack trained professionals to assess and manage climate-related risks, leading to emissions calculations and reporting discrepancies (Tang, 2024). To address these issues, researchers emphasize the urgent need for robust policy frameworks and integrated accounting procedures that can enhance data reliability and improve the utility of carbon accounting for decision-making (Marlowe & Clarke, 2022). So far, standardization efforts have proven challenging, as carbon accounting inherently serves multiple, often conflicting purposes, including scientific measurement, political negotiation, financial reporting, and corporate disclosure, resulting in competing methodologies and frameworks (Ascui & Lovell, 2011).

In addition, the treatment of carbon credits and GHG offsetting remains ambiguous, further complicating carbon accounting. While carbon credits are expected to be disclosed in emissions reports, there is no consensus on whether they should be classified as assets, expenses, or financial instruments. This lack of regulation creates challenges in ensuring consistent and transparent reporting practices (Tang, 2024). These limitations underscore the need for stronger regulatory frameworks, standardized reporting methodologies, and independent verification mechanisms to enhance the credibility and impact of carbon accounting.

As carbon accounting continues to evolve, hybrid approaches integrating physical and monetary data are expected to play an increasingly important role. With the expansion of carbon accounting beyond direct emissions to include supply chains and product life cycles, researchers are developing new methodologies to improve accuracy and comparability (Schaltegger & Csutora, 2012). Multi-sensor approaches integrating satellite imaging, LIDAR, and radar measurements are expected to play a more significant role in verifying emissions data. Unlike traditional carbon accounting, which relies on financial reporting structures, hybrid verification systems combining physical and digital carbon tracking could enhance compliance verification, emissions trading, and corporate sustainability assessments (Goetz et al., 2009). Moreover, there is increasing pressure for international standardization in carbon accounting frameworks. Currently, emissions reporting varies significantly across jurisdictions, making it challenging to compare emissions between companies, sectors, and countries (Schaltegger & Csutora, 2012). Technological advancements, regulatory requirements, and growing investor demand for transparency will likely improve future carbon accounting.

Empirical evidence suggests that regulatory frameworks can mitigate the financial penalties associated with carbon reductions, reinforcing the need for stronger carbon pricing mechanisms and mandatory reporting requirements to align emissions reductions with corporate financial incentives (Busch et al., 2022). Innovations such as real-time emissions tracking, AI-driven data validation, and satellite-based verification systems are expected to enhance the accuracy and credibility of reported emissions. Blockchain technology can radically transform emissions data's transparency, reliability, and traceability in supply chains (Kaur et al., 2023). Emerging frameworks, such as the Dynamic Integrated Carbon Accounting (DICA) model, aim to address the complexity of modern carbon accounting by integrating evolving environmental, financial, and regulatory factors (Tang, 2024). Unlike traditional approaches, DICA incorporates forward-looking elements, financial carbon accounting, and management tools to enhance decision-making and climate risk assessment.

Finally, independent verification mechanisms, such as ISAE 3410, are expected to improve data credibility (Tang, 2024). To improve emissions tracking and minimize errors from multiple-counting, the E-liability approach has been proposed as a novel methodology that applies financial accounting principles, such as double-entry bookkeeping, to carbon tracking (Alfieri et al., 2024). This method allows firms to record direct emissions, account for emissions transferred from suppliers, and allocate emissions to products more accurately, reducing reliance on rough estimates and assumptions. By integrating real-time, auditable carbon data at each value chain stage, E-liability offers a potential solution for enhancing carbon reporting transparency and ensuring a more precise measurement of cumulative emissions (Alfieri et al., 2024). As carbon accounting moves beyond compliance, integrated and adaptive frameworks will support businesses' transition to net-zero operations.

2.2 GHGRP: The EPA's Greenhouse Gas Reporting Program

The US GHGRP originated from growing concerns over the effects of GHG emissions on climate and the environment within various US agencies in the 1990s, including the US Environmental Protection Agency (EPA). In response, the United States government set goals to limit their emission of greenhouse gases, thus establishing the need for accurate and transparent GHG accounting (Schrock et al., 1994). Initial voluntary reporting frameworks provided the basis for establishing the mandatory GHGRP in October 2009 (codified at 40 CFR Part 98 (Environmental Protection Agency, 2009)). The program was further influenced by previous successful cap-and-trade programs in the US, like the SO2 trading program (Kruger, 2005).

GHGRP aims to enhance transparency and facilitate data-driven US climate policies through its detailed GHG emissions accounting (Miller et al., 2024). Additionally, its "data can be used by businesses and others to track and compare facilities' greenhouse gas emissions, identify opportunities to cut pollution, minimize wasted energy, and save money" (Environmental Protection Agency, 2024, p. 1).

To do so, GHGRP "requires reporting of greenhouse gas (GHG) data and other relevant information from large GHG emission sources, fuel and industrial gas suppliers, and CO2 injection sites in the United States" (Environmental Protection Agency, 2024, p. 1). It includes CO2 and non-CO2 gases, mainly methane, nitrous oxide, and fluorinated gases, with tailored methodologies to capture sector-specific emissions (Hanle et al., 2012). "Direct-emitting facilities report emissions from each source category (rule subpart) included in the GHGRP, and these emissions can generally be categorized as either combustion or process emissions. Combustion emissions comprise CO2, CH4, and N2O emitted from fossil fuel combustion or biomass feedstocks. Process emissions generally include emissions from chemical transformations of raw materials and fugitive emissions" (Environmental Protection Agency, 2017, p. 1). Significant sources are defined as emitting at least 25,000 tons of CO2e annually, while smaller sources, sinks of greenhouse gases, and specific sectors like agriculture are excluded (Environmental Protection Agency, 2024b). It encompasses 46 industrial categories and recently expanded to include Direct Air Capture facilities, thus allowing for comprehensive tracking of CO2 flows from capture to final sequestration (Miller et al., 2024). Both direct GHG emitters and upstream suppliers are required to report. Direct emissions alone account for around half of US GHG emissions and fall under Scope 1. Here, reporting is done on the facility level, with parent company information also being collected (Environmental Protection Agency, 2024b). "Suppliers report the amount of CO2e that would be released if the products they produce, import, or export (e.g., fossil fuels) were released, combusted, or oxidized. These are reported at the corporate level and fall under Scope 3 emissions" (Environmental Protection Agency, 2024, p. 1). Together, these cover 85-90% of US GHG emissions (Environmental Protection Agency, 2024b). As the program is federal, the location within the United States does not directly affect the reporting rules (U.S. National Archives and Records Administration, 2025). Emissions from fuel combustions, e.g., burning biomass for energy or heat production, are measured through a continuous emission monitoring system (CEMS), which is considered the most accurate approach, or by applying default emissions factors. An emissions factor is an average value that estimates pollutant emissions per unit of activity, typically expressed as the weight of emissions per unit of fuel burned or industrial output (Environmental Protection Agency, 2024a). On the other hand, process emissions (emissions released from chemical reactions or transformations of raw materials) can be accounted for by either CEMS, a mass balance approach, or site-specific or default emission factors (Environmental Protection Agency, 2017).

Firms are required to submit annual reports electronically by 31 March of the following year. Reporters are provided with a series of online training sessions to provide them with the necessary information to comply with the program's reporting requirements, depending on the industry sector of their facilities. The program's online reporting tool employs built-in data validation checks to ensure accuracy (Environmental Protection Agency, 2024e). GHGRP improves upon earlier programs, such as the Inventory of U.S. Greenhouse Gas Emissions and Sinks (Environmental Protection Agency, 2025a), by requiring direct measurements for specific sectors like methane recovery, reducing uncertainties and enhancing robustness in landfill emissions tracking (Bronstein et al., 2012). Generally, the program follows a flexible approach, balancing accurate reporting on the one hand with the need for cost-effectiveness and not overburdening firms with unnecessarily complex bureaucracy. For example, facilities can choose between direct measurement methods or applying EPA-approved emission factors to site-specific activity data in most industries (Hanle et al., 2012).

A previous study on the program showed that electric power plants affected by its reporting regulations consequently reduced CO2e emissions by 7% compared to non-reporting facilities (Yang et al., 2024). This reduction effect is even more prominent for companies with prior experience in pollution prevention, demonstrating how GHGRP builds upon pre-existing environmental knowledge systems (Lee & Bi, 2020). Furthermore, it leads to firms significantly improving their emission intensity, defined as Scope 1 specific emissions divided by total assets, even without substantially reducing absolute emissions. Through its mandatory and standardized reporting, the program also reduces impression management opportunities, meaning the ability of organizations to influence how they are perceived by others, thus increasing the credibility of reported data (Bauckloh et al., 2022). Recently, the program's introduction has also been linked to companies' significant reduction in greenwashing behavior by increasing their motivation to improve ESG performance and enhancing transparency and accountability, especially for large firms worried about public perception (Nguyen et al., 2025). GHGRP has further been shown to shape corporate behavior as firms mandated to report under its rules are 68% less likely to face environmental lawsuits, driven by increased community stakeholder pressure from enhanced reporting transparency (Huang et al., 2023). In general, the standardized and easily accessible data provided by the EPA allows for large-scale analyses of greenhouse gases, which offer significant opportunities for diverse scientific inquiries (Jain et al., 2021).

Several limitations of GHGRP have been evaluated so far. As mentioned, the program might not necessarily reduce absolute CO2e emissions (Bauckloh et al., 2022), as it is not directly linked to reduction mechanisms such as carbon pricing. Therefore, the program may incentivize operational efficiency without necessarily aligning with the

broader climate goal of reducing total emissions. GHGRP also demonstrates methodological limitations by relying on outdated emission factors and thresholds, excluding significant emission sources (Subramanian et al., 2015). Here, the authors stress the need for direct measurements of gases to guarantee accurate reporting, challenging indirect measurement systems like those satellite-based systems introduced in the following chapter. The program's cost-effectiveness has also been under scrutiny, arguing that its contribution to emission reductions is negligible for its high implementation costs (Sanchez et al., 2012). Looking at the facilities required to report, the associated incremental costs are as high as $1.42 per metric ton of CO2e, not considering potential tax credits that might exceed this amount (Godec et al., 2022). However, it should be noted that this study was released before the data used in this thesis started. Since then, the program has provided open access to its data, addressing one of the authors' main concerns. Due to data confidentiality concerns, facility-level data in specific sectors, such as suppliers of carbon dioxide, is withheld, limiting data consistency (Miller et al., 2024). Further criticism has focused on certain methodological flexibilities, which facilities are offered to compensate for their diverse operational and physical characteristics, which might lead to the systematic underestimation of the US methane inventory (Stark et al., 2024). More reporting inaccuracies have been attributed to GHGRP's often static and overly simplistic assumptions, as demonstrated through the discovery of the significant effect of rainfall on methane emissions from landfills, which is not considered by the program's emission factors (Jain et al., 2021). It should be noted that EPA continues to update its program, regularly adding new features to its tools and expanding its datasets (Environmental Protection Agency, 2024c).

2.3 Climate TRACE: Satellite-based emissions monitoring

Climate TRACE is a non-profit coalition of various organizations with the stated goal of making "meaningful climate action faster and easier by harnessing technology to track greenhouse gas (GHG) emissions with unprecedented detail and speed, delivering information that is relevant to all parties working to achieve net-zero global emissions" (Climate TRACE, 2025, p. 1). As of the writing of this paper, over 100 organizations, including Google.org, NASA, the US EPA, the European Space Agency, and various universities like Carnegie Mellon, are collaborating on the project. It was first started in May 2019 with funding from Google.org and the vision to use artificial intelligence to monitor power plant emissions from space (Climate TRACE, 2025a). Since then, the coalition has expanded to cover all GHG emissions and improve its technology, releasing its first emissions inventory in September 2021. In their second facility-level inventory, released in December 2023, CT covered over 352 million emissions sources worldwide (Climate TRACE, 2025a). This is also the inventory I used in the analysis of this paper. In November 2024, their third inventory was released, covering more than 660 million individual sources of GHG emissions (Climate TRACE, 2025a). The collected data is publicly available through their interactive website, API, and download packages.

Contrary to most other GHG accounting projects, such as the GHGRP discussed before, the non-profit started providing emissions inventories that were not based on self-reported data. Its primary purpose is to enable and facilitate "the creation of policies, programs, and campaigns aimed at limiting global temperature rise to 1.5°C as agreed to under the Paris Climate Agreement" (Climate TRACE, 2025, p. 1).

CT's data collection is based on over 300 satellites orbiting the Earth. The first step, thus, is identifying the locations of significant sources of emissions around the globe. These include power plants, factories, mines, airports, etc. Next, ground truth - direct on-site measurement - samples are collected for some of these, meaning precise emission measurements. These can then be used to train machine-learning algorithms together with satellite imagery. The resulting models can then be extrapolated to all sources of emissions that do not have ground truth samples (Climate TRACE, 2025a). This approach is not unique, with various other projects like Landsat 8 by NASA (NASA, 2025a) or PlanetScope by ESA (The European Space Agency, 2025) providing the satellite imagery necessary to estimate carbon emissions remotely (Couture et al., 2024).

The resulting facility-level data is categorized into 10 sectors (Figure 1), in decreasing order of total emissions: Power (the leading source with around 25% of global emissions), Manufacturing, Fossil Fuel Operations, Transportation, Agriculture, Buildings, Waste Fluorated Gases, Mineral Extraction, and Forestry And Land Use, as well as 67 sub-sectors. The project's database contains information on over 500 billion tons of CO2e between January 2015 and December 2024 (Climate TRACE, 2025b).

Source: (Climate TRACE, 2025a)

For power plants specifically, CT employs a dual-method approach to estimate facility-level emissions. The first method, applied to 12.5% of global power plants, utilizes machine learning (ML) and remotely sensed water vapor plumes to derive time-dependent capacity factors (CFs) at the facility level. The second method, which accounts for 87.5% of facilities, relies on national-average power plant properties to estimate emissions (Gurney et al., 2024). This distinction is crucial, as the ML-based approach offers higher precision by incorporating real-time satellite observations. In contrast, the national average method introduces more significant uncertainty due to its reliance on broad assumptions (Gurney et al., 2024).

Another advantage of satellite-based sensing is its potential to identify and locate previously unknown or underestimated emissions hotspots, such as methane-emitting solid waste landfills (Dogniaux et al., 2024). Generally, hybrid approaches combining activity-based and atmospheric measurements can "produce more complete and accurate estimates of GHG emissions and sinks" (National Academies of

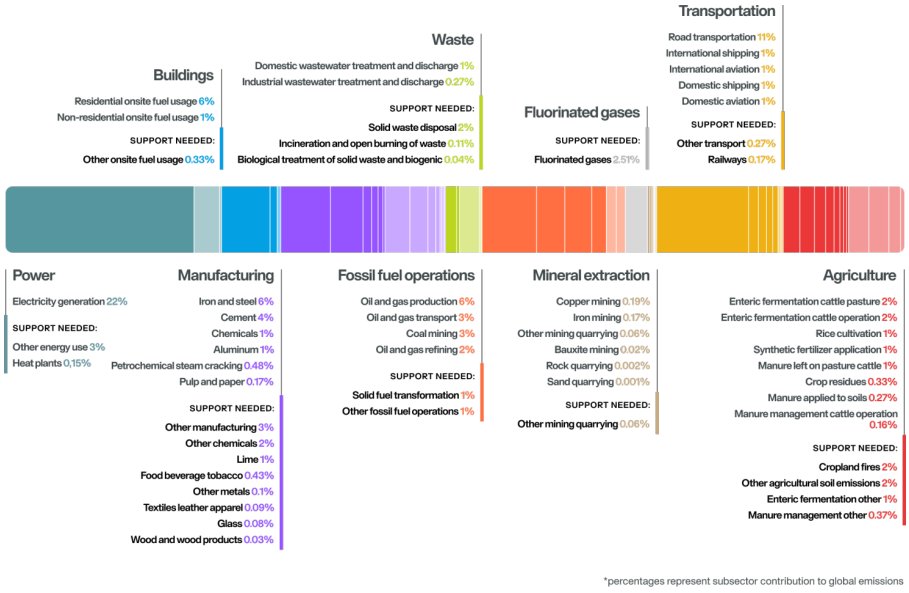


Figure 1: Climate Trace global emissions per category 2024

Sciences, Engineering, and Medicine, 2022) than individual approaches.

A 2024 study compared Climate TRACE's power plant emissions estimates with the Vulcan-power v1.1 dataset, revealing notable discrepancies. While CT's AI-based satellite detection produced relatively accurate estimates (mean relative difference = -1.1%, SD: 46.4%), its national-average capacity factor (CF) - the typical ratio of actual energy output to maximum possible output for a given energy source across a country, averaged over time - approach led to large underestimations (MRD = -50.0%, SD: 117.7%) (Gurney et al., 2024). The study highlights that CT's use of national-mean CFs does not account for plant-level variations, leading to significant biases. In contrast, emissions estimates derived directly from AI-based satellite observations showed far better agreement with Vulcan-power v1.1, suggesting that expanding direct satellite measurements rather than relying on generalized assumptions could improve CT's accuracy (Gurney et al., 2024). A – yet-to-be-peer-reviewed study - conducting a direct comparison between GHGSat-detected emissions (HGSat Inc., 2025) - another satellite-based GHG emissions monitor - and CT estimates for landfills found no correlation between the two datasets, raising concerns about the reliability of modeled emissions estimates (Dogniaux et al., 2024). This suggests that CT's landfill emissions estimates may diverge significantly from direct satellite observations, highlighting potential methodological limitations. While CT's approach integrates machine learning and global facility data, its reliance on modeled emissions factors rather than direct satellite plume detection could explain the observed discrepancies.

Furthermore, the study found substantial differences in magnitude between CT and GHGSat estimates, with CT overestimating or underestimating emissions (Dogniaux et al., 2024). Another point of criticism centers around the organization's transparency regarding its data and source code, which are mostly not publicly available. The National Academy of Sciences has categorized the website as Low on Usability and timeliness because there is little evidence that it is used by relevant decision-makers, Low on Information Transparency because of the mentioned limited access to data and source code, and Low on Evaluation and transparency resulting from few peer-reviewed evaluations of its data (National Academies of Sciences, Engineering, and Medicine, 2022).

These findings reinforce a key limitation of Climate TRACE: while its dataset provides unprecedented global coverage, discrepancies arise when compared to direct remote sensing measurements.

Looking ahead, the launch of ever more numerous and sophisticated satellites like the Copernicus CO2 Monitoring mission (Deutsches Zentrum für Luft- und Raumfahrt e.V., 2024) scheduled for 2026 is expected to further improve remote emissions monitoring (Couture et al., 2024), directly improving CT's data coverage and quality.

2.4 Existing research on emissions discrepancies

There have already been several other studies focused on understanding emissions discrepancies between different data sources. Recent - not yet peer-reviewed - research has shown that discrepancies in landfill methane emissions are particularly pronounced, with satellite observations revealing emissions up to six times higher than reported to the GHGRP (Balasus et al., 2025). This divergence is primarily driven by the systematic selection of a reporting methodology that yields lower emissions estimates, as landfills increasingly opt for the recovery-first model, which estimates emissions from recovered methane rather than total generation, often overestimating collection efficiency and underreporting actual emissions trends (Balasus et al., 2025). Discrepancies

have also been observed at national inventory levels, as a recent study comparing CT and Brazil's SEEG inventory found emissions estimates differing by up to 1 gigaton of CO2e in some years, primarily due to methodological differences in the forestry and land-use sectors (da Costa et al., 2025). Additionally, CT appears to better capture seasonal variability in emissions and removals, whereas traditional inventory methods rely on static emissions factors, potentially leading to underestimations of dynamic carbon sinks (da Costa et al., 2025). A further high-resolution satellite-based analysis confirms systematic underreporting of landfill methane emissions to GHGRP, with TROPOMI-based estimates showing landfill emissions are 77% larger than reported. Independent aircraft-based validation supports these findings, further reinforcing self-reported landfill methane data reliability concerns (Nesser et al., 2024). Another study compared CO2 emissions estimates from the ODIAC satellite-based dataset with city-reported inventories across 14 global cities, revealing significant differences. In some cities, such as Cape Town (+148%) and São Paulo (+43%), satellite-based estimates were much higher than reported data, likely due to industrial emissions being misallocated in developing countries (Chen et al., 2020). Further research on U.S. city-level emissions has found that self-reported inventories underestimate emissions by an average of 18.3%, with some cities misreporting emissions by up to 145.5%. If applied to all U.S. cities, this underreporting would exceed California's total annual emissions by 23.5%, highlighting the scale of uncertainty in urban carbon accounting (Gurney et al., 2021).

2.5 Comparing GHGRP and Climate TRACE

Accurate reporting of greenhouse gas emissions is critical for climate policy and corporate accountability. However, as discussed, discrepancies persist between self-reported emissions data, such as the U.S. Environmental Protection Agency's (EPA) Greenhouse Gas Reporting Program (GHGRP), and independent satellite-based estimates from sources such as Climate TRACE (CT). These discrepancies stem from methodological differences, operational limitations, and data verification challenges (Balasus et al., 2025; Chen et al., 2020; da Costa et al., 2025; Gurney et al., 2021; Nesser et al., 2024).

GHGRP relies on a bottom-up inventory-based approach in which facilities self-report emissions using standardized estimation techniques, such as fuel-based emissions factors, mass balance methods, or direct measurement via Continuous Emissions Monitoring Systems (CEMS) (Environmental Protection Agency, 2024b). While this framework allows for industry-specific reporting flexibility, it inherently depends on the accuracy of facility-reported data and the assumptions underlying emissions calculations (Hanle et al., 2012).

CT employs a "top-down" approach integrating satellite observations, remote sensing techniques, and artificial intelligence (AI) models to estimate facility-level emissions (Climate TRACE, 2025a). This methodology can independently verify carbon stock estimates and enhance emissions monitoring accuracy, potentially identifying discrepancies in self-reported inventories (Goetz et al., 2009). However, CT's reliance on model extrapolation for facilities without direct observations introduces uncertainty, particularly in cases where ground-truth verification is limited (Gurney et al., 2024).

A key distinction between the two methodologies is their ability to capture transient emissions events, meaning short-term, high-intensity releases of pollutants or greenhouse gases that occur intermittently rather than continuously. GHGRP, which requires facilities to submit annual emissions reports, provides estimates based on standardized methodologies rather than real-time measurement (Environmental Protection Agency, 2024b). As a result, short-term fluctuations due to operational changes, infrastructure failures, or meteorological conditions may not be fully reflected (Subramanian et al., 2015). In contrast, satellite-based monitoring can detect short-lived but significant emissions surges, such as methane "burps" from landfills or unintentional fugitive emissions from oil and gas facilities (Cusworth et al., 2021). This distinction is particularly relevant for methane accounting, as episodic emissions events can represent a significant portion of total emissions but remain undetected in self-reported inventories (Duren et al., 2019).

Beyond methodological differences, facility-specific factors are crucial in reporting discrepancies in emissions (Cusworth et al., 2021; Gurney et al., 2024; Nesser et al., 2024). GHGRP provides multiple methodologies for estimating emissions, including direct measurement, mass balance calculations (which track inputs and outputs to infer emissions), and default emissions factors, depending on the sector (Environmental Protection Agency, 2017). However, sectors relying on modeled emissions factors, such as landfills and oil and gas facilities, are prone to underreporting, as assumptions about operational efficiency and emissions capture rates do not always align with real-world conditions. This aligns with broader criticism that self-reported emissions inventories tend to underestimate non-combustion sources, particularly fugitive emissions (Subramanian et al., 2015).

In contrast, like other airborne spectrometer-based methods, CT's remote sensing approach can effectively detect fugitive methane emissions by directly observing plumes. While studies have primarily focused on oil and gas infrastructure, similar methods could be applied to landfills to improve emissions detection and characterization (Cusworth et al., 2021). This is supported by a growing body of research advocating for hybrid verification approaches that combine self-reported data with independent observational methods to improve emissions accuracy (Chan et al., 2024).

GHGRP and CT offer distinct advantages and limitations that shape their effectiveness for monitoring facility emissions. GHGRP provides legally mandated, structured reporting that aligns with federal regulations and compliance mechanisms. This ensures that all significant GHG emitters in the U.S. must adhere to a standardized methodology, improving industry comparability (Environmental Protection Agency, 2024b). The program also tailors emissions calculation methods to specific industries, allowing for

detailed sectoral tracking (Hanle et al., 2012). Another strength is its continuous dataset dating back to 2010, enabling longitudinal analysis of emissions trends at the facility level (Environmental Protection Agency, 2024b). On the other hand, emissions estimates are only as accurate as the reporting facility's methodology and compliance efforts, raising concerns about potential underreporting or misclassification (Subramanian et al., 2015). Furthermore, GHGRP excludes certain smaller facilities and unregulated emissions sources, potentially omitting significant emission contributions (Environmental Protection Agency, 2017).

CT offers an external validation mechanism for facility-reported emissions, reducing reliance on self-reported estimates (Climate TRACE, 2025b). Its satellite-based monitoring enables continuous emissions tracking, potentially detecting previously unknown sources and emissions spikes (Goetz et al., 2009). Finally, unlike GHGRP, which is limited to the U.S., CT provides global emissions data across multiple sectors, facilitating international climate monitoring efforts (Climate TRACE, 2025b; Environmental Protection Agency, 2017). While some facilities benefit from direct satellite observations, many CT estimates rely on machine-learning extrapolations, which introduce uncertainty in emission calculations (Gurney et al., 2024). CT's facility-level methane emission estimates show significant discrepancies from national reporting programs and independent satellite observations, with no strong correlation found. This suggests potential accuracy limitations, particularly in industrial sectors where site-level validation is critical (Dogniaux et al., 2024).

Given the strengths and limitations of both systems, a hybrid verification approach combining self-reported data (GHGRP) with independent observational data (Climate TRACE and other remote sensing systems) might offer the best path forward for improving emissions accuracy.

Despite their differences, GHGRP and CT provide facility-level emissions data that can be directly compared to enhance emissions verification efforts. GHGRP's structured reporting framework allows consistent facility-level tracking, while CT offers an independent validation mechanism using satellite-based estimates (Climate TRACE, 2024b; Environmental Protection Agency, 2024d). The availability of comparable facility-level emissions data from both sources enables cross-validation, offering a promising pathway for improving emissions monitoring and policy enforcement. The availability of comparable facility-level emissions data from both sources enables the comparative analysis conducted in this paper, providing a foundation for assessing discrepancies and improving emissions monitoring approaches.

2.6 Identified gaps in the literature

Despite growing efforts to improve emissions accounting, significant gaps remain in reconciling self-reported and independent emissions estimates (Dogniaux et al., 2024; Subramanian et al., 2015). Studies have highlighted systematic underreporting of methane and CO2 emissions across different sectors and geographic scales, yet the drivers of these discrepancies remain poorly understood (Balasus et al., 2025; Nesser et al., 2024; Subramanian et al., 2015). While previous research has compared self-reported inventories and independent estimates, these analyses often lack facility-level resolution, making it challenging to pinpoint systematic biases in emissions reporting (Dogniaux et al., 2024; Gurney et al., 2021). Furthermore, the effectiveness of hybrid verification approaches, which integrate bottom-up self-reported data with top-down satellite-based estimates, remains underexplored at the facility level (Chan et al., 2024; Goetz et al., 2009).

To address these gaps, I conducted a facility-level comparative analysis of this study's GHGRP self-reported dataset and CT satellite-based estimates. By systematically matching individual facilities across the two datasets and applying machine learning and Bayesian statistical methods, this research identifies key predictors of emissions discrepancies and assesses potential biases in facility-level reporting.

3 Methodology

3.1 Research design

In this study, I performed a comparative analysis to evaluate discrepancies between self-reported emissions data from the GHGRP and independent satellite-based estimates from CT. My primary objective was to identify key predictors of emissions discrepancies and assess systematic biases that may exist in facility-level GHG reporting. I systematically matched individual facilities across the two datasets to do so, directly comparing facility levels. With this approach, I provide new insights into patterns of over- or underestimation and the factors driving these discrepancies. For the complete analysis, see Appendix A.

To ensure a robust and reproducible analysis, I followed a structured four-step workflow:

1. Since GHGRP and CT do not share a common identifier, I developed a multi-step facility-matching algorithm based on geographical proximity, name similarity, and industry classification.

2. I measured emissions differences using absolute discrepancies in tons of CO2e, allowing me to perform a nuanced analysis of over- and under-reporting trends.

3. Given the skewed distribution of emissions discrepancies, I applied an appropriate statistical transformation to ensure reliable modeling.

4. To determine the predictors of discrepancies, I used two complementary approaches, namely Random Forest regression, to identify key predictors through machine learning and Bayesian Hierarchical Modelling to capture structured variability at multiple levels, such as facility, corporate ownership, and geographic region.

By integrating self-reported and independent satellite-based data with this study, I contribute to improving our understanding of reported emissions differences and furthering our understanding of the validity of satellite-based emissions accounting. The findings provide practical insights for policymakers, industries, and researchers, supporting efforts to enhance the accuracy and reliability of global emissions monitoring.

3.2 Data collection and matching process

3.2.1 GHGRP

I downloaded the GHGRP data from the EPA's Greenhouse Gas Reporting Program's website's Facility Level Information on GreenHouse Gases Tool (FLIGHT) (Environmental Protection Agency, 2024d) in CSV format, covering 2015-2022 and containing around 7,600 unique facilities. The resulting dataset held 61,358 rows of facility-level information on CO2e emissions, including an ID, the facility's name, latitude, longitude, county, parent company, industry, and emissions in tons. Next, I set out to preprocess the data, which included handling the 9.62% of rows missing a county. Addressing this was critical for later statistical modeling. To do so, I downloaded a dataset containing a lookup between US Zip Codes and counties from Kaggle (Ofer, 2018) and used this to fill in the missing counties, as all the GHGRP rows contained ZIP Code information. Following this approach, I reduced the percentage of missing counties down to 1.46% of total rows. Finally, I harmonized naming conventions for parent companies and counties using fuzzy matching (rapidfuzz, 2025), as many slight variations of the same names existed, which would have caused problems later.

3.2.2 Climate TRACE

The CT dataset I used in this analysis also covered 2015-2022, and I downloaded it in CSV format from the organization's website (Climate TRACE, 2024a). Like the first dataset, each row contained information on a facility's CO2e emissions for a single year, an ID, the facility's name, latitude, longitude, and industry sector. In total, the data frame contained 363,514 rows, covering the whole USA. Of these, 1.83% were missing emissions quantity information. As this information is critical for my analysis, I decided to omit the affected rows. Another challenge was that around 8,200 rows of data did not display yearly but monthly emissions data, so I needed to aggregate these to later compare them with the GHGRP dataset. Around 20,000 unique facilities were covered annually, significantly more than the GHGRP dataset because CT covers more industries than GHGRP and does not impose the 25,000 tons of CO2e threshold mentioned before.

3.2.3 Facility-level data matching

Now, both datasets were ready to be matched. The goal here was to find the CT equivalent for each row of GHGRP data so that the same facility information would be available for the same year. As no standard ID linked the two datasets, I needed to develop a matching algorithm based on multiple steps, including geographic information through the facility coordinates, similarity between facility names, and industry overlap. Another study has already validated a similar approach connecting the CT data to other facility-level emissions data sets (Gurney et al., 2024). This approach was designed to ensure accurate matches while minimizing false positives:

1. Coordinate-based matching: I used the facilities' latitude and longitude as key identifiers with a k-d tree algorithm (SciPy, 2024) to efficiently query facilities in the CT dataset within 5 meters of a given GHGRP facility. I chose this distance as a conservative threshold, as this would allow for variability while minimizing false positives.

2. Temporal filtering: Then, I restricted the resulting matches to the same reporting year to maintain temporal alignment between the datasets.

3. Category overlap filtering: Facilities were filtered further by comparing their broad emission categories. A match was only considered valid if the categories between the two datasets overlapped or the GHGRP facility was classified as "Other". I introduced this step as the industry sector in the two datasets was initially not aligned. The exact categorization can be seen in Table 1. This step reduced mismatches where geographically close facilities belonged to different industries, ensuring that only functionally comparable facilities were matched.

4. Name similarity scoring: Up to this point, multiple CT facilities could be valid matches for a GHGRP facility. To address this, I calculated name similarity scores using the fuzz.ratio function from the rapidfuzz library (rapidfuzz, 2025) for facilities that met the spatial and category overlap. This score, ranging from 0 to 100, quantified the textual similarity between facility names across datasets.

5. Combined scoring and ranking: I then integrated the spatial proximity and name similarity scores into a single combined score, calculated as:

$$Combined\ Score = 0.5 \times \left(1 - \frac{Distance}{5\,m}\right) + 0.5 \times \left(\frac{Name\ Similarity}{100}\right) \quad (1)$$

Where Distance is the distance in meters between possible matches. This formula equally weighted geographic proximity and textual similarity to prioritize accurate matches.

6. Selecting the best match: As a final step, for each GHGRP facility, I ranked all potential matches by

Table 1: Broad category mapping between GHGRP subparts and CT subsector

Broad category	GHGRP subparts	CT subsector
Manufacturing	Iron and Steel Production, Glass Production, Cement Production, Pulp and Paper Manufacturing, Lime Manufacturing, Ammonia Manufacturing, Petrochemical Production, Hydrogen Production, Aluminum Production, Electronics Manufacturing, Silicon Carbide Production, Soda Ash Manufacturing, Titanium Dioxide Production, Manufacture of Electric Transmission and Distribution Equipment, Lead Production, Phosphoric Acid Production, Nitric Acid Production, Ferroalloy Production, Zinc Production, Magnesium Production, Adipic Acid Production	other-manufacturing, aluminum, petrochemicals, steel, cement
Mineral Extraction	Underground Coal Mines	coal-mining, copper-mining, iron-mining
Power	Electricity Generation	electricity-generation
Transportation	Road Transportation, Domestic Aviation, International Aviation	road-transportation, international-aviation, domestic-aviation, domestic-shipping, international-shipping
Waste	Municipal Solid Waste Landfills, Industrial Waste Landfills, Industrial Wastewater Treatment	solid-waste-disposal, wastewater-treatment-and-discharge
Fossil fuel operations & Fluorinated Gases	Petroleum and Natural Gas Systems	oil-and-gas-production-and-transport, oil-and-gas-refining
Other	Geologic Sequestration of Carbon Dioxide, Use of Electric Transmission and Distribution Equipment, Importers and Exporters of Equipment Pre-Charged with Fluorinated GHGs or Containing Fluorinated GHGs in Closed-Cell Foams, HCFC-22 Production and HFC-23 Destruction, Injection of Carbon Dioxide, Fluorinated Gas Production, Petroleum Refineries, Suppliers of Natural Gas and Natural Gas Liquids, Suppliers of Petroleum Products, General Stationary Fuel Combustion Sources, Suppliers of Coal-based Liquid Fuels, Suppliers of Carbon Dioxide, Miscellaneous Uses of Carbonate, Suppliers of Industrial Greenhouse Gases	

Source: Own illustration modified from (Environmental Protection Agency, 2024f)

their combined score and selected the highest-ranking match as the most likely pair. This way, I could employ a transparent, scalable, empirically validated method for linking GHGRP and CT facilities while minimizing false matches.

The resulting data frame now contained 17,762 rows, which meant I found matches for 28.9% of all GHGRP rows. To further improve this number, I employed an additional fuzzy name-matching algorithm for the unmatched data:

1. Integrating ownership data: The GHGRP dataset held information for a facility's parent company. However, the original CT dataset did not have this information for all categories. Since ownership information can help resolve ambiguous matches, I merged ownership data from a separate CT dataset from the organization's website (Climate TRACE, 2024a).

2. Fuzzy name matching: The actual matching was done by applying fuzzy matching (rapidfuzz, 2025) using a facility's name and its parent company's name. This resulted in a text similarity score between 0 and 100 for all unmatched facilities of the same reporting year. I balanced the importance of facility name and parent company name equally, using the following formula:

$$Combined\ Score = \frac{Facility\ Name\ Similarity + Ownership\ Similarity}{2} \quad (2)$$

This way, I could account for facilities with similar names belonging to different companies. I considered two facilities a potential match if the score was above 70.

3. Extending potential matches with distance calculations: For the identified potential matches, I next added the distance between the facilities, using their Haversine distance – the distance between two points on Earth using their latitude and longitude - (PyPi, 2024) through the coordinates, as I did in the coordinate-based matching before. This allowed me to evaluate whether fuzzy-matched facilities were also geographically reasonable.

4. Manually selecting valid matches: The previous steps identified 38 potential matches. After manually checking each of these, considering their similarity score and distances, I used the GHGRP FLIGHT tool (Environmental Protection Agency, 2024d) and the interactive CT emissions map (Climate TRACE, 2024b) to inspect them. Doing so resulted in only 10 more valid matches, combined with the 17,762 rows from before.

3.2.4 Validation checks on matched data

I employed several validation checks on the final matched data to understand how well my matching process worked. An overview of the relevant metrics can be seen in Table 2 below.

I suspect the reason for the relatively low percentage of original GHGRP rows that found a match (28.96%) in the CT dataset is the different scopes of the two data sources ex-

Table 2: Overview of evaluation metrics of matched data

Number of matched rows	Fraction of rows matched	Fraction of unique facilities matched	Mean distance between facilities	Median distance between facilities	Mean facility name similarity	Median facility name similarity
17,772	28.96%	36.40%	0.87m	0.00m	92.39	100.00

Table 3: Percentage of matches per year

Year	Number of GHGRP rows	Fraction of rows matched (%)	Fraction of unique facilities matched (%)
2015	7,983	2.43	14.90
2016	7,643	2.41	14.79
2017	7,569	2.39	14.69
2018	7,676	2.42	14.92
2019	7,683	4.47	27.80
2020	7,643	4.38	27.27
2021	7,615	5.24	32.66
2022	7,546	5.22	32.53

Source: Own illustration

plained before and the conservative choice of using a maximum allowed distance between potential facilities of only 5 meters. I will address this tradeoff in a later sensitivity analysis exploring different distance thresholds. My primary goal for the principal analysis was to minimize the prevalence of false matches. I concluded that I succeeded by looking at the low distances between matches and the high similarity between matched facility names.

Further investigation of matches per year, shown in Table 3, reveals that while the number of GHGRP facilities per year is stable, more recent years resulted in a higher percentage of matches. To ensure this increase does not indicate flaws in my matching process, I investigated further to identify potential explanations for the trend.

I came up with several hypotheses for this increase:

1. Improved data coverage by CT over time: The organization improved its technology and coverage in more recent years, supported by its first emissions inventory released in 2021 (Climate TRACE, 2025a). For later years, this could mean more complete reporting, including more facilities covered, leading to better facility matching.

2. Improved industry representation in CT: If specific industries were underrepresented in earlier CT datasets but later improved, this would explain the increasing number of matches.

3. Changes in matching accuracy: If name similarity or location accuracy changes over time, this could lead to higher rates later.

Testing this hypothesis revealed several drivers for the increase in matches in later years:

1. The Facility coverage over time graph (Figure 2) shows that CT increased its number of tracked facilities significantly from 2019 onwards. Thus, it is only logical that the percentage of matches increases as the number of GHGRP facilities remains stable.

2. The Industry match rates over time graph (Figure 3) further supports this by showing shifts in the distribution of matched industries, suggesting better industry representation over time. This means that from 2019 onwards, more industry sectors were present in the CT dataset, leading to increased matches.

3. The increased distance between matched facilities (Figure 4) and the decrease in name similarity suggests a slight decline in match quality, likely leading to more matches as the criteria for determining a match became less strict.

In conclusion, there is data to support all three of my hypotheses for the increase in matches in later years. The primary reason for the increase is the expansion of the CT dataset after 2018. However, this is not evidence of a fundamental issue or a significant reduction in the quality of the matched data.

3.3 Calculation of GHG differences

My next goal was to quantify the discrepancies in reported emissions between matched facilities. Both datasets report metric tons of CO2e emissions using a Global Warming Potential (GWP) of 100 years(Climate TRACE, 2024a; Environmental Protection Agency, 2025b), making unit conversion unnecessary. I calculated absolute discrepancies in CO2e emissions using the following formula:

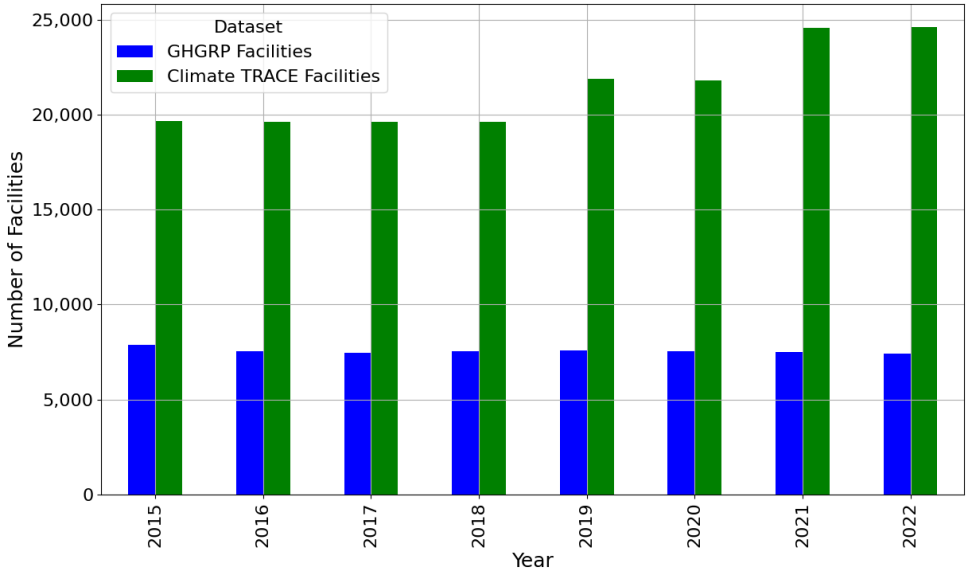


Figure 2: Facility coverage over time

Source: Own illustration

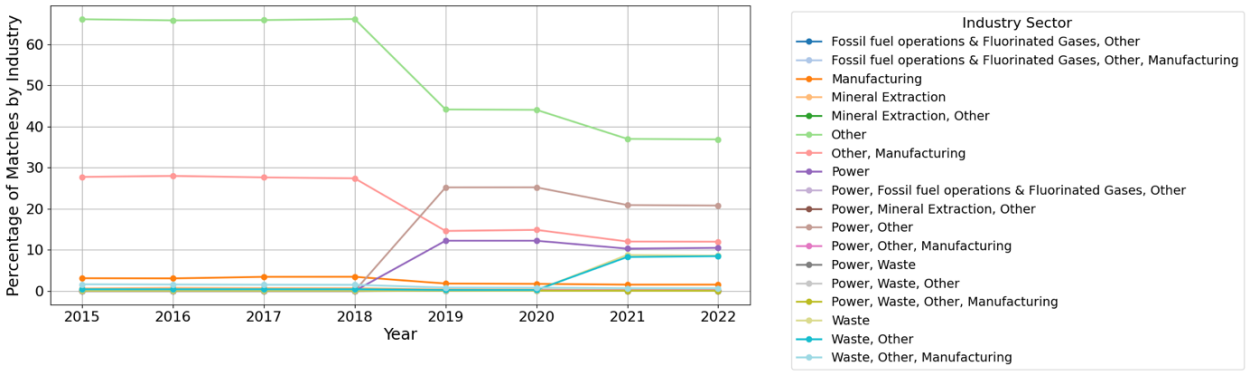


Figure 3: Industry match rates over time

Source: Own illustration

$$\begin{aligned} &Emissions\ Discrepancy \\ &= GHGRP\ Emission - CT\ Emissions \end{aligned} \tag{3}$$

resulting in direct differences in metric tons of CO2e. Using an absolute difference is consistent with previous comparative emissions studies (Gurney et al., 2024) to quantify over- or underestimation trends. I decided to use the emissions reported under GHGRP as the ground truth, as it is a legally mandated reporting system, following standardized methodologies, as explained before. CT is a relatively new initiative subject to fewer studies than GHGRP, so I assumed its data to be less reliable.

To enable additional analyses, I also calculated relative discrepancies, expressed as percentage deviations from GHGRP emissions, to normalize facility-scale effects:

$$\begin{aligned} &Percentage\ Discrepancy \\ &= \left(\frac{Emissions\ Discrepancy}{GHGRP\ Emissions} \right) \times 100 \end{aligned} \tag{4}$$

allowing for comparability across facilities with varying emission magnitudes. It should be noted that significant Emissions Discrepancies for low-emission facilities will disproportionately affect these results.

Of the 17,767 matched facility rows, 275 rows reported zero emissions discrepancies. I decided to exclude these from further calculations as they do not contribute to the analysis of discrepancies and could skew statistical interpretations by artificially lowering the observed variance in emissions differences.

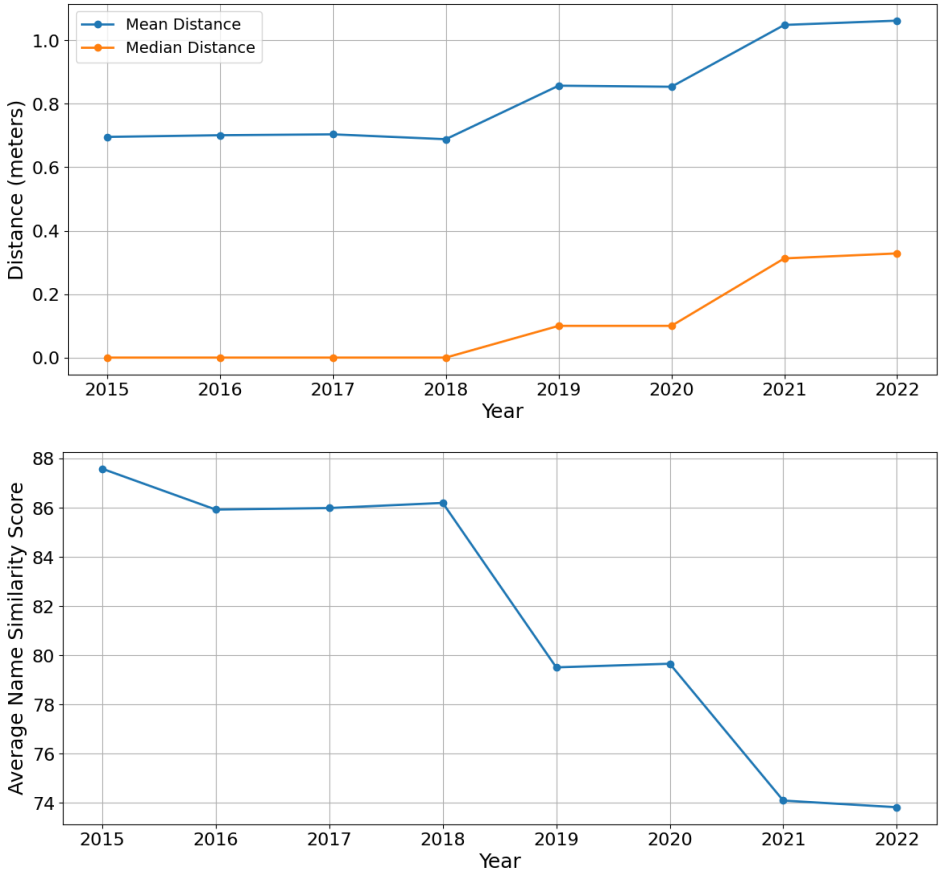


Figure 4: Distance between matched facilities (top) and name similarity over time (bottom)

Source: Own illustration

3.4 Outlier investigation

As a final data validation step, I manually reviewed the top 1% highest and top 1% lowest emission discrepancies. This was crucial in investigating the validity of these outliers, where discrepancies were extreme and potentially resulted from incorrect matches rather than genuine differences in reported emissions. I identified 15 facility-year pairs - less than 0.1% of the total matched data - that needed to be excluded from the analysis. My decision was based on manually comparing their GHGRP and CT facility names. I removed these mismatches to prevent outliers from disproportionately influencing statistical trends. This step enhanced the validity of subsequent analyses by ensuring that discrepancies are primarily driven by systemic reporting factors rather than data inconsistencies. Their low number speaks to the quality of the matching process, even before removing these false positives.

3.5 Variables and measures

Consistent with prior studies examining emission differences (Gurney, et al., 2024; Gurney, et al., 2021; He, et al., 2024; Papadopoulos, 2022), I used the absolute difference in reported emissions between the GHGRP and CT data sets, expressed as Emissions Discrepancies as the dependent variable of this paper. It is an absolute number of metric tons of CO2e, showing the extent of reporting deviations.

To explain these differences, I included a set of independent variables. The facility's GHGRP ID remains consistent across reporting years and, together with the Year of the report, enables the identification of within-facility variations over time. This allowed me to analyze temporal trends in discrepancies, particularly for facilities that report emissions in multiple years. Additionally, the reported CO2e emissions of a facility in the GHGRP dataset (Emission Tons) served as a proxy for its emission intensity, reflecting the scale of its GHG output. Each facility is classified into one or more industry categories (Subparts) according to the GHGRP classification system to capture potential industry-specific patterns. The County Name, representing the U.S. county where the facility is located, was included for potential geographic effects, such as environmental and climate conditions or infrastructure variations that may influence emissions reporting discrepancies. Finally, the parent company to which the facility belongs (Parent Company) is recorded to account for corporate-level influences on emissions reporting and potential systematic discrepancies across different organizations.

3.6 Data transformation and preprocessing

Before conducting statistical modeling, I applied multiple transformations to the dataset to address key data distribution issues, ensuring robustness in the analysis. To validate the appropriateness of parametric modeling approaches, I first examined the distribution of the dependent variable and its residuals (Figure 5). Tests for normality, including the Shapiro-Wilk test (González-Estrada, 2019) and D'Agostino's K-squared test (Statext LLC, 2024), revealed significant deviations from normality ($p < 0.001$). Additionally, the Breusch-Pagan (Bobbitt, 2020) test for heteroscedasticity indicated a significant variance instability ($p < 0.001$), suggesting that raw emissions discrepancy values were not homoscedastic across predictor variables.

These results indicated standard parametric models would be biased if applied to untransformed data. To address these issues of non-normality and heteroscedasticity, I explored several transformations on the dependent variable, including Box-Cox transformation, Log transformation, Quantile transformation, and Yeo-Johnson transformation. Each transformation was evaluated based on its impact on skewness, kurtosis, and normality test p-values. The resulting distributions can be seen in Figure 6 below.

Table 4 delivers further insights into these transformations.

Table 4: Transformation metrics

Transformation	Skewness	Kurtosis	Shapiro-Wilk (p)	Breusch-Pagan (p)
Original	0.37	49.50	2.19e-111	0.00
Box-Cox	1.16	44.96	3.16e-111	0.00
Log	-9.32	318.96	3.21e-116	0.00
Quantile	0.04	0.11	1.03e-04	0.00
Yeo-Johnson	-1.25	60.43	1.42e-111	0.00

The results of the transformation analysis, as presented in Table 4, demonstrate that the Quantile Transformation yields the most effective normalization of the emissions discrepancy variable. Compared to the original distribution, which exhibits high skewness (0.37) and extreme kurtosis (49.50), the Quantile Transformation reduces skewness to near zero (0.04). It drastically lowers kurtosis (0.11), resulting in a distribution that closely approximates normality. While alternative transformations such as Box-Cox and Yeo-Johnson improve distributional properties to some extent, they do not achieve the same level of symmetry and kurtosis reduction. The log transformation introduces extreme negative skewness (-9.32) and inflates kurtosis (318.96), making it unsuitable for this analysis. Furthermore, despite the Quantile Transformation's Shapiro-Wilk p-value (1.03e-04) indicating a statistically significant deviation from perfect normality, it remains superior to other transformations in reducing skewness and achieving a well-behaved distribution. This transformation ensured better model validity by improving adherence to the assumptions of parametric statistical methods while maintaining interpretability.

3.7 Modeling approach

Unlike previous studies on emissions discrepancies, which relied on conventional statistical techniques such as linear regression (Wang et al., 2022; Zhou et al., 2021), I implemented modern machine learning and Bayesian approaches to explain my dependent variable. My motivation is their ability to handle non-linear relationships and complex interactions that traditional regression methods may not capture ("Predictive Species and Habitat Modeling in Landscape Ecology", 2011). Additionally, as demonstrated in the previous section, parametric assumptions such as normality and homoscedasticity are violated in the raw emissions discrepancy data, so I chose methods less reliant on strict distributional assumptions (Kern et al., 2019).

A Random Forest Model is a non-parametric ensemble learning method that leverages multiple decision trees to improve predictive accuracy and reduce overfitting (Breiman, 2001). Unlike linear models, a Random Forest can automatically capture non-linear relationships between independent variables and emissions discrepancies. The method offers feature importance ranking, which allowed me to assess which predictors contribute most significantly to explaining discrepancies (Breiman, 2001). Furthermore, the model is robust to outliers and missing data, which is relevant to the data used in this paper, as identified earlier (Roy & Larocque, 2012).

I trained the model using Randomized Search Cross-Validation through the RandomizedSearchCV module from the sklearn library (scikit-learn developers, 2025), using the following hyperparameters:

- Number of trees (n_estimators = 500)
- Maximum depth of trees (max_depth = 50, None)
- Minimum samples per split (min_samples_split = 2, 5)
- Minimum samples per leaf (min_samples_leaf = 2, 3)

I employed a standard 70% training and 30% testing data split to evaluate the model's generalizability and used Mean Squared Error (MSE), R^2 score, and Mean Absolute Error (MAE) for model evaluation. After trying and comparing multiple different hyperparameter configurations, I saved the best model with the parameters above for further analysis.

Bayesian modeling allows for explicit uncertainty quantification in parameter estimates (Helleckes et al., 2022), which is particularly relevant to this paper's case of emissions accounting, where discrepancies may stem from multiple sources. Unlike frequentist models, Bayesian inference incorporates prior information, enabling more robust estimation with limited data in specific facility categories (Li et al., 2016). Using hierarchical effects allowed me to capture facility-specific, corporate-level, and regional influences on reporting discrepancies. Specifically, my model included:

- Facility-level effects to account for systematic biases in reporting at the facility level

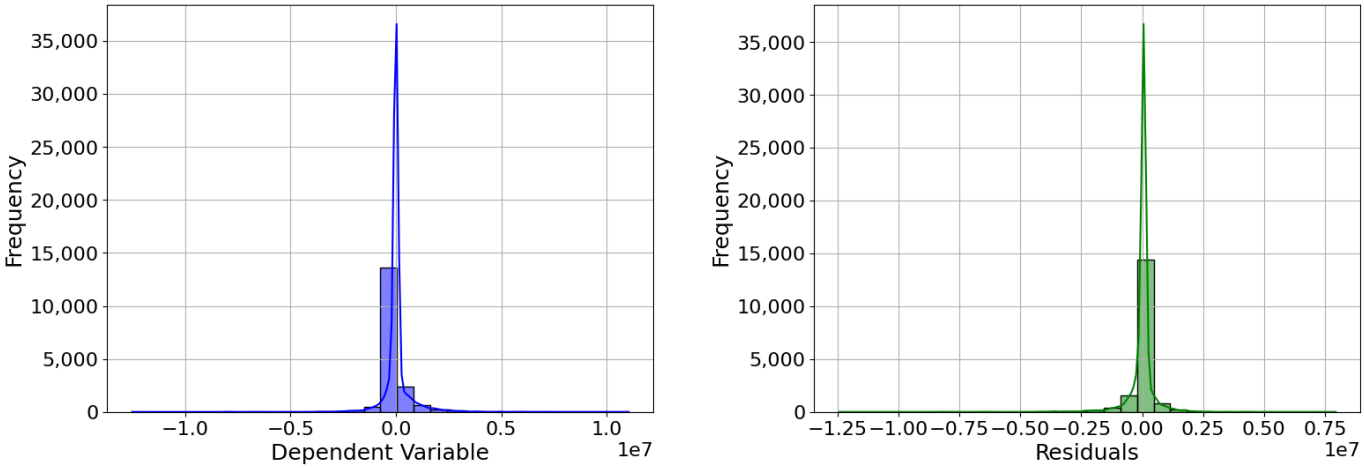


Figure 5: Original dependent variable and residual distributions

Source: Own illustration

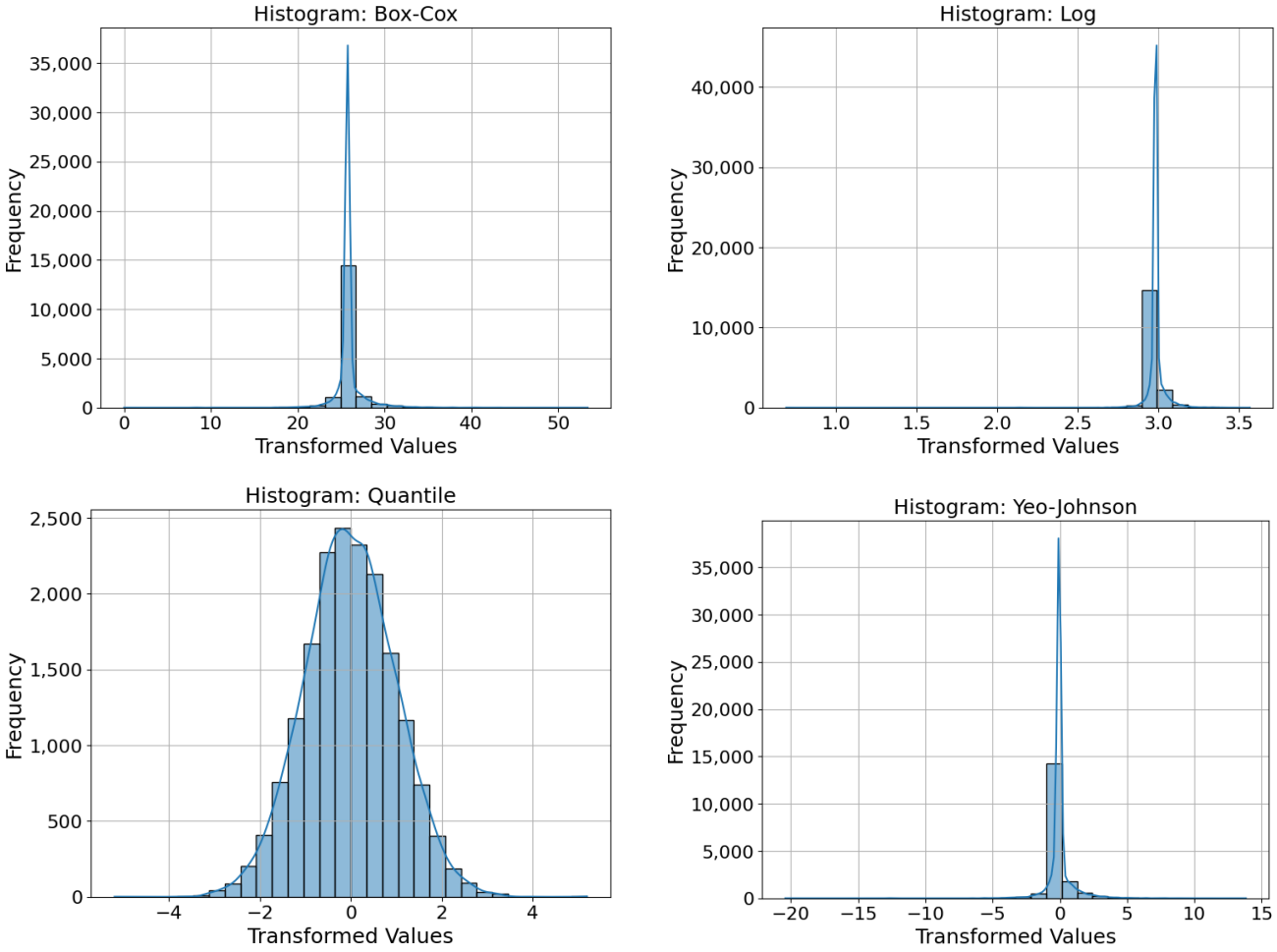


Figure 6: Overview of different transformations on the dependent variable

Source: Own illustration

- Parent company effects, as corporate structures may influence how emissions are reported

- County-level effects, capturing geographic influences on discrepancies
- Industry (Subpart) effects, accounting for sector-specific variations

I chose a Student-T likelihood function instead of a Gaussian likelihood because of the non-normal data distribution discussed earlier to account for potential heavy-tailed distributions in emissions discrepancies (Vogel et al., 2024). I then performed Markov Chain Monte Carlo (MCMC) sampling using the No-U-Turn Sampler (Hoffman & Gelman, 2014) with 1,500 posterior draws and 3,000 tuning steps, running four chains in parallel to ensure convergence, with diagnostics such as the Gelman-Rubin statistic (R-hat) checked for stability (Vehtari et al., 2021).

Comparing the two models, both offer unique advantages, further adding to the robustness of their insights. While Random Forest provides a black-box predictive model, Bayesian inference allows for interpretable parameter estimates and uncertainty quantification. By incorporating facility-specific priors, the Bayesian model can adjust for structural biases in reporting that the purely data-driven Random Forest model might not capture. Finally, I evaluated and compared both models' predictive accuracy. My Random Forest model identified key predictors of emissions discrepancies, while the Bayesian Hierarchical model provided a structured, interpretable framework for understanding uncertainty and variability in reporting discrepancies. This combined approach allowed me to develop a more comprehensive understanding of emissions discrepancies.

4 Results

4.1 Preliminary data exploration and descriptive statistics

In this chapter, I present an in-depth exploration of the dataset, focusing on patterns, trends, and key insights into discrepancies between GHGRP self-reported emissions and CT satellite-based estimates. My goal here is to describe the key statistical properties of the dataset before applying modeling techniques. This includes visualizations of the distribution of emissions discrepancies to identify patterns and systematic biases in different sectors, facility types, and geographic regions.

First, looking at the distribution of matched facilities over the distance between them (Figure 7) indicates that most matches have close to no space (0.000m) to one another.

This strongly supports the quality of the matched data. Next, the scatterplot comparing distance and emissions discrepancies (Figure 8) shows high variance at near-zero distances, suggesting discrepancies are not purely due to location and facility mismatches.

Variability decreases with distance, but discrepancies persist, indicating distance alone is not a key driver, showing no strong trend, and implying other factors drive the discrepancies. I thus concluded that the observed differences are not due to false matches.

As mentioned before, my primary variable of interest is the difference in reported emissions between GHGRP and CT, expressed in absolute metric tons of CO2e. For deeper insights, I also calculated percentage discrepancies to provide a relative difference, calculated as the absolute discrepancy divided by GHGRP-reported emissions. Finally, I computed weighted percentage discrepancies, which provide a measure accounting for the influence of facility size by weighting discrepancies based on each facility's total emissions, ensuring that large-emission facilities do not dominate the analysis disproportionately. The results for the matched data set are shown in Table 5 below.

The median absolute discrepancy of -2,175 tons CO2e suggests that at least half of the matched facilities exhibit only minor differences between the two data sources. This difference appears even more minor when considering that the median annual emissions per facility are 76,610 tons of CO2e for GHGRP facilities and 85,413.65 tons of CO2e for CT facilities. However, the mean absolute discrepancy is substantially higher (56,865 tons), indicating that extreme outliers skew the distribution. The median percentage discrepancy is close to zero (-0.0044%), but the mean percentage discrepancy is highly negative (-2,328.71%), reinforcing the influence of facilities with extreme percentage deviations. The median weighted percentage discrepancy is near zero (-3.99e-7%), while the mean weighted percentage discrepancy is -1.14%, revealing that extreme outliers strongly impact unweighted percentage calculations. The divergence between median and mean suggests that while most facilities have relatively small discrepancies, a subset of high-emission facilities accounts for disproportionately large absolute differences. The weighted percentage discrepancy further indicates that, when properly adjusted for facility size, the overall discrepancies are much more minor, suggesting that the extreme percentage deviations in the unweighted metric are not representative of the dataset. In the following sections, I explore these patterns in more depth.

I started by examining the distribution of emissions discrepancies. The histogram in Figure 9 - which I restricted to the 1st–99th percentile range of the emissions discrepancy to improve visual clarity - shows a slightly right-skewed distribution, meaning that while most discrepancies are minor, a few facilities exhibit extreme deviations. This further supports the findings from the raw discrepancy figures discussed before.

I concluded that most facilities report discrepancies close to zero but with a long tail in both positive and negative directions. Figure 10's Kernel Density Estimate plot provides a smoothed density representation of absolute discrepancies, further confirming the skewed distribution.

When looking at total emissions per data source, Figure 11 shows that GHGRP consistently reports higher total emissions per year for the matched facilities than CT. This is likely due to GHGRP's reliance on direct measurements and standardized reporting methodologies. In contrast, CT often extrapolates emissions using sector-wide models and satellite-derived estimates, which may not fully cap-

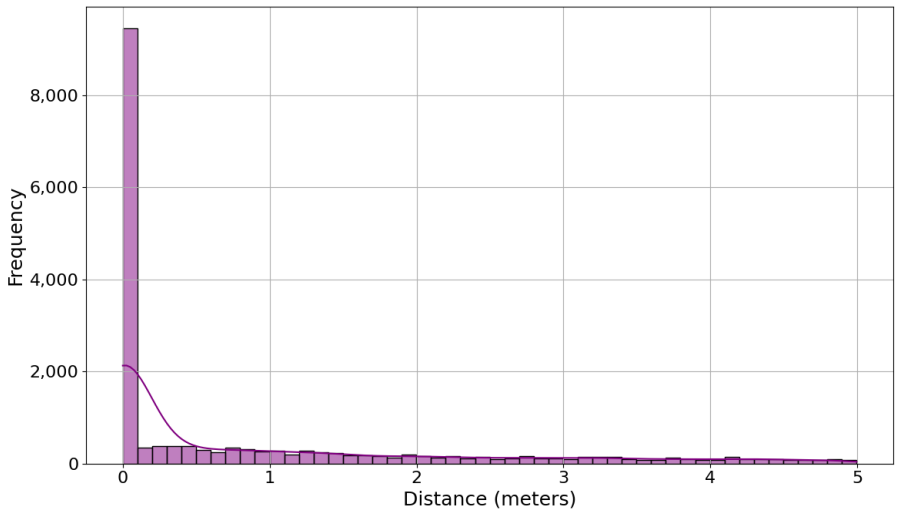


Figure 7: Distribution of matches by distance between them

Source: Own illustration

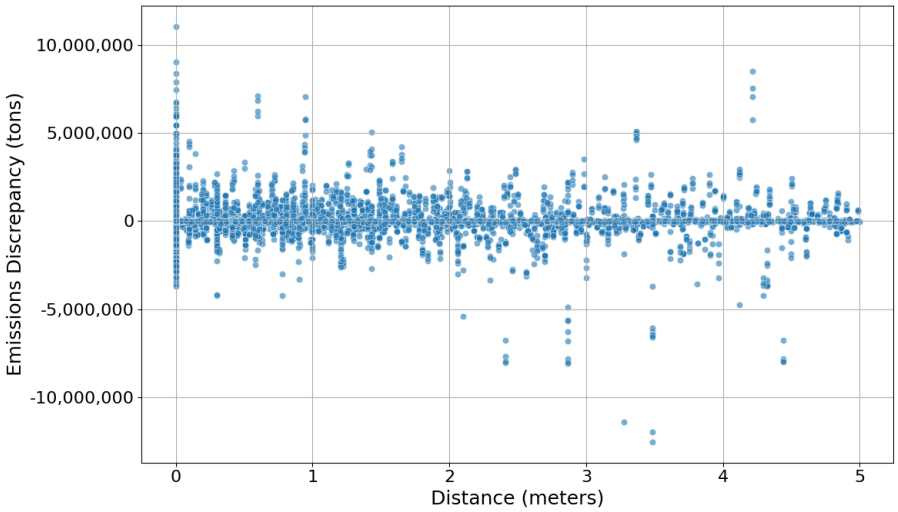


Figure 8: Distance vs. emissions discrepancy

Source: Own illustration

Table 5: Matched data emissions discrepancies

Statistic	Absolute discrepancy	Percentage Discrepancy	Weighted Discrepancy
Median	-2,175 tons	-0.0044%	-3.99e-7%
Mean	56,865 tons	-2328.71%	-1.14%

Source: Own illustration

ture specific facility-level emissions (Environmental Protection Agency, 2024c; Gurney et al., 2024). GHGRP also includes some fugitive and supplier-based emissions, which CT may not comprehensively account for, further contributing to the discrepancy (Dogniaux et al., 2024; Hanle et al., 2012).

The sharp increase in total emissions in both datasets from 2018 onwards can be explained through the combination of an increase in the number of matched facilities from 1494 in 2015 to 2,738 in 2019 and finally, 3,200 facilities in 2022, as well as the increase in the average emissions per facility, shown in Figure 12.

Looking at how discrepancies fluctuate over time, see Fig-

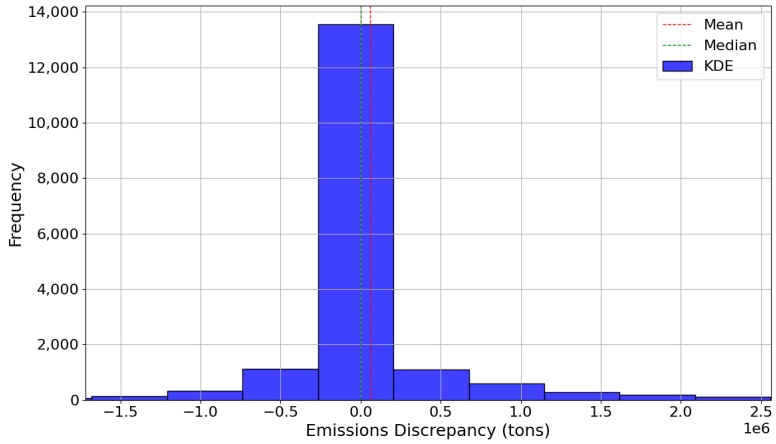


Figure 9: Histogram of absolute emissions discrepancies (1st to 99th percentile)

Source: Own illustration

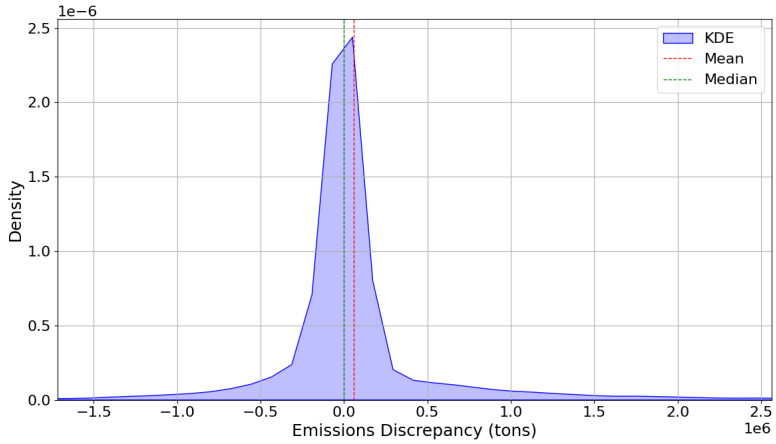


Figure 10: Kernel Density Estimate plot

Source: Own illustration

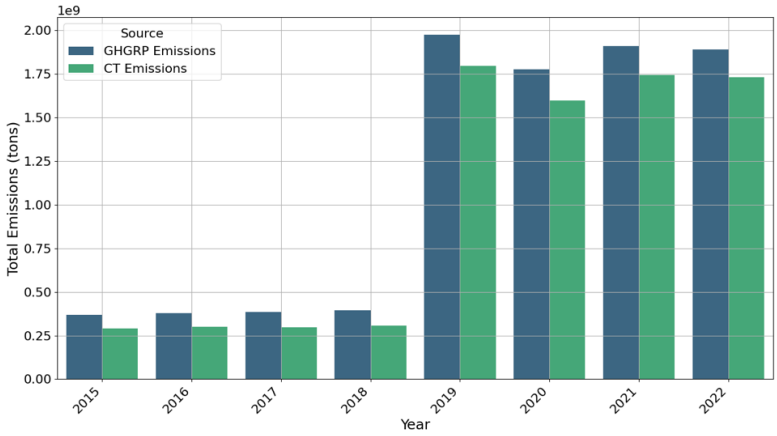


Figure 11: Total emissions by year (GHGRP vs Climate TRACE)

Source: Own illustration

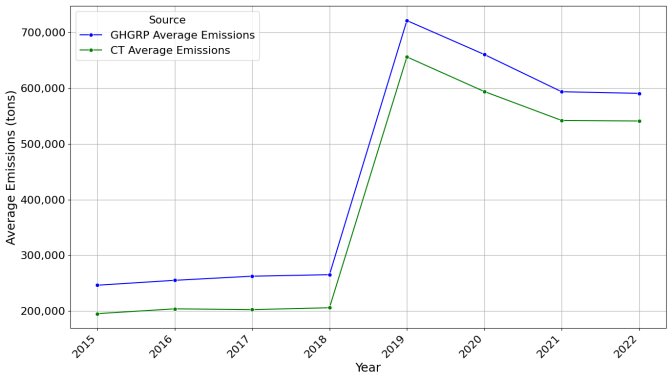


Figure 12: Average facility emissions per year (GHGRP vs Climate TRACE)

Source: Own illustration

ure 13. For absolutes - the left graph - shows a steady trend from 2015 to 2016, followed by a sharp increase until a peak in 2020. From then on, discrepancies declined in 2020 to below even in 2015. The discrepancies are always positive, in line with the total emissions per data source, showing higher emissions for GHGRP on average. However, this flips for relative discrepancies, and the percentage discrepancies become negative. This further shows how extreme outliers strongly impact unweighted percentage calculations.

Figure 14 displays the distribution of emissions discrepancies for each year to further examine how discrepancies evolved. While the previous trend analysis highlighted mean shifts, this visualization provides deeper insights into the dispersion and presence of outliers across different years. The boxplots reveal that while the median discrepancy remains relatively stable, the spread of discrepancies remains vast, with strong outliers in both directions across all years.

This distributional view strengthens prior observations that a subset of facilities consistently exhibits extreme discrepancies rather than random fluctuations affecting all facilities equally.

The next step was to examine how weighing emissions discrepancies by facility size affects these patterns (Figure 15). By adjusting for facility size, I could better assess whether the observed patterns hold across all emitters or if a subset of facilities primarily drives them.

This adjustment confirms that the extreme percentage discrepancies are primarily driven by small facilities, where even minor absolute differences lead to large relative deviations. When accounting for facility size, the overall trend in discrepancies remains more stable, indicating that the observed variations are less pronounced among larger emitters.

Finally, the impact of high-emissions facilities is made more evident by the direct relationship between discrepancies and the emissions reported under GHGRP (Figure 16).

A clear pattern emerges here: more emission-intensive facilities tend to have more significant absolute discrepancies. Notably, some high-emission facilities report extreme underestimations or overestimations.

I concluded that the divergence between absolute and percentage discrepancies stems from facility size effects. Smaller facilities often exhibit extreme percentage discrepancies despite minor absolute differences, while larger-emitting facilities dominate absolute discrepancies but show more minor relative errors. The weighted percentage discrepancy corrects for this imbalance, revealing that extreme unweighted percentage errors are largely statistical distortions. When adjusted for facility size, the overall discrepancies are much more minor, indicating that the high negative mean percentage discrepancy is driven by a few small emitters rather than systemic issues.

This showed that the differences in reported emissions are not static and have evolved significantly over time. As I discussed before, this increase coincided with an expansion of CT's dataset, suggesting that improvements in satellite-based monitoring may be influencing discrepancies.

Next, I explored if geographic patterns emerge, which might help understand the observed discrepancies.

The graphs in Figure 17 clearly show significant differences between the states, with certain states (e.g., North Dakota, Wyoming) exhibiting substantial positive discrepancies, suggesting GHGRP may systematically underreport emissions or CT may overestimate them. Other states (e.g., Georgia and Puerto Rico) show negative discrepancies, indicating potential over-reporting in GHGRP or under-detection by CT. This observation further stresses the need to consider the geographical location of facilities, which I did in the later statistical modeling.

At this point, the facilities' industry, or subsector, remained the last unexplored data parameter. To better understand its potential impact on emissions discrepancies, see Figure 18.

The first key observation is that the total amount of emissions reported is not equal between the different sectors, with a few industries like energy generation, manufacturing, or cement production dominating the reported emissions.

The plot in Figure 19 displays the relative percentage discrepancies across sectors.

International and domestic aviation exhibit the most negative discrepancies, suggesting GHGRP underreports emis-

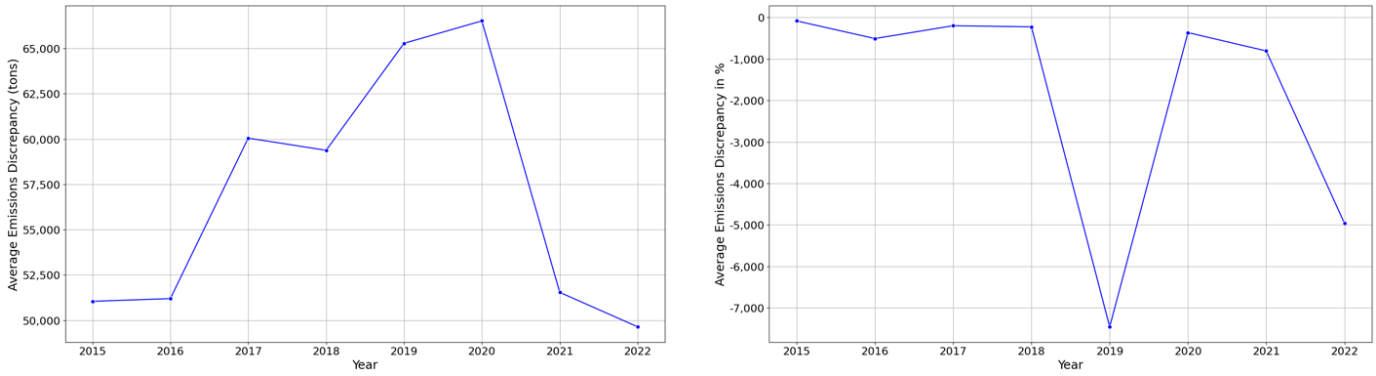


Figure 13: Temporal trend in emissions discrepancies (left: absolute, right: relative)

Source: Own illustration

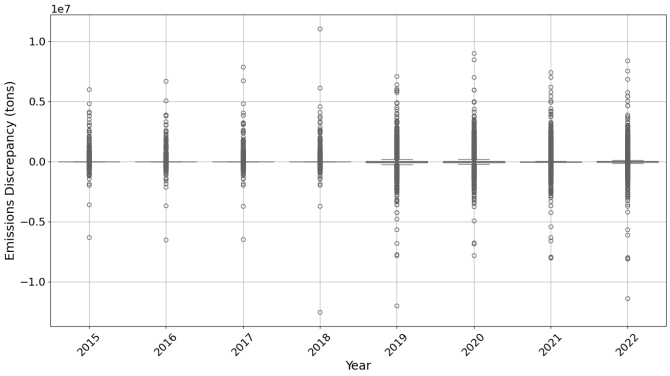


Figure 14: Distribution of emissions discrepancies by year

Source: Own illustration

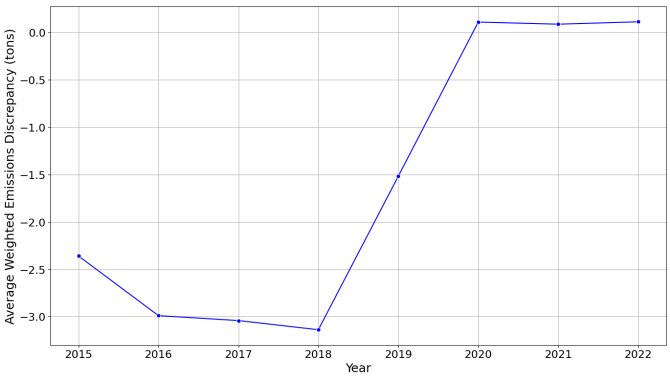


Figure 15: Temporal trends in weighted emissions discrepancies

Source: Own illustration

sions or CT overestimates them. As these percentages are unweighted and Figure 18 shows that these sectors have little total emissions, their significant percentage gaps dominate the plot, even though the total discrepancies are minimal. This point is further underscored by looking at the outliers for these sectors in Figure 20 below.

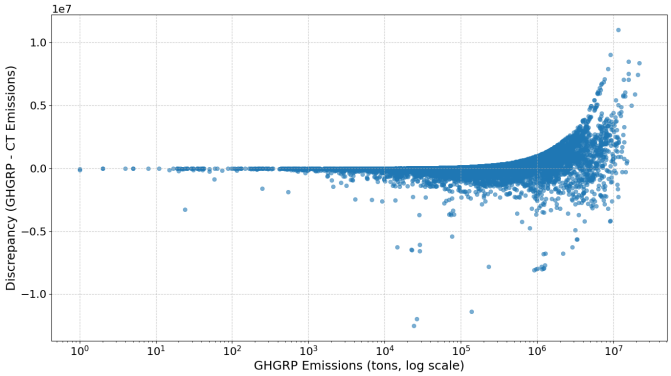


Figure 16: Emissions discrepancies vs facility size

Source: Own illustration

Electricity generation and international aviation also show substantial negative discrepancies, while sectors like aluminum and domestic shipping have relatively small percentage deviations. These differences highlight sector-specific biases that must be accounted for in further analysis.

Figure 21 shows the top 10 highest emitting facilities by average GHGRP emissions as a final insight.

Interestingly, all these facilities belong to the energy generation subsector.

In summary, this preliminary data exploration showed that emissions discrepancies, with sectoral and geographic differences, are not random. Furthermore, extreme outliers strongly influence mean values, necessitating careful statistical treatment, which I performed in the following section.

4.2 Main model results

This section presents the results from my Random Forest and Bayesian Hierarchical modeling. Both were applied to assess the key predictors of discrepancies between GHGRP self-reported and CT satellite-based emissions. As mentioned, the models differ in their approach: Random Forest captures non-linear interactions and produces substan-

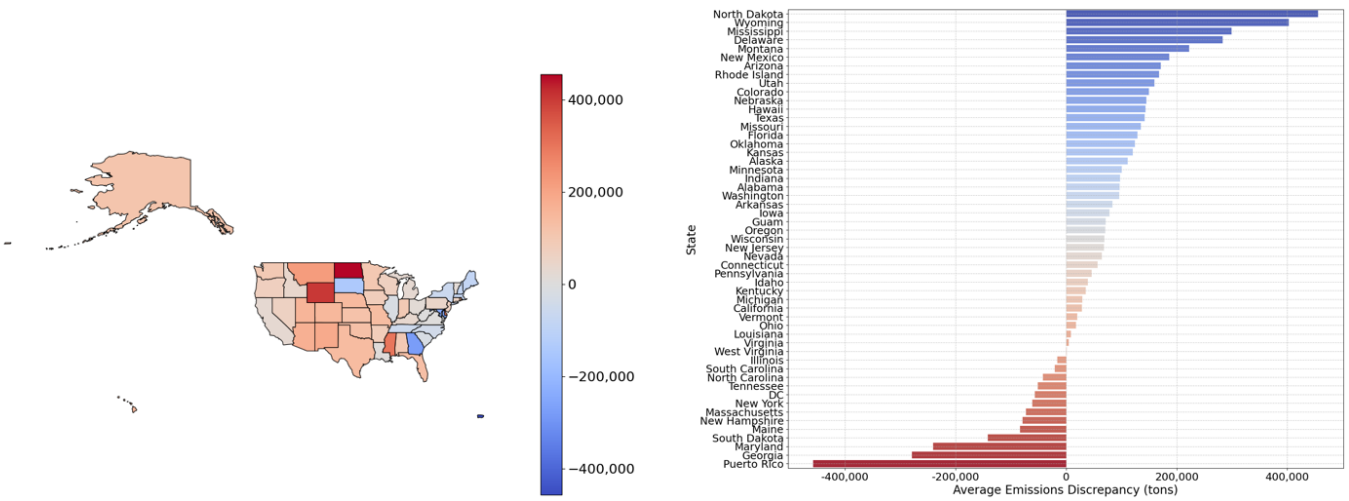


Figure 17: Emissions discrepancies by US State

Source: Own illustration

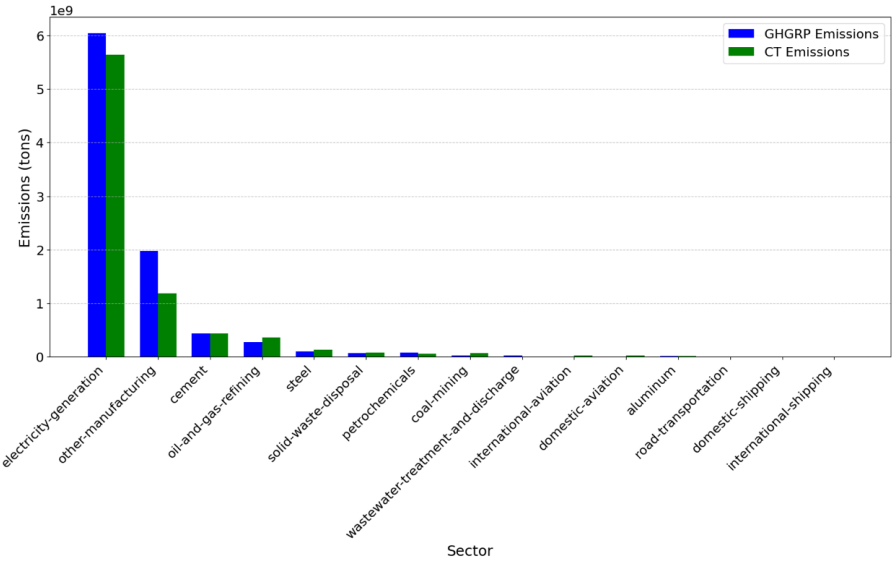


Figure 18: Sector-wise emissions comparison

Source: Own illustration

tial predictive accuracy, while Bayesian Hierarchical Modeling provides uncertainty quantification and variance decomposition. I first present each model individually, followed by a comparative analysis of their performance and insights.

The Random Forest model was evaluated using R^2, RMSE, and MAE to assess predictive accuracy. Cross-validation results were also analyzed to confirm robustness. The model exhibits strong predictive performance (Train $R^2 = 0.8926$, Test $R^2 = 0.8677$). Cross-validation across five folds confirms the model's generalizability (Mean $R^2 = 0.8410$) with minimal overfitting. The model also predicts discrepancies with a relatively small error (Test MAE = 0.2821); see Table 6 below.

These results suggest that the key predictors identified by the model generalize well beyond the training data, making this approach reliable for understanding emissions discrepancies.

Figure 22 presents the residuals vs. predicted values (left) and a Q-Q plot (right) to assess model performance beyond summary metrics.

The residual plot suggests a pattern of heteroskedasticity, where prediction errors tend to be more significant for extreme values, indicating that the model struggles more with high discrepancy cases. This suggests that the model may not fully capture some important factors influencing discrepancies. Additionally, the Q-Q plot shows deviations from a normal distribution, particularly at the tails, further confirming the presence of outliers that the Random Forest

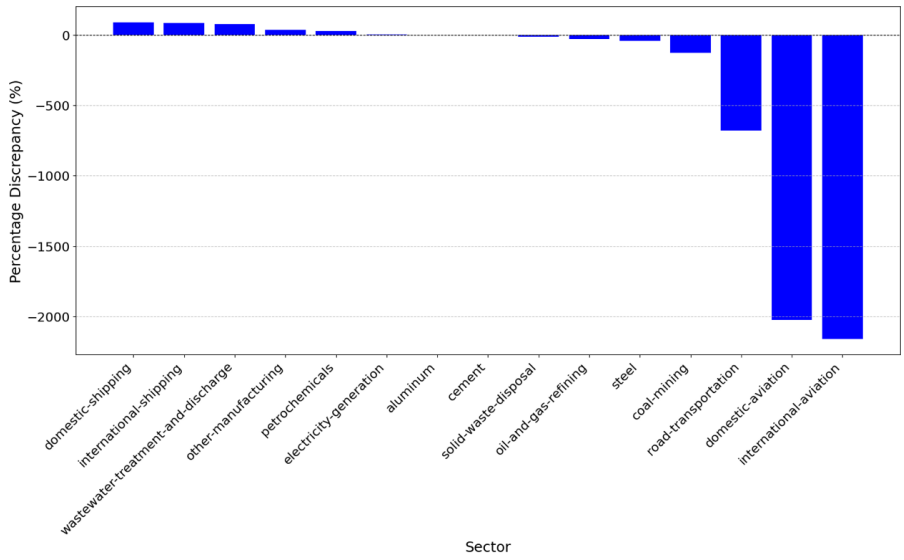


Figure 19: Percentage discrepancies by sector

Source: Own illustration

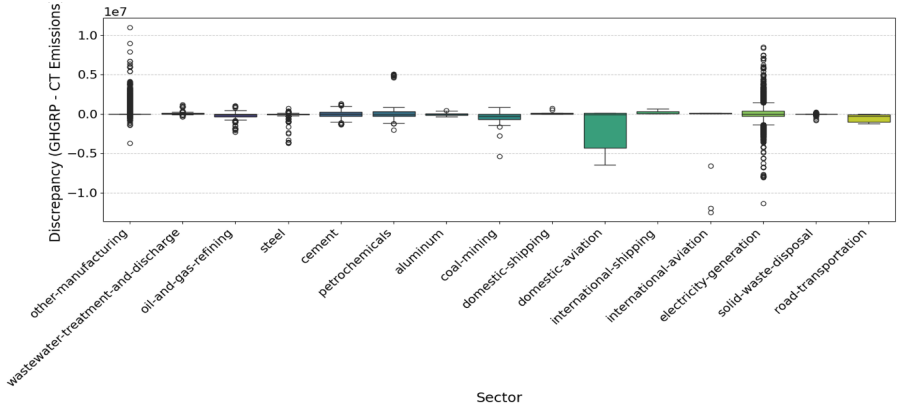


Figure 20: Discrepancies in emissions by sector

Source: Own illustration

model does not fully account for. These findings indicate that while the model performs well overall, it may be less reliable for predicting extreme discrepancies and could benefit from additional feature engineering or alternative modeling approaches. These diagnostics highlight the need for a robust statistical treatment of outliers and uncertainty quantification, which I later addressed with my Bayesian model.

I conducted a bootstrapped permutation importance analysis – which estimates feature importance by repeatedly shuffling the outcome variable and using bootstrap sampling to correct biases and improve reliability (Altmann et al., 2010) - to understand what drives emissions discrepancies according to my model, summarized in Table 7.

Facility-specific effects (GHGRP ID) dominate (40.78%) the explained differences in reported emissions between GHGRP and CT. This suggests that emission discrepancies depend highly on the individual facility rather than purely sectoral or geographic trends. Differences in reporting practices, operational factors, or data quality at specific facilities likely play a significant role. County and Parent Company Effects (24.23% and 22.56%) also contribute significantly, hinting at geographic or climate effects and corporate-level report-

Table 6: Random Forest performance metrics

Metric	Value
Train R^2	0.8926
Test R^2	0.8677
Cross-val. Mean R^2	0.8410
Cross-val. R^2	0.8419
Test MAE	0.2821

Source: Own illustration

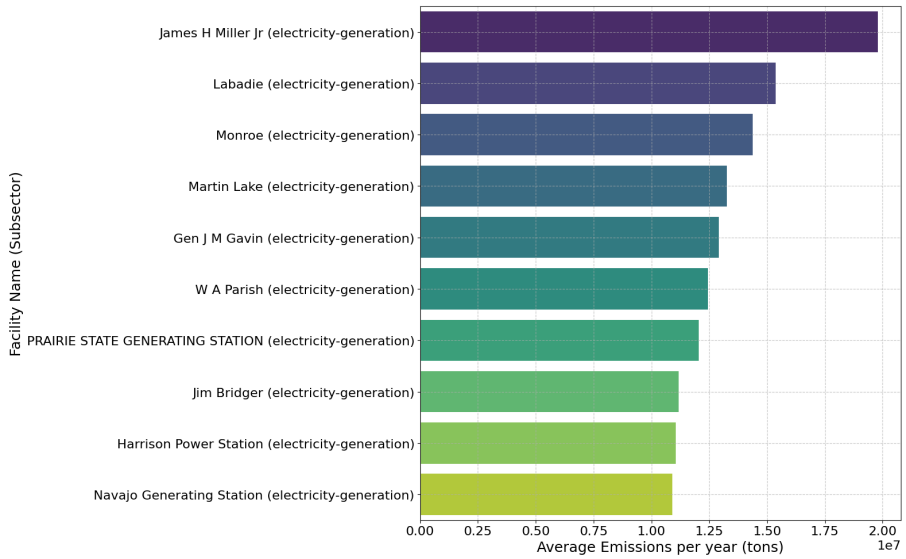


Figure 21: Top 10 emitters by average GHGRP emissions

Source: Own illustration

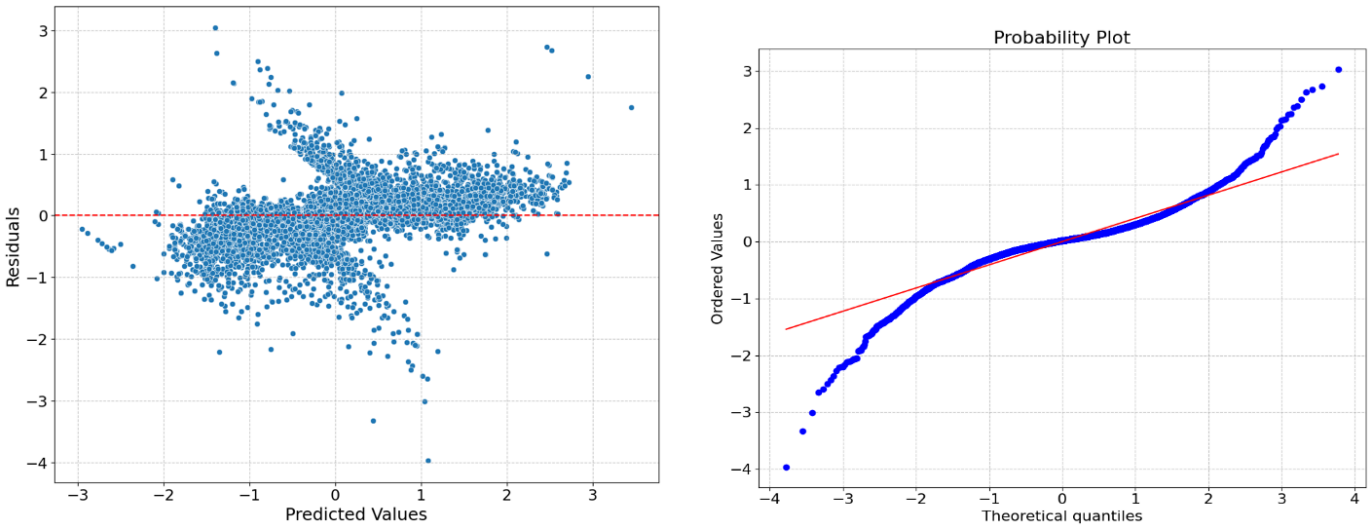


Figure 22: Residuals vs. predicted values (left) and Q-Q plot of residuals (right)

Source: Own illustration

ing strategies affecting emissions data. The industry Subpart (10.33%) to which a facility belongs also contributes considerably, indicating that specific industries or their methodologies may lead to systematic discrepancies. The reporting Year (2.11%) plays a minor role, showing that discrepancies do not change significantly over time, contrary to potential expectations that improved satellite monitoring might reduce them. Interestingly, after accounting for all other factors, the total Emissions Volume (0%) does not directly contribute to discrepancy predictions in this model. This contradicts prior assumptions that absolute emissions alone would drive discrepancies.

In summary, the Random Forest Model is highly predictive, but residual diagnostics reveal systematic variance in extreme cases, necessitating additional modeling approaches. Observed facility-level characteristics dominate the variance, while county and parent-company factors also play a decisive role. Emission volume is not a meaningful predictor, reinforcing the need to consider facility-specific biases and reporting practices. These findings suggest that discrepancies are not random but instead driven by structured facility, corporate, and regional factors - highlighting the importance of multilevel modeling.

I employed the Bayesian model to complement the Random Forest analysis by incorporating uncertainty quantification and variance decomposition. This approach enabled me

to better understand the contribution of different hierarchical levels to the observed discrepancies.

Table 7: Random Forest model results

Normalized group contributions	Mean importance	Contribution (%)
Facility-Specific Effects (GHGRP ID)	0.16	40.78
County-Level Effects	0.09	24.23
Parent Company Effects	0.09	22.56
Subparts (Industry)	0.04	10.33
Year of Reporting	0.00	2.11
Emission Volume (tons)	0.00	0.00

Source: Own illustration

The Bayesian model exhibited strong predictive performance, with an overall Bayesian R^2 of 0.8660 and a test R^2 from cross-validation of 0.8659 ± 0.0141, closely aligning with the Random Forest model. The RMSE and MAE were 0.429 and 0.176, respectively, indicating relatively low error levels; see Table 8 below.

Table 8: Bayesian performance metrics

Metric	Value
Bayesian R^2	0.8660
RMSE	0.4290
MAE	0.1760
Cross-Validation R^2	0.8659 ± 0.0141
R-hat (Convergence)	$\leq$ 1.01 (Stable)

Source: Own illustration

To assess the reliability of the estimates, I examined convergence diagnostics (R-hat values) and adequate sample sizes (ESS). Most parameters demonstrated good convergence (R-hat $\approx$ 1.0), but some parameters exhibited slight deviations, suggesting minor instability. Furthermore, I detected two divergences during sampling, indicating that future refinements (e.g., adjusting the step size or target acceptance rate) may improve model stability.

Figure 23 presents the residual distribution across cross-validation folds, indicating that residuals are symmetrically distributed around zero, with some variation at the tails. Compared to the Random Forest residuals, which exhibited clear heteroskedasticity, the Bayesian model's residuals appear more stable, suggesting that the hierarchical structure helps capture variance at different levels more effectively.

The variance decomposition analysis revealed that facility-level effects dominate, explaining approximately 93.95% of the variance, significantly higher than the 40.78% found in the Random Forest model. This suggests that the Bayesian model captures a more hierarchical structure, reinforcing that emissions data discrepancies are highly facility-dependent.

Table 9 summarizes the variance contributions across different hierarchical levels.

These findings confirm that individual facilities account for most observed discrepancies (93.95%), while County (3.67%) and Parent Company (1.6%) effects play secondary roles. Interestingly, the Subpart (0.01%) contribution is negligible, indicating that systematic discrepancies are more facility-specific than sector-wide. I omitted the emissions variable altogether, as it previously proved insignificant, and drastically reduced the model's computational burden.

Figure 24 presents the posterior distributions of key parameters, including uncertainty intervals for beta (fixed effects) and variance components.

The relatively narrow spread of these distributions suggests model stability.

Having analyzed both models separately, I now compared their performance and interpretability.

Both models exhibit similar R^2 values ($\sim$0.86), suggesting they are equally effective at capturing the key predictors of emissions discrepancies. However, the Bayesian model achieves slightly lower RMSE and MAE, indicating marginally better predictive accuracy. Table 10 presents a side-by-side comparison of factor contributions and an average between them.

Table 10: Random Forest and Bayesian model comparison

Group contributions	Random Forest (%)	Bayesian Model (%)	Mean (%)
Facility-Specific Effects (GHGRP ID)	40.78	93.95	67.37
County-Level Effects	24.23	3.67	13.95
Parent Company Effects	22.56	1.60	12.08
Subparts (Industry)	10.33	0.01	5.17
Year of Reporting	2.11	-	1.05
Emission Volume (tons)	0.00	-	0.00

Source: Own illustration

Table 9: Bayesian variance contributions

Variance contribution	Mean %	Lower CI (2.5%)	Upper CI (97.5%)
Facility Effect	93.95	79.53	99.04
County Effect	3.67	0.45	17.74
Parent Effect	1.60	0.002	7.90
Subparts Effect	0.01	0.00	0.07
Residual	0.77	0.65	0.84

Source: Own illustration

While the reporting year is a minor factor in both models, this comparison highlights key differences:

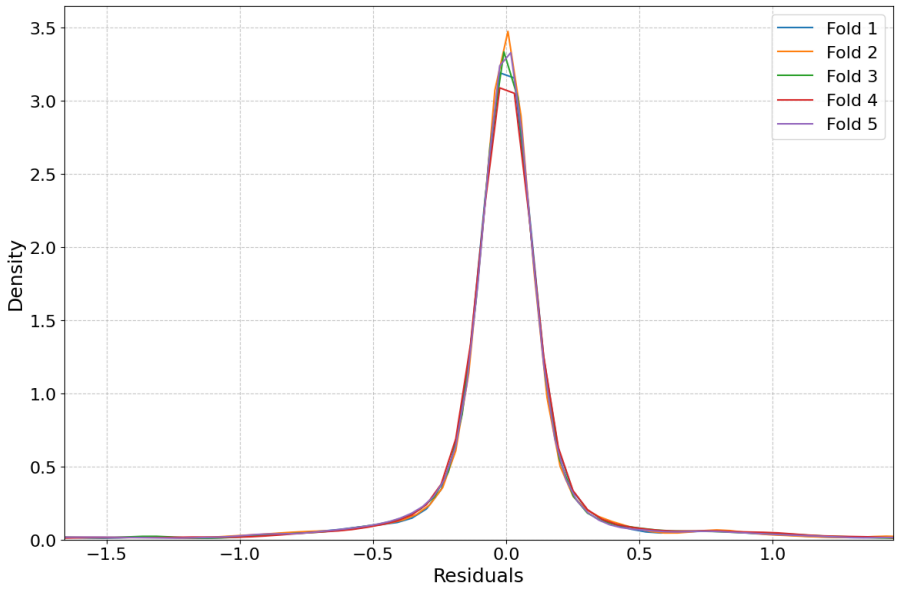


Figure 23: Bayesian residual distribution across folds

Source: Own illustration

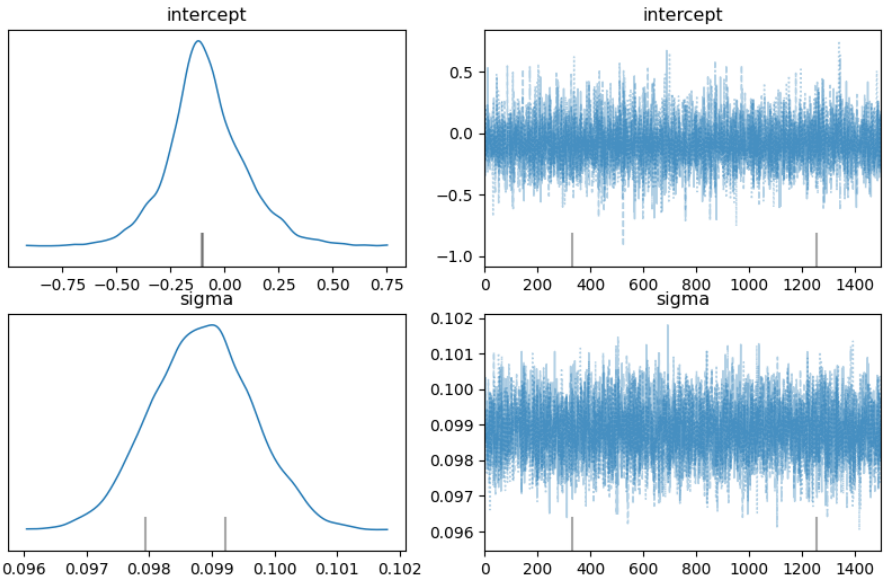


Figure 24: Bayesian variance contributions

Source: Own illustration

- The Bayesian model attributes most variance to facility-specific effects (93.95%), while the Random Forest model suggests that county- and parent-level effects also play a prominent role (24.23% and 22.56%, respectively).
- Industry (subpart) effects are negligible in the Bayesian model but contribute 10.33% in Random Forest, indicating that some industries may have systematic reporting differences.

However, both models identified facility-specific (GHGRP ID) effects as the major contributor, followed by County and Parent Companies. The Subpart (industry) of the facility, the Year of the report, and the Emissions volume are negligible.

4.3 Sensitivity analyses

To validate the results of this paper and ensure they are robust to alternative data and model configurations, I performed several sensitivity analyses.

4.3.1 Using different distance thresholds for coordinate-based matching

As mentioned, I decided on a 5-meter radius for the coordinate-based matching during my primary analyses, as this distance represents a conservative approach. The primary rationale behind this choice was to reduce the number of false-positive matches to a minimum while allowing for reasonable distances between potential matching facilities. However, this was still an individual selection and not empirically validated, so I set out to find the computationally optimal matching distance for the dataset. First, I reran the coordinate-based matching between GHGRP and CT data on multiple distance thresholds. Next, I performed a similarity evaluation of the names of the matched facilities of the two datasets using the token_set_ratio function from the rapidfuzz module, which returns a similarity score between 0 and 100 (rapidfuzz, 2025). I did this assuming that, on average, a higher name similarity between the two datasets' facility names implied a better match. Finally, I calculated the distances between matched facilities for the different thresholds, again assuming that the closer the two matched facilities are, the higher the likelihood of a correct match. An overview of the matching results can be seen in Table 11.

Table 11: Results of different distance thresholds on coordinate matching

Thresholds (m)	% Unique facilities	% Total rows	Median facility distance (m)	Median facility name similarity
5	36.40	28.96	0.00	100.00
10	42.04	33.07	0.28	100.00
15	45.52	36.11	0.56	100.00
20	47.60	38.17	0.78	100.00
25	49.19	39.91	0.95	100.00
30	50.76	41.80	1.20	100.00
35	52.36	43.60	1.43	100.00
40	53.94	45.60	1.80	100.00
45	55.02	47.00	2.09	100.00
50	56.04	48.40	2.43	100.00
75	60.68	55.05	4.35	100.00
100	63.32	58.69	6.35	84.21
500	77.27	72.62	25.35	54.55
1000	85.23	79.56	44.64	46.15

Source: Own illustration

I combined the three factors (percentage of matched rows, distance between facilities, and name similarity) into one score to calculate the optimal distance thresholds. First, I normalized them between 0 and 1 to compare them on the same scale. Then, I assigned weights to decide how important each factor was. For example, one set of weights might assign 40% to matched percentage, 40% to median distance, and 20% to name similarity. All weight combinations always added up to 1. For each weight combination:

1. I calculated an overall score for each threshold by multiplying each normalized factor by weight and adding them.
2. I identified the best threshold for that specific weight combination (the one with the highest score).

Figure 25 shows the effect of the different weight combinations on the threshold score.

Finally, I calculated the average overall score for each threshold by averaging its scores across all weight combinations. As shown in Figure 26, this allowed me to rank the thresholds and identify the one that performed best.

For 14 of 18 possible weight combinations, the threshold of 75m delivered the highest score. Since it also resulted in the highest average score, I concluded that this distance threshold delivers the best possible matches.

This new threshold increased the number of total matches by 90.07% from 17,772 to 33780 rows. Compared to the 5m data, the mean distance between matches increased from 0.87m to 16.84m, and the mean facility name similarity decreased from 92.39 to 72.01.

To assess if this increased maximum distance between facilities would change the results of my analysis, I reran the Random Forest Model used before on the new dataset. To ensure comparability with the 5-meter dataset, I did not change any of the model's hyperparameters. Table 12 shows the results of this analysis.

Consistent with earlier findings, facility-specific effects (GHGRP ID) remain the dominant contributors to the explained variance in the model. The relative contributions of predictors show some minor shifts when comparing the 5-meter and 75-meter datasets. GHGRP Parent Companies, for instance, increased from 22.6% to 24.1%, while GHGRP ID decreased slightly from 40.8% to 37.7%. Similarly, County

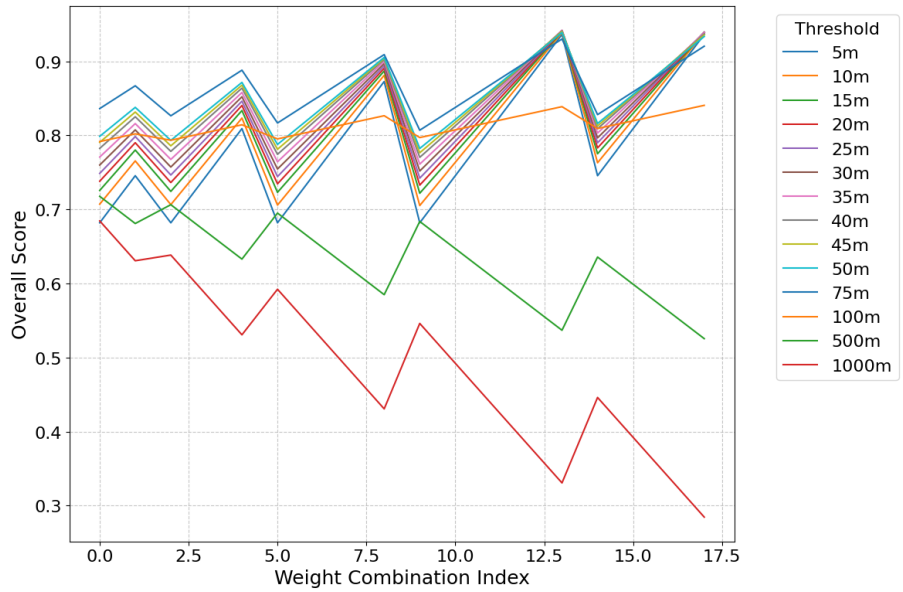


Figure 25: Variation in overall scores across weight combinations

Source: Own illustration

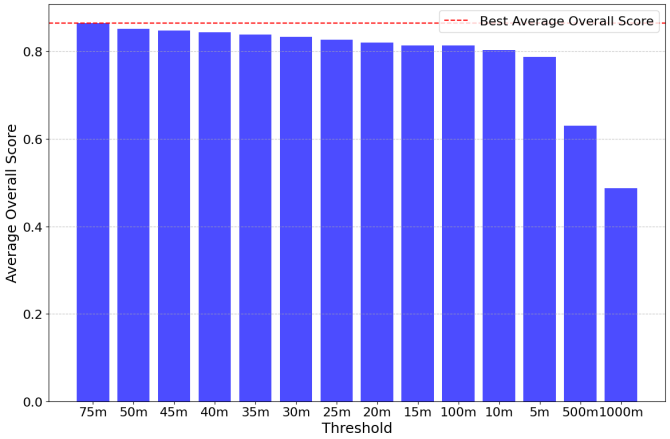


Figure 26: Average overall score across all weight combinations

Source: Own illustration

Table 12: Comparison of Random Forest results: 5m and 75m radius data

	5m	75m
Train R^2	0.8926	0.8779
Test R^2	0.8677	0.8498
CV Mean R^2	0.8410	0.8292
Mean Std Dev R^2	0.0419	0.0408
Contributions to explained variance:		
GHGRP ID	40.8%	37.7%
County Name	24.2%	20.1%
Parent Companies	22.6%	24.1%
Subparts	10.3%	14.8%
Year	2.1%	3.2%
Emissions tons	0.0%	0.0%

Source: Own illustration

Name decreased from 24.2% to 20.1%, while Subparts increased from 10.3% to 14.8%. These shifts likely reflect the inclusion of additional matches in the larger dataset, which introduces a broader range of facility and industry characteristics.

However, it is essential to note that these changes do not significantly alter the conclusion that facility-specific effects still dominate the explained variance. The contributions of shared effects, such as geographic location (County Name) and industry-specific characteristics (Subparts), remain secondary and show only marginal adjustments in their relative importance.

Moreover, the performance metrics of the Random Forest model remain robust. Train R^2 and Test R^2 show slight declines, from 0.8926 to 0.8779 and 0.8677 to 0.8498, respectively, while the cross-validated mean R^2 drops marginally from 0.8410 to 0.8292. The consistency of these metrics, even when the dataset size nearly doubles, demonstrates that the model's predictions are stable and not highly sensitive to changes in the spatial matching radius.

These results reinforce the robustness of my earlier findings. The slight shifts in predictor contributions do not undermine the central conclusion: emissions discrepancies are primarily driven by facility-specific factors, with shared effects such as geographic and industry-level characteristics contributing to a lesser extent.

4.3.2 Excluding matches with minor emission differences

Figure 27 shows the differences between the emissions reported by GHGRP and CT in absolutes.

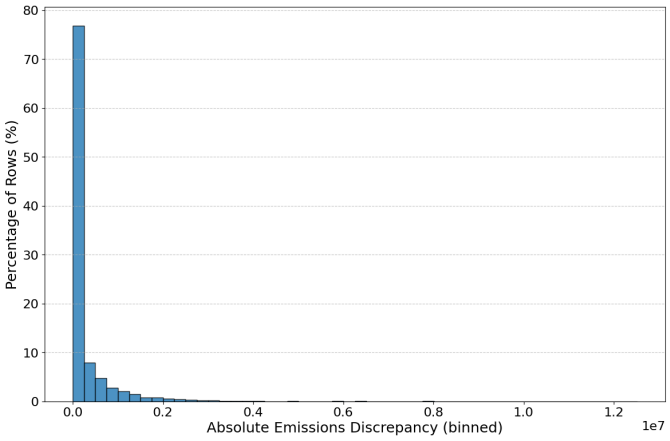


Figure 27: Distribution of absolute emissions discrepancies

Source: Own illustration

Here, most (70.23%) of the matches - the first bin - have an absolute difference of between 0 and 125,147.31 tons, which is only 0.009% of the maximum difference of 12,514,731.42 tons of CO2e. Since my earlier analyses identified facility-specific effects as the main drivers of the emissions discrepancies, I tested if this trend upholds when excluding matches with minimal emission differences from the dataset. I hypothesized that excluding matches with minimal differences may reveal more apparent patterns in the predictors of the dependent variable, as these minor differences might mask the proper drivers of variance in the dataset. To do so, I ran two additional Random Forest models, using the same hyperparameters again but excluding the data either ±1 or ±2 standard deviations around the mean of the above distribution. I chose this as it offers a data-driven approach to excluding minimal discrepancies, avoiding the arbitrariness of setting a fixed threshold, and ensuring consistency. The resulting two distributions are displayed in Figure 28 below.

This reduced the number of matches from 17,772 to 1,979 and 801 rows, respectively. Table 13 compares the results of the Random Forest models on the complete data with the results on the two reduced datasets.

Excluding matches with minimal discrepancies led to some shifts in the results, but the core findings remain robust. I observed increases in Train R^2, Test R^2, and Cross-Validation Mean R^2, with the Test R^2 rising from 0.8677 in the entire dataset to 0.9117 and 0.9118 in the ±1 and ±2 standard deviation datasets, respectively. This suggests that removing minor discrepancies reduces noise, allowing the model to better capture meaningful patterns. Since the gap between train and test R^2 did not significantly increase, overfitting did not either.

The contributions of predictors also shifted. GHGRP ID's importance decreased slightly (from 40.8% to 39.6% and 35.6%), while County Name's contribution increased significantly (24.2% to 31.0% and 33.0%). Parent Companies remained relatively stable, with a slight rise in the ±2 dataset (24.6%). Subparts decreased noticeably, from 10.3% in the whole dataset to below 5% in both reduced datasets. Year and Emissions Tons showed only minimal changes, with the latter remaining negligible.

These shifts suggest that removing minor discrepancies amplifies the role of shared factors like geography (County Name) while slightly reducing the dominance of facility-specific effects (GHGRP ID). However, the overall conclusions remain consistent: facility-specific effects still drive discrepancies in emissions reporting, with County Name and Parent Companies playing complementary roles. The robustness of the Random Forest model across datasets, as shown by stable performance metrics, supports the reliability of my earlier findings.

4.3.3 Using traditional models: Comparing results with Random Forest and Bayesian Models

As a final test of robustness, I applied two more traditional modeling approaches: ANOVA and Elastic Net regression. These models provide alternative perspectives on the drivers of emissions discrepancies and allow for a comparison with the results obtained from the primary analysis models. While Random Forest models excel at capturing complex, nonlinear relationships and Bayesian models offer probabilistic insights, ANOVA and Elastic Net offer interpretability and insight into variable significance through linear and variance decomposition frameworks (Breiman, 2001; Hans, 2012; Helleckes et al., 2022; Ståhle & Wold, 1989).

The ANOVA model explains 91.0% of the variance in emissions discrepancies ($R^2 = 0.910$, F(4732, 24298)=1.114*10**6, $p < 0.001$). Consistent with my primary analysis results, facility-specific effects dominated the explained variance. GHGRP ID accounted for 84.6% of the variance, significantly higher than all other predictors. Parent Companies contributed 14.5%, while County Name accounted for 0.9%. Subparts and Year's contributions were negligible, explaining less than 0.1% of the variance. Emissions Tons had no significant contribution. Table 14 shows these results.

The ANOVA results align closely with my main findings, confirming that facility-specific effects drive emissions discrepancies. These results underscore the robustness of the primary analysis and highlight the consistency of the insights across different modeling approaches.

Elastic Net regression, optimized with an alpha of 0.00047 and an l1_ratio of 1.0 (indicating a purely LASSO approach), further supported these findings. The model achieved a Train R^2 of 0.7715 and a Test R^2 of 0.7237, corresponding Mean Squared Errors (MSE) of 0.3822 and 0.4517, respectively. Cross-validation results indicated stability, with a mean R^2 of 0.6909 and a standard deviation of 0.0420, see Table 15.

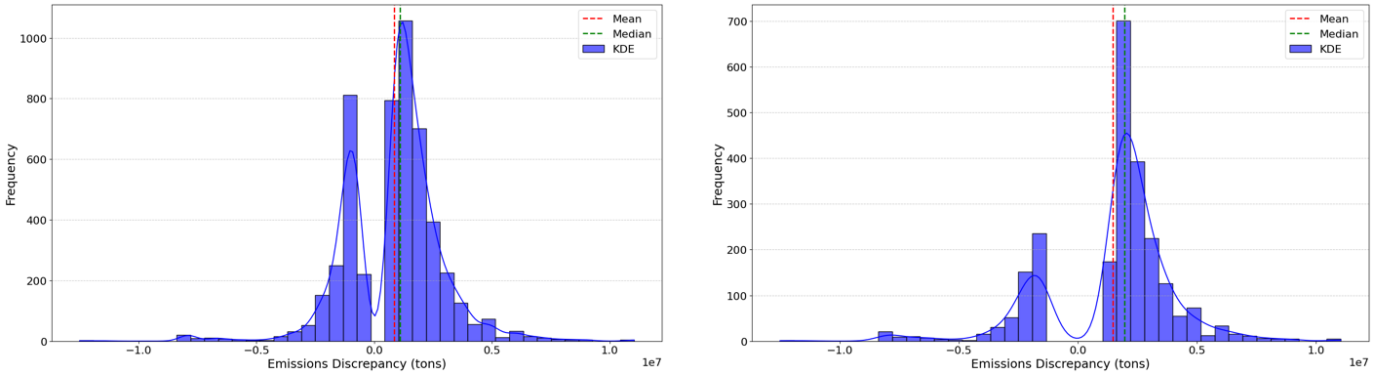


Figure 28: Distributions of emissions discrepancies with 1 (left) and 2 (right) SDs excluded

Source: Own illustration

Table 13: Random Forest models, original and minimal discrepancies removed

	Full data	±1 Std Dev removed	±2 Std Dev removed
Train R^2	0.8926	0.9405	0.9437
Test R^2	0.8677	0.9117	0.9118
CV Mean R^2	0.8410	0.8766	0.8527
Mean Std Dev R^2	0.0419	0.0380	0.0663
Contributions to explained variance:			
GHGRP ID	40.8%	39.6%	35.6%
County Name	24.2%	31.0%	33.0%
Parent Companies	22.6%	22.3%	24.6%
Subparts	10.3%	4.4%	3.9%
Year	2.1%	2.6%	2.9%
Emissions tons	0.0%	0.0%	0.0%

Source: Own illustration

Table 14: ANOVA results

Predictor	Sum of Squares	df	F	p-value	Partial Eta Squared
GHGRP ID	181,347,700	3483	350,239.38	< 0.001	0.846
County Name	1,864,718	985	12,734.54	< 0.001	0.009
Parent Companies	31,165,200	2250	93,173.79	< 0.001	0.145
Subparts	1.88	36	0.35	1.000	< 0.001
Year	12.22	7	11.74	< 0.001	< 0.001
Emissions tons	66.39	1	446.57	< 0.001	< 0.001

Source: Own illustration

Table 15: Elastic Net model results

Predictor	Total Contribution	Percentage (%)	Significant Predictors	Significance (%)
GHGRP ID	214.21	70.83	68	1.92
County Name	46.83	15.48	33	3.35
Parent Companies	39.23	12.97	47	2.05
Subparts	1.82	0.60	9	25.00
Year	0.01	0.003	1	100.00
Emissions tons	0.00	0.00	0	0.00

Source: Own illustration

Regarding predictor contributions, GHGRP ID again emerged as the dominant factor, accounting for 70.83% of the explained variance. Parent Companies contributed 12.97%, and County Name added 15.48%. Subparts and Year contributed minimally, explaining 0.60% and 0.34%, respectively, while Emissions Tons again had no significant impact.

The Elastic Net results confirm the dominance of facility-specific effects while providing additional nuance to the roles of shared factors. Although the contribution of County Name (15.48%) was slightly higher compared to the Random Forest results of my primary analysis, the overall pattern remained consistent: emissions discrepancies are primarily determined by facility-specific characteristics, with geographic and organizational factors playing supporting roles.

ANOVA and Elastic Net results align closely with the Random Forest and Bayesian models, consistently showing that facility-specific effects (GHGRP ID) dominate the explained variance. At the same time, Parent Companies, County Name and Subparts exert a weaker influence. The negligible impact of Year and Emissions Tons is also consistent across models.

The two linear models confirm the robustness of the more complex machine learning models, reinforcing the reliability of my conclusions. These additional analyses confirm that the main findings are not model-dependent but reflect authentic patterns in the data, further supporting their generalizability.

5 Discussion

In this study, I examined discrepancies between self-reported emissions from the Greenhouse Gas Reporting Program (GHGRP) and satellite-based estimates from Climate Trace (CT), identifying key predictors of these disparities. Understanding these differences is critical for improving emissions accounting accuracy, informing climate policy, and advancing hybrid verification frameworks integrating self-reported and independent monitoring approaches.

My analysis showed that facility-specific factors are the primary driver of discrepancies between GHGRP self-reported emissions and CT satellite-based estimates. County-level geographic factors and corporate (parent company) influences also contribute, but to a lesser extent. Notably, industry-wide effects are negligible, suggesting discrepancies stem from individual facility-level reporting practices rather than sector-wide methodological differences. Furthermore, the facility's total emissions volume and the report's year do not influence GHG discrepancies, which remained relatively stable over time. The sustained increase in emissions discrepancies from 2018 to 2020 suggests an influence of CT dataset expansion, possibly due to satellite monitoring and data coverage improvements.

These findings align with existing literature on GHG emissions; for example, they are consistent with Gurney et al. (2024), who found that facility-level biases significantly influence reported emissions. They also support research by Subramanian et al. (2015), which showed self-reported inventories underestimating fugitive emissions, particularly in oil and gas. My results align with those of Cusworth et al. (2021), who found no strong industry-wide effects, emphasizing reporting variability at the facility level. This is similarly supported by other research suggesting high facility-level variability in methane emissions (Papadopoulos, 2022).

However, my results also contradict some existing research. They contrast Nesser et al. (2024), who found systematic underreporting in landfill methane emissions, whereas this study finds both positive and negative discrepancies. While my study and Dogniaux et al. (2024) identify discrepancies between satellite-based methane estimates and facility-reported emissions, my study finds that CT's model-based approach introduces additional uncertainties, mainly due to extreme outliers and facility-specific factors. In contrast, Dogniaux et al. (2024) found no significant correlation between GHGSat satellite observations and reported emissions at all.

My analysis could not directly answer why facility-level factors dominate the observed discrepancies between reported GHG emissions from the two data sources. However, based on the reviewed literature and analysis results, the dominance of facility-level factors in explaining emissions discrepancies might stem from variability in emissions estimation methodologies, operational conditions, and regulatory flexibility. Facilities employ different calculation approaches, ranging from direct measurement (e.g., Continuous Emissions Monitoring Systems) to model-based estimation using default emissions factors, which inherently introduces variability. This methodological flexibility is a key limitation of carbon accounting systems like GHGRP, as facilities often select the approach that minimizes reported emissions, leading to systematic biases (Stark et al., 2024; Subramanian et al., 2015). Beyond methodology, operational conditions - such as equipment maintenance, process efficiency, and calibration practices - may also impact emissions measurement. For example, landfill methane emissions are significantly underestimated under "recovery-first" models, which assume optimistic gas capture rates despite empirical evidence suggesting substantial fugitive emissions(Nesser et al., 2024).

Additionally, local environmental factors, such as temperature and humidity, can influence emissions variability, particularly for landfills and petrochemicals, where atmospheric conditions affect methane oxidation and volatile organic compound emissions (Jain et al., 2021). This could explain my finding that a facility's geographic location (county) does influence the observed discrepancies. While corporate governance and geographic factors also contribute, their effects are secondary to facility-level variability. I suspect parent company policies and sustainability strategies shaping reporting transparency, as some firms engage in more rigorous voluntary disclosures than others. Similarly, county-level regulatory enforcement and state policies - although GHGRP is federally mandated - might indirectly influence reporting behaviors.

In contrast, industry-wide effects are minimal, suggesting that standardized sectoral methodologies do not eliminate intra-sectoral variability. While carbon accounting frameworks such as GHGRP provide industry-specific reporting guidelines, facilities might implement them differently, leading to inconsistent reporting practices within the same sector. As mentioned, this is supported by research indicating that discrepancies in reported emissions are primarily facility-driven rather than sector-wide trends (Cusworth et al., 2021; Gurney et al., 2024).

My findings underscore the need for hybrid verification systems, integrating bottom-up facility reporting with top-down satellite-based estimates, to reduce reliance on self-reported emissions data and enhance the accuracy of carbon accounting frameworks (Chan et al., 2024).

This study is not without limitations. First, I could only match 28.96% of GHGRP facilities to CT, meaning most discrepancies remain unaccounted for. Expanding the spatial matching threshold showed increased coverage but may introduce false positives. Thus, a more refined matching algorithm could increase the coverage and expand to the newest datasets from GHGRP and CT, including emission inventories for 2023, which were released after I finished my analysis. Second, as my study focuses solely on U.S.-based facilities, its generalizability to other countries using different reporting standards is limited. These might exhibit different patterns of discrepancies between their self-reporting and satellite-based emission data. Third, as the industry classifications between the two data sources do not perfectly align, I needed to introduce a manual mapping between them - see Table 1. I based this new categorization only on the existing category names, potentially introducing mismatches in their methodologies. This is especially relevant since earlier studies have found systematic underreporting in specific sectors (Nesser et al., 2024).

My study offers broad implications for the scientific literature, policy regulations, and practical applications. First, I contribute to scientific discussions on the need for hybrid verification approaches (Chan et al., 2024; Dogniaux et al., 2024; Goetz et al., 2009; National Academies of Sciences, Engineering, and Medicine, 2022) and show the potential of combining self-reported inventories with independent satellite-based monitoring. Second, my findings indicate that regulators should prioritize facility-level audits rather than industry-wide assessments, as discrepancies are facility-driven. Integrating third-party verification into programs like GHGRP, using satellite-based sensing approaches like CT, could enhance real-time detection of emissions events and improve data accuracy (Balasus et al., 2025; Cusworth et al., 2021; Duren et al., 2019). Finally, facility operators should use insights from third-party programs such as CT to refine their emissions monitoring practice and validate their data when creating reports, potentially lowering their risk of inaccurate reporting.

Building on the findings of my study, future research should expand and refine methodologies for assessing discrepancies between self-reported and satellite-based emissions data. A significant limitation of this study was the relatively low facility-matching rate (28.96%). Future research should refine matching algorithms to improve coverage while maintaining accuracy. This could involve integrating more detailed facility metadata, including geospatial attributes like temperature, humidity, and precipitation, and extending GHGRP and CT to include their newest inventories. Future research should additionally expand the comparison between CT and self-reported emissions to include other regulatory frameworks, such as the European Union Emissions Trading System (EU ETS), which covers over 11,000 power plants and industrial facilities (European Commission, 2025), Brazil's SEEG Inventory, which provides sectoral and facility-level emissions estimates (SEEG Community, 2025) and other national emissions reporting frameworks, such as China's Corporate GHG Inventory (UNFCCC, 2024) and Canada's Greenhouse Gas Reporting Program (Government of Canada, 2025). Likewise, comparing CT's emission data to other remote sensing systems, such as GHGSat (HGSat Inc., 2025) and NASA's OCO-2 and OCO-3 satellites (NASA, 2025a), could provide deeper insights into the reliability and consistency of satellite-based emissions monitoring. Beyond this, a thorough investigation into why specific sectors, namely aviation and road transport, exhibited far higher relative discrepancies in my analysis could result in a better understanding of potential weaknesses in CT's estimation algorithms. Finally, another promising avenue for future research is linking emissions discrepancies to corporate financial data. This would enable an exploration of whether financial incentives or governance structures influence emissions reporting accuracy. Specifically, this might include financial variables of facility-matched parent companies such as revenue, profitability, R&D investment, and ESG ratings. By expanding the analytical scope in these directions, future research can enhance the robustness of emissions verification methodologies, inform climate policy, and support the development of more reliable hybrid emissions monitoring frameworks.

6 Conclusion

With this study, I provide the first large-scale facility-level comparison of GHGRP self-reported emissions and CT satellite-based estimates. My analysis shows that facility-specific factors dominate emissions discrepancies, with geographic (county) and corporate (parent company) effects playing secondary roles. Industry-wide differences were minimal and remained stable over time, suggesting inconsistencies arise from individual reporting practices rather than systemic methodological flaws.

My findings underscore the limitations of self-reported emissions inventories and highlight the need for independent satellite-based verification to improve emissions monitoring. The results support hybrid verification systems integrating self-reported and remote-sensing data to improve accountability. Policymakers and facility operators should prioritize

facility-level audits and verification mechanisms leveraging satellite-based insight to refine reporting accuracy and compliance strategies.

Future research should expand on my facility-matching methodology to improve dataset coverage, compare CT to other emissions inventories outside the US, assess discrepancies using additional remote sensing datasets, and investigate financial drivers of reporting discrepancies by linking emissions data with corporate financial performance.

Strengthening independent emissions verification is crucial for ensuring the accuracy of emissions accounting and advancing effective climate policies. My study highlights the need for integrated monitoring approaches, leveraging self-reported and satellite-based data to improve transparency, accountability, and climate action.

References

Alfieri, A., Holloway, L., von Kalckreuth, U., Moll, S., Schmieder, C., & Tissot, B. (2024). Empowering Carbon Accounting: From Data to Action. *Bank for International Settlements*. https://ideas.repec.org/p/bis/bisifr/16.html

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation Importance: A Corrected Feature Importance Measure. *Bioinformatics*, 1340–1347.

Ascui, F. (2014). A Review of Carbon Accounting in the Social and Environmental Accounting Literature: What Can It Contribute to the Debate? *Social and Environmental Accountability Journal*, 6–28.

Ascui, F., & Lovell, H. (2011). As Frames Collide: Making Sense of Carbon Accounting. *Accounting, Auditing and Accountability Journal*, 978–999.

Balasus, N., Jacob, D. J., Maxemina, G., Jenks, C., Nesser, H., Maasakkers, J. D., Cusworth, D. H., Scarpelli, T. R., Varon, D. J., & Wang, X. (2025). Satellite Monitoring of Annual US Landfill Methane Emissions and Trends. https://acmg.seas.harvard.edu/sites/projects.iq.harvard.edu/files/balasus-landfill-2024.pdf

Bauckloh, T., Klein, C., Pioch, T., & Schiemann, F. (2022). Under Pressure? The Link Between Mandatory Climate Reporting and Firms' Carbon Performance. *Organization & Environment*, 126–149.

Bobbitt, Z. (2020). Statology. https://www.statology.org/breusch-pagan-test/

Breiman, L. (2001). Random Forests. *Machine Learning*, 5–32.

Bronstein, K., Coburn, J., & Schmeltz, R. (2012). Understanding the Inventory of U.S. Greenhouse Gas Emissions and Sinks and the Greenhouse Gas Reporting Program for Landfills: Methodologies, Uncertainties, Improvements, and Deferrals. *International Emission Inventory Conference*.

Busch, T., Bassen, A., Lewandowski, S., & Sump, F. (2022). Corporate Carbon and Financial Performance Revisited. *Organization & Environment*.

Chan, E., Vogel, F., Smyth, S., Barrigar, O., Ishizawa, M., Kim, J., Neish, M., Chan, D., & Worthy, D. (2024). Hybrid Bottom-Up and Top-Down Framework Resolves Discrepancies in Canada's Oil and Gas Methane Inventories. *Nature*.

Chapple, L., Clarkson, P. M., & Gold, D. L. (2013). The Cost of Carbon: Capital Market Effects of the Proposed Emissions Trading Scheme (ETS). *Abacus*.

Chen, J., Zhao, F., Zeng, N., & Oda, T. (2020). Comparing a Global High-Resolution Downscaled Fossil Fuel CO2 Emission Dataset to Local Inventory-Based Estimates Over 14 Global Cities. *Carbon Balance and Management*.

Climate TRACE. (2024a). Climate TRACE. https://climatetrace.org/data

Climate TRACE. (2024b). Climate TRACE. https://climatetrace.org/explore

Climate TRACE. (2025a). Climate TRACE. https://climatetrace.org/about

Climate TRACE. (2025b). Climate TRACE. https://climatetrace.org/sectors

Couture, H., Alvara, M., Freeman, J., Davitt, A., Koenig, H., Kargar, A., O'Connor, J., Söldner-Rembold, I., Ferreira, A., Jeyaratam, J., Lewis, J., McCormick, C., Nakano, T., Dalisay, C., & Lewi. (2024). Estimating Carbon Dioxide Emissions from Power Plant Water Vapor Plumes Using Satellite Imagery and Machine Learning. *Remote Sensing*.

Csutora, M., & Harangozo, G. (2017). Twenty Years of Carbon Accounting and Auditing – A Review and Outlook. *Society and Economy*.

Cusworth, D. H., Duren, R. M., Thorpe, A. K., Olson-Duvall, W., Heckler, J., Chapman, J. W., Eastwood, M. L., Helmlinger, M. C., Green, R. O., Asner, G. P., Dennison, P. E., & Miller, C. E. (2021). Intermittency of Large Methane Emitters in the Permian Basin. *Environmental Science & Technology Letters*.

da Costa, L. M., Davitt, A., Volpato, G., Costa de Mendonça, G., Panosso, A. R., & La Scala Jr, N. (2025). A Comparative Analysis of GHG Inventories and Ecosystems Carbon Absorption in Brazil. *Science of The Total Environment*.

Deutsches Zentrum für Luft- und Raumfahrt e.V. (2024). CO2M - Copernicus Anthropogenic Carbon Dioxide Monitoring. https://www.d-copernicus.de/daten/satelliten/satelliten-details/news/co2m-copernicus-anthropogenic-carbon-dioxide-monitoring/

Dogniaux, M., Maasakkers, J., Girard, M., Jervis, D., McKeever, J., Schuit, B., Sharma, S., Lopez-Noreña, A., Varon, D., & Aben, I. (2024). Satellite Survey Sheds New Light on Global Solid Waste Methane Emissions. *EarthArXiv*.

Duren, R. M., Thorpe, A. K., Foster, K. T., Rafiq, T., Hopkins, F. M., Yadav, V., Bue, B. D., Thompson, D. R., Conley, S., Colombi, N. K., Frankenberg, C., McCubbin, I. B., & Eastwood, M. L. (2019). California's Methane Super-Emitters. *Nature*.

Environmental Protection Agency. (2009). Mandatory Reporting of Greenhouse Gases; Final Rule [Federal Register, 74(209)].

Environmental Protection Agency. (2017). EPA. https://www.epa.gov/sites/default/files/2017-12/documents/ghgrp_methodology_factsheet.pdf

Environmental Protection Agency. (2024a). Basic Information of Air Emissions Factors and Quantification. https://www.epa.gov/air-emissions-factors-and-quantification/basic-information-air-emissions-factors-and-quantification

Environmental Protection Agency. (2024b). EPA. https://www.epa.gov/ghgreporting/what-ghgrp

Environmental Protection Agency. (2024c). EPA. https://www.epa.gov/ghgreporting/ghgrp-newsletters

Environmental Protection Agency. (2024d). EPA FLIGHT. https://ghgdata.epa.gov/ghgp/main.do

Environmental Protection Agency. (2024e). GHGRP Overview Fact Sheet. https://www.epa.gov/system/files/documents/2023-08/ghgrp-overview-fact-sheet.pdf

Environmental Protection Agency. (2024f). Revisions and Confidentiality Determinations for Data Elements Under the Greenhouse Gas Reporting Rule. https://www.epa.gov/system/files/documents/2024-04/ghgrp-final-preamble-and-rule-april-2024.pdf

Environmental Protection Agency. (2025a). Inventory of U.S. Greenhouse Gas Emissions and Sinks. https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks

Environmental Protection Agency. (2025b). Overview of Greenhouse Gases. https://www.epa.gov/ghgemissions/overview-greenhouse-gases

European Commission. (2024). GHG Emissions of All World Countries. https://edgar.jrc.ec.europa.eu/report_2024

European Commission. (2025). EU Emissions Trading System (EU ETS). https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets_en

European Commission, Joint Research Centre. (2023). JRC SCIENCE FOR POLICY REPORT GHG EMISSIONS OF ALL WORLD COUNTRIES. https://edgar.jrc.ec.europa.eu/booklet/GHG_emissions_of_all_world_countries_booklet_2023report.pdf

Fawzy, S., Osman, A. I., Doran, W. J., & Rooney, D. (2020). Strategies for Mitigation of Climate Change: A Review. *Environmental Chemistry Letters*, 2069–2094.

Godec, M. L., de Figueiredo, M. A., & Miller, M. L. (2022). Assessment of the Costs Associated with Monitoring, Reporting and Verification for Greenhouse Gas Reporting of Geologic Sequestration of Carbon Dioxide. *16th International Conference on Greenhouse Gas Control Technologies*.

Goetz, S. J., Baccini, A., Laporte, N. T., Johns, T., Walker, W., Kellndorfer, J., Houghton, R. A., & Sun, M. (2009). Mapping and Monitoring Carbon Stocks with Satellite Observations: A Comparison of Methods. *Carbon Balance and Management*.

González-Estrada, E. (2019). Shapiro–Wilk Test for Skew Normal Distributions Based on Data Transformations. *Journal of Statistical Computation and Simulation*, 3258–3272.

Government of Canada. (2025). Greenhouse Gas Reporting Program (GHGRP) - Facility Greenhouse Gas (GHG) Data. https://open.canada.ca/data/en/dataset/a8ba14b7-7f23-462a-bdbb-83b0ef629823

Gurney, K. R., Aslam, B., Dass, P., Gawuc, L., Hocking, T., Barber, J. J., & Kato, A. (2024). Assessment of the Climate Trace Global Powerplant CO2 Emissions. *Environmental Research Letters*, *19*(11).

Gurney, K. R., Liang, J., Roest, G., Song, Y., Mueller, K., & Lauvaux, T. (2021). Under-Reporting of Greenhouse Gas Emissions in U.S. Cities. *Nature Communications*.

Hanle, L., Kocchi, S., & Ragnauth, S. (2012). Developing a Facility-Level Reporting Programme for Sources of Non-CO2 Gases: The United States Experience. *Greenhouse Gas Measurement and Management*, 33–42.

Hans, C. (2012). Elastic Net Regression Modeling With the Orthant Normal Prior. *Journal of the American Statistical Association*, 1383–1393.

Hazaea, S., Al-Matari, E. M., Alosaimi, M., Farhan, N., Abubakar, A., & Zhu, J. (2023). Past, Present, and Future of Carbon Accounting: Insights from Scholarly Research. *Frontiers in Energy Research*, 1.

He, J., & Huang, B. (2023). Direct Estimation of Global Anthropogenic CO2 Emissions Using Satellite Data. *EGUsphere*.

He, M., Ditto, J., Gardner, L., Machechsky, J., Hass-Mitchell, T., Chen, C., Khare, P., & Sahin, B. (2024). Total Organic Carbon Measurements Reveal Major Gaps in Petrochemical Emissions Reporting. *Science*, 426–432.

Helleckes, L. M., Osthege, M., Wiechert, W., von Lieres, E., & Oldiges, M. (2022). Bayesian Calibration, Process Modeling and Uncertainty Quantification in Biotechnology. *Computational Biology*.

HGSat Inc. (2025). GHGSat. https://www.ghgsat.com/en/

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 1593–1623.

Huang, C., Patsika, V., Triantafylli, A., & Zhang, Y. (2023). Mandatory Greenhouse Gas Emissions Reporting and Firm Environmental Litigation Risk. *Accounting Forum*, 249–277.

Huertas, J. I., Huertas, M. E., & Díaz, J. (2012). Assessing Precision and Accuracy of Atmospheric Emission Inventories. *International Journal of Environmental Science and Technology*, 195–202.

IEA. (2024). CO2 Emissions in 2023. https://iea.blob.core.windows.net/assets/33e2badc-b839-4c18-84ce-f6387b3c008f/CO2Emissionsin2023.pdf

IPCC. (2023). CLIMATE CHANGE 2023 Synthesis Report Summary for Policymakers. https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_SPM.pdf

Jain, P., Wally, J., Townsend, T., Krause, M., & Tolaymat, T. (2021). Greenhouse Gas Reporting Data Improves Understanding of Regional Climate Impact on Landfill Methane Production and Collection. *PLoS ONE*.

Kaur, R., Patsavellas, J., Haddad, Y., & Salonitis, K. (2023). The Concept of Carbon Accounting in Manufacturing Systems and Supply Chains. *Energies*.

Kennelly, C., Berners-Lee, M., & Hewitt, C. (2019). Hybrid Life-Cycle Assessment for Robust, Best-Practice Carbon Accounting. *Journal of Cleaner Production*, 35–43.

Kern, C., Klausch, T., & Kreuter, F. (2019). Tree-Based Machine Learning Methods for Survey Research. *Surv Res Methods*, 73–93.

Kiemle, C., Ehret, G., Amediek, A., Fix, A., Quatrevalet, M., & Wirth, M. (2017). Potential of Spaceborne Lidar Measurements of Carbon Dioxide and Methane Emissions from Strong Point Sources. *Remote Sensing*.

Klaaßen, L., & Stoll, C. (2021). Harmonizing Corporate Carbon Footprints. *Nature Communications*.

Kruger, J. (2005). From SO2 to Greenhouse Gases: Trends and Events Shaping Future Emissions Trading Programs in the United States. http://www.rff.org/

Lee, S., & Bi, X. (2020). Can Embedded Knowledge in Pollution Prevention Techniques Reduce Greenhouse Gas Emissions? A Case of the Power Generating Industry in the United States. *Environmental Research Letters*.

Lewandowski, S. (2017). Corporate Carbon and Financial Performance: The Role of Emission Reductions. *Business Strategy and the Environment*.

Li, B., Sun, Z., He, Q., Zhu, Y., & Qin, Z. (2016). Bayesian Inference with Historical Data-Based Informative Priors Improves Detection of Differentially Expressed Genes. *Bioinformatics*, 682–689.

Marlowe, J., & Clarke, A. (2022). Carbon Accounting: A Systematic Literature Review and Directions for Future Research. *Green Finance*.

Miller, M., Clark, T., Hadden, O., & Konovalchik, N. (2024). United States Greenhouse Gas Reporting Framework for CCS. *17th International Conference on Greenhouse Gas Control Technologies, GHGT-17*.

Moran, D., & Wood, R. (2014). Convergence Between the Eora, WIOD, EXIOBASE, and OpenEU's Consumption-Based Carbon Accounts. *Economic Systems Research*.

NASA. (2025a). Landsat 8. https://landsat.gsfc.nasa.gov/satellites/landsat-8/

NASA. (2025b). Orbiting Carbon Observatory-2. https://ocov2.jpl.nasa.gov/

National Academies of Sciences, Engineering, and Medicine. (2022). *Greenhouse Gas Emissions Information for Decision Making: A Framework Going Forward*. The National Academies Press.

Nesser, H., Jacob, D. J., Maasakkers, J. D., Lorente, A., Chen, Z., Lu, X., Shen, L., Qu, Z., Sulprizio, M. P., Winter, M., Ma, S., Bloom, A. A., Worden, J. R., Stavins, R. N., & Randles, C. A. (2024). High-Resolution US Methane Emissions Inferred from an Inversion of 2019 TROPOMI Satellite Data: Contributions from Individual States, Urban Areas, and Landfills. *Atmospheric Chemistry and Physics*.

Nguyen, H. L., Chau, L., Hiep, N. L., & Nguyen, D. T. (2025). Does Mandatory Greenhouse Gas Emissions Reporting Program Deter Corporate Greenwashing? *Journal of Environmental Management*.

Ofer, D. (2018). Kaggle. https://www.kaggle.com/datasets/danofer/zipcodes-county-fips-crosswalk

Papadopoulos, G. (2022). Discrepancies in Corporate GHG Emissions Data and Their Impact on Firm Performance Assessment. *SSRN Electronic Journal*.

Predictive Species and Habitat Modeling in Landscape Ecology. (2011). In *Predictive Species and Habitat Modeling in Landscape Ecology* (pp. 140–144). Springer.

Provaty, S., Hasan, M., & Luo, L. (2024). Organization Capital and GHG Emissions. *Energy Economics*.

PyPi. (2024). PyPi Haversine. https://pypi.org/project/haversine/

rapidfuzz. (2025). RapidFuzz 3.11.0 Documentation. https://rapidfuzz.github.io/RapidFuzz/Usage/fuzz.html

Rogelj, J., Huppmann, D., Krey, V., Riahi, K., Clarke, L., Gidden, M., Nicholls, Z., & Meinshausen, M. (2019). A New Scenario Logic for the Paris Agreement Long-Term Temperature Goal. *Nature*.

Roy, V., & Larocque, D. (2012). Robustness of Random Forests for Regression. *Journal of Nonparametric Statistics*, 993–1006.

Sanchez, M., Matthews, S., & Fischbeck, P. (2012). How Much Is United States Greenhouse Gas Emissions Certainty Worth? *Energy Policy*, 259–263.

Schaltegger, S., & Csutora, M. (2012). Carbon Accounting for Sustainability and Management. Status Quo and Challenges. *Journal of Cleaner Production*, 1–16.

Schrock, D., Stoops, J., Meier, A., Vine, E., & Solomon, B. (1994). Developmental Issues in Environmental Reporting Protocols. *American Council for Energy-Efficient Economy 1994 Summer Study on Energy Efficiency in Buildings*.

scikit-learn developers. (2025). RandomizedSearchCV. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html

SciPy. (2024). cKDTree. https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.cKDTree.html

SEEG Community. (2025). SEEG. https://seeg.eco.br/en/home/

Sitompul, M., Suroso, A. I., Sumarwan, U., & Zulbainarni, N. (2023). Revisiting the Impact of Corporate Carbon Management Strategies on Corporate Financial Performance: A Systematic Literature Review. *Economies*.

Ståhle, L., & Wold, S. (1989). Analysis of Variance (ANOVA). *Chemometrics and Intelligent Laboratory Systems*, 259–272.

Stark, B. M., Tian, K., & Krause, M. J. (2024). Investigation of U.S. Landfill GHG Reporting Program Methane Emission Models. *Waste Management*, 86–93.

Statext LLC. (2024). Normality Test. https://www.statext.com/practice/NormalityTest04.php

Stechemesser, K., & Guenther, E. (2012). Carbon Accounting: A Systematic Literature Review. *Journal of Cleaner Production*, 17–38.

Subramanian, R., Williams, L. L., Vaughn, T. L., Zimmerle, D., Roscioli, J. R., Herndon, S. C., Yacovitch, T. I., Floerchinger, C., Tkacik, D. S., Mitchell, A. L., Sullivan, M. R., Dallmann, T. R., & Robinson, A. L. (2015). Methane Emissions from Natural Gas Compressor Sta-

tions in the Transmission and Storage Sector: Measurements and Comparisons with the EPA Greenhouse Gas Reporting Program Protocol. *Environmental Science & Technology*, 3252–3261.

Tang, Q. (2017). Framework for and the Role of Carbon Accounting in Corporate Carbon Management Systems: A Holistic Approach. *SSRN Electronic Journal*.

Tang, Q. (2024). Managing the Shift from Voluntary to Mandatory Climate Reporting: The Role of Carbon Accounting. *SSRN*.

The European Space Agency. (2025). PlanetScope. https://earth.esa.int/eogateway/missions/planetscope

Tuesta, Y., Soler, C. C., & Feliu, V. R. (2020). Carbon Management Accounting and Financial Performance: Evidence from the European Union Emission Trading System. *Business Strategy and the Environment*.

UNFCCC. (2015). Paris Agreement. https://unfccc.int/sites/default/files/english_paris_agreement.pdf

UNFCCC. (2024). The People's Republic of China Fourth Biennial Update Report on Climate Change. https://unfccc.int/sites/default/files/resource/The%5C%20People%5C%E2%5C%80%5C%99s%5C%20Republic%5C%20of%5C%20China%5C%20Fourth%5C%20Biennial%5C%20Update%5C%20Report%5C%20on%5C%20Climate%5C%20Change.pdf

United Nations. (1998). Kyoto Protocol to the United Nations Framework Convention on Climate Change. https://unfccc.int/resource/docs/convkp/kpeng.pdf

U.S. National Archives and Records Administration. (2025). Code of Federal Regulations, Title 40, Chapter I, Subchapter C, Part 98, Subpart A. https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-98/subpart-A

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bärkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC. *Bayesian Analysis*.

Vogel, R. M., Papalexiou, S. M., Lamontagne, J. R., & Dolan, F. (2024). When Heavy Tails Disrupt Statistical Inference. *The American Statistician*.

Wang, X., Chen, L. W., Lu, M., Ho, K.-F., Lee, S.-C., Ho, S. S. H., Chow, J., & Watson, J. G. (2022). Apportionment of Vehicle Fleet Emissions by Linear Regression, Positive Matrix Factorization, and Emission Modeling. *Atmosphere*.

World Resources Institute. (2024). Greenhouse Gas Protocol. https://www.itf-oecd.org/sites/default/files/1.2._ghg_protocol.pdf

Yang, L., Muller, N. Z., & Liang, P. J. (2024). The Real Effects of Mandatory CSR Disclosure on Emissions: Evidence from the Greenhouse Gas Reporting Program. *National Bureau of Economic Research (NBER)*. https://www.nber.org/system/files/working_papers/w28984/w28984.pdf

Zhou, Y., Zhang, J., & Hu, S. (2021). Regression Analysis and Driving Force Model Building of CO2 Emissions in China. *Scientific Reports*.

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Experimental Analysis of Gender Bias in Hiring within STEM Fields

Roya Kazimova

Technical University of Munich

Abstract

This study investigates the presence of gender bias in students' hiring evaluations within STEM fields using a survey-based experimental design with systematically varied resumes. Students were randomly assigned to evaluate one of four fictional CVs, differing by candidate gender and qualification strength. Participants rated the applicants on competence, hireability, salary recommendation, and willingness to offer professional support. The results reveal no consistent bias in favor of male candidates; however, strong male applicants received higher salary suggestions, while female candidates were more likely to be offered professional support. These patterns are primarily driven by interactions between candidate gender and resume strength. The effects are most pronounced among participants from Eastern Europe and older participants, suggesting a role of social background in shaping perceptions. The findings highlight how implicit biases may form prior to labor market entry and underscore the need for early educational interventions to promote equitable hiring perceptions in STEM.

Keywords: gender bias; hiring discrimination; survey experiment; resume evaluation

1 Introduction

In recent years, there has been growing attention to diversity and inclusion in the workplace, especially in areas where women are still underrepresented. Many companies, universities, and policymakers are trying to better understand and reduce hidden biases that can affect hiring and promotion decisions (Wilkinson & Barry, 2020). These efforts are especially important in STEM, where gender gaps remain despite efforts to improve equality (Evagorou et al., 2024). One of the key questions in labor economics and gender studies is whether gender bias continues to influence hiring decisions in these male-dominated sectors. These biases not only limit women's access to employment but also affect long-term career growth, salary progression, and leadership opportunities.[1] According to the Gender Social Norms Index by United Nations Development Programme (2023), 85% of the global population has biased views against women, and more than half believes that men are better suited for leadership positions.

While there is strong evidence that gender bias affects real-world hiring decisions, most existing research has focused on the behavior of employers or experienced professionals. For example, Quadlin (2018) conducted a large-scale field experiment and found that women with high academic performance were rated as competent but less likable and hirable than men with moderate academic records. This suggests that even when qualifications are equal or superior, gender stereotypes can significantly shape hiring outcomes. However, much less is known about how university students, who represent the next generation of hiring professionals, perceive and internalize such biases. This gap is particularly relevant in STEM contexts, where common stereotypes (e.g., "math is a male domain") shape both other people's opinions and personal confidence (Correll, 2001).

This study investigates how university students with diverse academic and professional backgrounds perceive gender bias in candidate evaluation. By focusing on students in interdisciplinary programs combining technical and managerial studies, the research captures perspectives that are highly relevant to future decision-making in both STEM and organizational contexts. Most participants have prior work experience, adding practical insight to their evaluations. The study uses a survey-based experimental design in which participants assess fictional job applicants, allowing a systematic investigation of how gender might affect candidate evaluation. This approach shows how students think about gender bias before they enter professional roles and start making hiring decisions.

Preliminary findings from the survey experiment reveal a more complex picture than anticipated. Contrary to expectations, male candidates were not consistently favored and were in some cases rated lower on hireability compared to equally qualified female applicants. While direct gender effects were limited, interaction patterns suggested that strong male candidates were more likely to receive higher salary rec-

[1] McKinsey & Company, *Women in the Workplace 2024: The 10th-Anniversary Report*, 2024, https://www.mckinsey.com/featured-insights/diversity-and-inclusion/women-in-the-workplace.

DOI: https://doi.org/10.5282/jums/v11i3pp598-610

© The Author(s) 2026. Published by Junior Management Science.
This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.

ommendations but were less likely to receive offers of professional support. Additionally, participants' demographic characteristics, particularly age and region of origin, significantly influenced candidate evaluations. These results point to a nuanced understanding of bias, shaped by social norms, cultural background, and prior work experience.

This study contributes to existing research in three ways. First, it examines whether university students recognize gender biases in hiring, filling a gap in research focused on future professionals rather than current employees. Second, it analyzes the role of demographic factors such as gender, field of study, and work experience in shaping students' hiring evaluations. Third, it explores whether students' perceptions of gender bias translate into biased hiring decisions in a controlled experimental setting. By identifying the extent to which students recognize hiring biases and how these biases affect candidate evaluations, this research contributes to developing educational interventions and hiring policies that promote gender equality in STEM fields. Understanding how future professionals perceive and respond to gender biases can inform diversity initiatives, helping to create more equitable hiring practices in STEM and beyond.

The remainder of the paper is structured as follows. Section 2 presents the literature review on gender bias in hiring, experimental methodologies, and perception research. Section 3 describes the methodological approach, including the survey experiment design and data collection. Section 4 outlines the steps of the data analysis. Section 5 presents the empirical results. Finally, Section 6 offers a discussion of the findings and concludes the paper.

2 Literature Review

Research shows that gender bias still affects hiring decisions, even when men and women have the same qualifications. These biases are often subtle and may not be immediately apparent, but they can still influence how candidates are evaluated. In their famous study of orchestra auditions, Goldin and Rouse (2000) demonstrated that "blind" evaluation – when the screen hides the musician's identity – significantly increases the chance that women will be selected. The results of the study show how even minor cues related to gender can influence decision-making and how removing identity-related information can reduce bias. Bias is especially likely to appear in roles that are typically associated with one gender. Employers may not discriminate against women overall, but may be more hesitant to hire them into male-dominated positions, and vice versa for men. This pattern reflects the persistence of gender stereotypes in the labor market (Rich, 2014). When a woman enters a role in a male-dominated field, her competence may be judged not only by her resume but also by stereotypical perceptions of who is "naturally suited" for the role. Over time, even these small patterns of judgment can create unequal starting points in careers. Moreover, they can accumulate and lead to long-term negative consequences for women's careers, such as harsher evaluations, lower starting salaries, or fewer opportunities for advancement (Hardy III et al., 2022).

Despite growing attention to gender equality, women remain underrepresented in STEM fields, especially at higher academic and professional levels. Large-scale studies have shown that women in STEM often face slower career progression. Holman et al. (2018), for example, analyzed over 36 million scientific publications and found that women are less likely to be listed as lead authors. Since being a lead author is important for academic recognition and promotion, the lack of women in these positions shows how gender inequality limits career growth in STEM. Bias in hiring further reinforces inequality. STEM managers may apply different evaluation criteria depending on the candidate's gender. For instance, men in decision-making roles often emphasize long working hours when reviewing female applicants, while giving less weight to core skills. Even small cues, such as including a note about childcare support in a resume, can shift these perceptions and reduce bias (Friedmann & Efrat-Treister, 2023).

Students already form opinions about gender, fairness, and who is the most suitable for certain roles. These beliefs can also shape what they expect from future workplaces. If students believe that bias and discrimination are common, this can discourage them from applying to certain jobs or make them hesitant about entering male-dominated fields like STEM. Previous research has shown that even experienced professionals, such as university professors, tend to rate identical resumes more favorably when they have a male name (Moss-Racusin et al., 2012). If such bias exists among academics, it raises important questions about how early these patterns develop. Studies show that female students are more likely to see gender discrimination as a serious barrier to career success, while male students often believe it is less common or less important. These differences are influenced by personal experiences, education, and exposure to diversity (Sipe et al., 2009). Since today's students will become tomorrow's professionals and decision-makers, it is important to understand their views now.

Discrimination in hiring is often difficult to measure through surveys or self-reported data because people may not be fully aware of their biases or may not feel comfortable admitting them. For this reason, many researchers use experimental methods to study biased behavior in a more controlled and reliable way, like the resume evaluation method. In this design, researchers send out fictional job applications that are identical in qualifications but differ in one key aspect, such as gender or ethnicity. This makes it possible to see the effect of that variable on hiring decisions. Bertrand and Mullainathan (2004), for example, found that job applicants with white-sounding names received significantly more callbacks than those with black-sounding names, even though their resumes were otherwise the same. In another study, Milkman et al. (2015) applied the same method in academia and found that professors responded less often to requests from female and minority students.

When studying hiring discrimination, researchers commonly employ two main types of experimental designs: field

experiments and lab-based experiments. The former involve realistic scenarios conducted in natural settings, such as resume studies, where researchers submit fictitious resumes to real job openings. The latter, on the other hand, are conducted in controlled environments or online, allowing researchers to precisely manage variables and conditions. An important distinction exists between these two methodologies. Field experiments effectively capture real-world hiring decisions and are more likely to reveal actual biases because participants are unaware that they are being studied. In contrast, lab and online experiments provide researchers greater control over the experimental conditions but may not fully reflect real-life decisions, as participants typically face no genuine consequences for their choices. As a result, individuals in lab settings might provide answers they believe are socially desirable rather than truthful (Neumark, 2018). Despite this limitation, lab and online experiments remain valuable for understanding perceptions, attitudes, and cognitive processes within a carefully controlled environment.

The aforementioned studies have clarified how gender discrimination operates at different personal development stages, noting, for instance, that girls' interest in leadership declines during adolescence (Alan et al., 2019), and that women are less likely to negotiate salaries early in their careers (Babcock et al., 2003). These early patterns partially explain why fewer women reach leadership positions.[2] However, it remains unclear how early these tendencies form. Specifically, little is known about whether such biases are already present among students before entering the professional workforce. Addressing this gap is crucial, as understanding students' perceptions could help mitigate biases before they become deeply rooted.

3 Methodology

This chapter outlines the methodology used in this study, detailing the research design, survey structure, data collection process, and analytical approach. Section 3.1 presents the research design, explaining the rationale for using an experimental approach and describing the independent and dependent variables. Subsection 3.1.1 provides a detailed account of the experimental setup and procedure, including how candidate resumes were developed and structured. Section 3.2 describes the survey design, outlining the structure of the questionnaire and the key measures used to assess participants' evaluations and perceptions. Section 3.3 explains the data collection process, detailing the recruitment methods, survey distribution, ethical considerations, and response rates. This structure provides a clear and organized overview of the research process, forming the basis for analyzing gender bias in hiring.

3.1 Research Design

This study consists of two distinct methodological components: experimental survey design and descriptive survey analysis. The first part of the study employs a 2×2 between-subjects factorial experimental design to examine gender bias in hiring evaluations. Participants were randomly assigned to evaluate one of four candidate resumes, varying in gender (male vs. female) and qualification strength (strong vs. weak). This design ensured that each participant assessed only one candidate, preventing potential biases that could arise from direct comparisons. The design simulates a realistic hiring scenario, measuring hireability, competence, salary expectations, and professional support willingness. This method aligns with previous research on labor market discrimination and provides a structured approach to testing whether gender bias influences candidate evaluations. The study manipulates two independent variables: candidate gender and candidate qualification strength. The dependent variables include hireability (likelihood of hiring the candidate), competence (perceived ability and skills), and professional support (willingness to offer help in academic or professional settings). In addition, five demographic characteristics (participant age, gender, country of origin, field of study, and work experience) were included as control variables to account for their potential influence on candidate evaluations.

The second part of the study employs a descriptive survey analysis to examine participants' perceptions of gender discrimination in hiring, professional mentoring, and promotion opportunities. Unlike the experimental section, this part does not manipulate variables but instead presents a summary of responses, calculating mean discrimination perception scores across different participant groups (e.g., age, gender, country of origin, educational background, and work experience). Additionally, three control variables such as discrimination against women, men, and oneself, were derived from participants' responses. A total of 92 completed responses were included in the final analysis. The data were analyzed using *R*. Descriptive statistics were used to summarize the sample's demographic characteristics, reporting means, and standard deviations. Additionally, a series of OLS and OLR models with robust standard errors were performed to assess the influence of participant demographics and gender bias on candidate assessment scores. The survey experiment was structured to isolate the effects of candidate gender and qualifications while capturing nuanced differences in participant evaluations. Combined with demographic data, this approach allows for a multifaceted analysis of potential biases and perceived discrimination in STEM hiring contexts. The experimental procedures are described in Subsection 3.1.1, while the structure of the survey instrument and question formats is outlined in Section 3.2.

3.1.1 Experimental Setup and Procedure

Building on the structure outlined in the Research Design section, the experiment was designed to examine how

[2] S&P Global, *Women in Leadership: What's the Holdup?*, accessed March 22, 2025, https://www.spglobal.com/esg/insights/featured/special-editorial/women-in-leadership-what-s-the-holdup?utm_source=chatgpt.com.

participants evaluate job candidates and reflect on gender-related challenges in hiring. Each participant took on the role of a hiring manager and reviewed a resume on various aspects related to their suitability for the job. To keep the evaluation process realistic, participants were told that the resumes belonged to real applicants and that their feedback would help improve hiring decisions. After completing the evaluation, participants answered a post-experiment questionnaire focused on perceptions of workplace discrimination. The survey was conducted online via *Qualtrics*. The following section explains how the resumes were developed to systematically vary candidate gender and qualification strength. The template for the resumes was selected based on online research as one of the most popular templates in the job market. This approach replicates the methodology used by Bertrand and Mullainathan (2004), who constructed resumes based on real templates and systematically varied candidate qualifications to analyze hiring discrimination. Like their study, resume differentiation in this experiment was achieved through variations in academic performance, technical expertise, and certification levels.

The candidate profiles were presented in a standardized format, including consistent sections on education, skills, work experience, and certifications. The content emphasized a STEM background with a focus on electrical engineering, complemented by some coursework in business-related subjects. Each profile reflected a balance of theoretical knowledge and practical experience, typical of entry-level applicants in a technical-oriented field. To show differences in candidate qualifications, their GPAs were adjusted to reflect different levels of academic and analytical ability. Candidates with lower GPA scores were portrayed as having weaker theoretical skills, while those with higher GPAs emphasized stronger academic performance. *MATLAB*, an important tool for electrical engineers, was added to the strong resumes to highlight stronger technical expertise (Stanculescu et al., 2009). Weak resumes included fewer technical skills. Differences in certifications (e.g. IELTS and GMAT) and scores were designed to indicate varying levels of ambition and competence. Strong resumes included higher scores and additional certifications. The experience levels of the candidates were differentiated by length of employment and range of responsibilities. Strong profiles showed longer work experience and more hands-on tasks, suggesting a higher level of professional development. Since applicants with more work experience are often seen as more competent, experience was used as a key difference between profiles (Paloniemi, 2006). The described candidate resumes used in the experiment are provided in Appendix B.

The fictitious candidates were named Sophie and Max Schneider. These names were carefully chosen to convey a clear gender identity without causing confusion, which was crucial, given the study's focus on gender perceptions in hiring decisions. Another deciding factor in selecting these names was their popularity among German newborns between 1997 and 2005, which aligned with the expected age range of recent university graduates.[3] This ensured that the names would feel familiar and realistic to survey participants. Schneider, one of the most common surnames in Germany, particularly in the southern regions, was selected to further enhance the authenticity of the candidate profiles.[4]

Participants received an introduction explaining the purpose of the study – to simulate a hiring decision – without revealing its specific focus on gender bias to avoid influencing answers. They were instructed to carefully review the candidate's resume and assess them on key dimensions, including hireability, competence, and potential starting salary. After evaluating the candidate, participants completed a short follow-up questionnaire on gender discrimination. On average, participants spent approximately 8 to 10 minutes completing the entire survey, including both the resume evaluation and post-experiment questionnaire (Kost & da Rosa, 2018). This duration is in line with the best practices for online surveys, where engagement typically declines after 10 minutes.[5] The experiment was carefully controlled to ensure the data could be used to draw valid conclusions about participants' views on gender-related challenges in hiring.

3.2 Survey Design

The survey is designed to collect data in a structured format, comprising three main sections: Demographic Questions, Experimental Evaluation, and Perception of Discrimination Questions. Each section is designed to ensure comprehensive data collection while maintaining data quality and participant engagement. The full version of the survey is provided in Appendix C.

In the first section, demographic data is collected to understand the background and characteristics of the participants. These include 7 questions based on age, gender, country of origin, pursued university degree, field of study, previous field of study, and work experience. Age was included to analyze possible generational differences, such as attitudes toward gender roles (Meagher & Shu, 2019; Yigit & Ali, 2024). Country of origin was included to capture cultural and societal influences, recognizing the diversity of norms across countries, which might be an interesting topic for future research (Cuddy et al., 2015). Questions on current and previous fields of study helped compare STEM and non-STEM participants, focusing on discipline-specific trends. Work experience provided insight into the extent to which participants were affected by workplace dynamics

[3] The names Max and Sophie were among the top 20 most popular baby names in Germany during the 2000s, accessed March 19, 2025, https://www.babymed.com/unassigned/top-german-baby-names-germany-2000s.

[4] Wikipedia, *List of the Most Common Surnames in Germany*, last modified March 19, 2025, https://en.wikipedia.org/wiki/List_of_the_most_common_surnames_in_Germany.

[5] SurveyMonkey, *Survey Completion Times: How Long Do People Take to Complete a Survey?*, accessed March 19, 2025, https://www.surveymonkey.com/curiosity/survey_completion_times/.

that may have influenced their perceptions of bias (Funk & Parker, 2018).

Demographic questions helped to group independent variables and conduct descriptive statistics to better characterize the sample (e.g., gender distribution, average age). In addition, these variables were used to analyze group differences and find relationships between independent and dependent variables, such as the possible influence of gender on hireability ratings. Special attention was given to gender, work experience, and fields of study, as these were the main interests in identifying trends and understanding how background factors contribute to participants' perceptions and evaluations. At this stage, a factual attention check question is included to ensure data quality by identifying inattentive or low-effort responses that could compromise the reliability and integrity of the results (Kung et al., 2018; Muszyński, 2023).

In the second section, participants are introduced to the experimental part of the survey. This study used an experimental audit research design typically used in field studies to analyze discrimination, which simulated hiring scenarios to examine gender bias in candidate evaluations of university students. Students were asked to rate applicants for entry-level electrical engineering roles. The electrical engineering role was chosen with the goal that resumes did not contain specific qualifications unfamiliar to students, as most of the respondents had STEM backgrounds. Students were provided with clear instructions on how to evaluate a candidate's CV, emphasizing that they should carefully review the qualifications and assess them based on several criteria. To improve the quality of responses, participants were asked to give honest evaluations, regardless of whether they thought the application was strong or weak. A request for commitment was also included, as research has shown that securing commitment from participants promotes greater engagement and increases the reliability of the data.[6]

The Candidate Evaluation section assessed whether students exhibit gender biases when evaluating job applicants based on resumes. Specifically, participants rated candidates' competence, hireability, salary expectations, and willingness to assist them in a professional or academic setting. This evaluation provided insights into whether gender influences students' hiring-related perceptions and decision-making. The assessment questions were adapted from Moss-Racusin et al. (2012), who investigated gender bias in hiring decisions among science faculty. This study included the competence evaluation criteria, which is a key determinant of hiring decisions and is often influenced by implicit gender biases (Heilman, 2012). Participants rated the applicant's competence using three questions:

1. Did the applicant strike you as competent?
2. How likely is it that the applicant has the necessary skills for this job?
3. How qualified do you think the applicant is?

Responses were recorded on a 5-point Likert scale (1 = Not at all, 5 = Very much), with higher scores indicating greater perceived competence.

Hireability was evaluated to determine whether the gender of an applicant influenced participants' likelihood of selecting them for a job. Prior research has shown that gender bias often manifests in lower hireability ratings for equally qualified female candidates (Rudman & Glick, 2001). To assess this, participants responded to three Likert-scale questions:

1. How likely would you be to hire the applicant?
2. How likely would you be to recommend the applicant for the job?
3. How likely is it that the applicant would actually be hired for the position?

These items were rated on a 5-point Likert scale (1 = Not at all likely, 5 = Very likely), with higher scores reflecting greater perceived hireability.

Salary allocation indicates how much the evaluator values the candidate's contributions (Gerhart & Rynes, 2003). To examine this, participants were asked: "*If you had to choose a starting salary for this applicant, which of the following would you select?*". Response options ranged from less than € 35,000 to more than € 70,000 per year (before taxes). This salary range was chosen based on market research, which identified € 50,000 as the average entry-level salary for an electrical engineering position.[7] This measure helped determine whether participants were assigned different salaries based on the applicant's gender, reflecting perceived value and career potential.

Professional support was also adapted from the mentoring measure used by Moss-Racusin et al. (2012). Unlike faculty hiring managers, students are not in a position to mentor applicants. Therefore, the mentoring question was changed to better reflect student evaluators' experiences. Instead of asking whether participants would mentor the applicant, the question was: "*If you were working alongside this candidate at your university or in a similar professional setting, how likely would you be to offer assistance if they were struggling with a challenging concept or project?*". The response was recorded using a 5-point Likert scale ranging from Extremely unlikely to Extremely likely.

In the third section, participants answer questions related to their perception of gender-specific challenges in hiring, mentoring, and promotion. More specifically, they answered questions related to perceived discrimination against

[6] Qualtrics, *Attention Checks and Data Quality: How to Improve Survey Responses*, accessed March 19, 2025, https://www.qualtrics.com/blog/attention-checks-and-data-quality/.

[7] PayScale, *Electrical Engineer Salary in Germany*, accessed March 19, 2025, https://www.payscale.com/research/DE/Job=Electrical_Engineer/Salary.

women, discrimination against men, and discrimination that might affect them personally (self-related discrimination). These 5-point Likert-scale questions from Strongly Disagree to Strongly Agree are designed to measure participants' beliefs and attitudes toward gender discrimination and bias within workplace settings, particularly in STEM fields. These questions were adapted from Sipe et al. (2009), who previously examined university students' perceptions of gender discrimination in the workplace. These questions were modified to focus specifically on STEM fields, aligning with the central research question of this study on gender bias in STEM hiring. For example, the original item "Women/Men are less likely to be promoted due to their gender" was adapted to "Women/Men are less likely to be promoted within STEM fields due to their gender". Such modifications ensured that the questions directly reflected the hiring environment relevant to the participants and the study's objectives. The specific questions included:

1. Women are less likely to be hired within STEM fields due to their gender / Men are less likely to be hired within STEM fields due to their gender.
2. Women are less likely to be promoted within STEM fields due to their gender/ Men are less likely to be promoted within STEM fields due to their gender.
3. Women face more challenges in networking and mentorship opportunities within STEM fields due to their gender / Men face more challenges in networking and mentorship opportunities within STEM fields due to their gender.
4. Women are more likely to be underpaid within STEM fields compared to men with similar qualifications / Men are more likely to be underpaid within STEM fields compared to women with similar qualifications.

3.3 Data Collection

Data were collected over a two-month period, from 2 December 2024 to 3 February 2025, using an online survey hosted on the *Qualtrics* platform. Participants accessed the survey through an automatically generated anonymous link, which was distributed via multiple channels, including the university mailing list, *Moodle* platform, class announcements, and *WhatsApp* groups. Additionally, the university's Student Counseling department assisted with promoting the survey to students.

Participants were invited to take part in the study through a message that explained the survey was anonymous and confidential (see Appendix B). After clicking the link, they saw a short description of the study, which included its purpose, the voluntary nature of participation, and the estimated time to complete it. By continuing with the survey, participants gave their consent to take part and allowed their responses to be used for research. *Qualtrics* settings ensured that no personal or sensitive information was collected. Participants could leave the survey at any time without any consequences. To eliminate bias in resume distribution, a randomization tool in *Qualtrics* was used to ensure that each participant received only one of the four possible resumes.

A total of 127 responses were recorded during the data collection period. Of these, 92 were fully completed and included in the final analysis. Participants who did not answer all survey questions, including one who skipped the final item, were excluded. To assess the comparability of the groups, a balancing table was constructed. Table 1 presents descriptive statistics across the four experimental conditions: strong/weak resumes and female/male candidate gender. It includes demographic characteristics of the participants (age, gender, work experience, and STEM background), as well as information on education and region of origin. The results indicate that the groups are generally well balanced, with only minor differences that do not suggest any systematic bias in participant allocation. This supports the conclusion that the random assignment procedure was successful and allows for valid data analysis to follow.

4 Data Analysis

This chapter outlines the analytical strategy used to examine participants' evaluations of hypothetical job candidates and their perceptions of gender-based discrimination in hiring. The analysis includes both descriptive and inferential components, beginning with an overview of variable construction, followed by descriptive insights into perceived discrimination, and concluding with a series of regression models designed to test the main question of the study.

The analysis focuses on four primary outcome variables: competence, hireability, professional support, and recommended starting salary. Two key experimental variables, the candidate's gender, and qualification, form the core of the independent variables. Candidate gender is coded as a binary variable (0 = female, 1 = male), with female as the baseline category. Candidate qualification is likewise binary, where strong candidates are coded as 1 and weak candidates as 0 (reference category). Several control variables are included to account for participants' demographic characteristics. These include gender (binary variable), age (categorized with age groups), country of origin (grouped into broader regional categories), academic background (STEM vs. non-STEM), and work experience (with grouped categories for years of experience). All categorical variables are dummy-coded for regression analyses to isolate the effects of candidate characteristics while accounting for participant background.

To complement the analysis of candidate evaluations, the study also explores how students perceive gender-based discrimination in hiring and workplace dynamics. These attitudes provide a critical context for interpreting evaluation behaviors. Three perception variables were used for this purpose: Discrimination Against Self, Discrimination Against Men and Discrimination Against Women. These

Table 1: Balancing Table of Resume Types

Variable	Strong Female	Strong Male	Weak Female	Weak Male
Mean Age	23.48	24.27	27.13	25.55
	(3.96)	(3.60)	(5.36)	(4.53)
Mean Gender (1 = Male)	0.43	0.43	0.65	0.48
	(0.51)	(0.51)	(0.49)	(0.51)
Mean Work Experience (1 = Yes)	0.83	0.87	0.86	0.95
	(0.39)	(0.34)	(0.35)	(0.21)
Bachelor's Degree (%)	65.22%	52.17%	26.09%	43.48%
Master's Degree (%)	34.78%	47.83%	69.57%	43.48%
PhD (%)	0.00%	0.00%	4.35%	13.04%
Mean STEM (1 = STEM)	0.52	0.52	0.65	0.61
	(0.51)	(0.51)	(0.49)	(0.50)
Regional Distribution				
Caucasus (%)	43.48%	65.22%	21.74%	13.04%
Western Europe (%)	4.35%	8.70%	21.74%	17.39%
Eastern Europe (%)	13.04%	8.70%	17.39%	0.00%
MENA (%)	17.39%	13.04%	21.74%	52.17%
Asia (%)	13.04%	4.35%	17.39%	8.70%
Americas (%)	8.70%	0.00%	0.00%	8.70%
Sample Size	23	23	23	23

Notes: This table presents the balancing of the four resume types used in the experiment. Variables include key participant characteristics such as age, gender, work experience, education level, and STEM, along with regional distribution. STEM is coded as 1 if the participant is from a STEM field and 0 otherwise. MENA refers to the Middle East and North Africa region. Sample sizes represent the number of respondents per resume type.

questions asked about perceived bias in hiring, promotion, salary offers, and mentoring or networking opportunities. While these variables are not included in the regression models as outcomes, they offer descriptive insights into how gender bias is perceived by different student subgroups. Group-based comparisons (e.g., by gender, study field, or work experience) help identify patterns in perception that may inform or parallel evaluation behaviors.

To investigate how candidate characteristics and participant demographics influence candidate evaluations, a series of regression models were estimated. Competence, hireability, and professional support were analyzed using OLS regression. These outcomes were measured on 5-point Likert scales and treated as continuous variables to simplify analysis and increase statistical power. However, for salary recommendations, an OLR model was used. This choice reflects the ordered nature of the salary data, which was collected in categories (e.g., € 35k–€ 45k, € 45k–€ 55k). OLR preserves the ranking of these categories without assuming equal distances between them, making it more suitable than OLS in this case. Robust standard errors were used in all models to correct for potential heteroskedasticity and ensure reliable statistical inference. This approach allows for appropriately analyzing both interval-like and ordinal outcomes within the survey experiment.

The central regression model is specified as follows:

$$Y_i = \beta_0 + \beta_1 \cdot \text{CandGender} + \beta_2 \cdot \text{CandQual} + \beta_3 \cdot (\text{CandGender} \times \text{CandQual}) + X\gamma + \epsilon_i \quad (1)$$

In this specification, the outcome variable Y_i represents one of four participant evaluations of a hypothetical job candidate: competence, hireability, professional support, or recommended starting salary. β_0 captures the baseline evaluation score for a female candidate with weak qualifications, assessed by a participant from the reference group. CandGender is a binary indicator equal to 1 if the candidate is male and 0 if female. CandQual equals 0 if the CV is weak and 1 if strong. The interaction terms CandGender × CandQual are included to assess whether the effects of candidate gender vary depending on the qualification strength of the respondent. X is a vector of control variables capturing participant characteristics (gender, age category, country of origin, study field, and work experience), γ is the associated vector of coefficients, and ϵ_i is the error term. The results of analyses are presented and interpreted in Chapter 5.

5 Results

This chapter presents the findings from both the descriptive analysis and the regression models introduced in Chapter 4. It begins by exploring how participants perceive gender-based discrimination in STEM fields. The second part of the chapter examines how candidate characteristics and participant demographics influence participants' decision-making in the simulated hiring scenarios.

5.1 Descriptive Statistics

This section examines students' perceptions of gender-based discrimination across three dimensions: discrimi-

nation against women, against men, and against oneself. Results vary by gender, work experience, academic background, and regional origin. Full descriptive statistics are provided in Table A 6. Discrimination against women was perceived as the most prevalent dimension of all. Furthermore, female students reported substantially higher awareness in this regard than male students (M = 3.54 vs. 2.65), and non-STEM participants also showed stronger agreement. These results align with prior research suggesting that women may be more attuned to structural gender inequalities. In terms of regional differences, students from the Americas expressed the highest levels of perception with respect to bias against women, while those from Western Europe were the least aware of the matter. In contrast, discrimination against men received the lowest overall ratings, though male students rated it slightly higher than female students (M = 2.26 vs. 2.07). Perceptions were somewhat elevated among STEM students and those with 5–7 years of work experience. When it comes to discrimination perceived on a personal level, students with 1–3 years of work experience reported the highest scores. Female students and those from non-STEM fields again reported greater concern, while participants with more than seven years of experience expressed the lowest levels. These results suggest that early-career students may feel more vulnerable to bias, whereas longer experience may be associated with greater confidence or adaptation to workplace dynamics. Taken together, these findings highlight the diversity of students' beliefs and expectations surrounding gender discrimination in STEM. The differences observed across gender, experience, academic field, and regional background provide important context for the experimental results presented in the next section, where students' actual evaluation behaviors are examined.

5.2 Regression Analyses

This section presents the results of regression models assessing how candidate gender and qualifications, along with participant demographics, influenced evaluations across four outcomes: competence, hireability, professional support, and salary recommendations. Results are grouped by a dependent variable, the key coefficients of each model are discussed in the text, and detailed regression tables are provided below. The regression analyses were performed in multiple stages. First, bivariate models estimated the effect of candidate gender or qualifications on the outcome variable without including participant controls. These were followed by multivariate models that controlled for respondent characteristics such as age, gender, origin, academic background, and work experience. Finally, the full models included interaction terms between candidate gender and qualifications to assess whether evaluation patterns change depending on the strength of the candidate's profile.

Candidate Competence

As shown in Table 2, none of the experimental manipulations significantly affected competence scores. In the bivariate model, male candidates were rated similarly to female candidates ($\beta = -0.13$), and this remained the case after including participant covariates ($\beta = -0.11$). Candidate qualifications also had no significant influence on perceived competence either in isolation or when adjusting for covariates. The interaction between gender and qualifications was consistently non-significant, including in the full model. In contrast, some participant characteristics did shape competence evaluations. Participants aged 23–27 and over 33 provided significantly higher ratings than those under 23. Additionally, respondents from Eastern Europe gave lower ratings to all candidates, suggesting regional variation in evaluation standards. No other demographic covariates had a notable impact.

Candidate Hireability

Table 3 illustrates regression estimates for hireability evaluations. Candidate gender significantly influenced perceived hireability in the bivariate model, with male candidates rated less hireable than female candidates ($\beta = -0.326$). After controlling for participant demographics, this effect remained marginally significant ($\beta = -0.276$), indicating a potential bias favoring female candidates. Candidate qualifications did not significantly predict hireability in any model. However, the interaction between candidate gender and qualification in Model 5 resulted in a marginally significant positive coefficient ($\beta = 0.565$), suggesting strong male candidates may be evaluated more favorably than weak male candidates. This effect weakened in the full model ($\beta = 0.404$) and was no longer significant. Among participant covariates, those aged 23–27 and those over 33 gave higher hireability scores. Additionally, participants from Eastern Europe tended to assign lower ratings. Other demographic variables were not statistically significant.

Professional Support

Table 4 presents the results of OLS regression models of participants' willingness to offer professional support to candidates. In Model 1, which includes only candidate gender, the coefficient for male candidates is negative ($\beta = -0.109$) but not statistically significant, indicating that gender alone does not predict the level of support. When demographic covariates are added in Model 2, the coefficient remains negative ($\beta = -0.147$) and non-significant. In Model 3, the effect of candidate qualifications alone on professional support did not show a significant correlation. However, when covariates are added in Model 4, age shows a significant predictor, especially for older participants. Model 5 introduces the interaction between candidate gender and qualifications. The main effect of being a strong candidate is marginally positive ($\beta = 0.435$), but this is moderated by a marginally negative interaction term ($\beta = -0.652$), suggesting that while strong candidates generally receive more support, this effect may not apply equally across genders. In the full model (Model 6), the interaction term remains negative ($\beta = -0.728$) and marginally significant, indicating that strong male candidates receive less support than expected based on the additive effects of gender and qualifications.

Table 2: OLS Regression – Dependent Variable: Competence Score

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Candidate Male	-0.130	-0.112	–	–	-0.174	-0.064
	(0.165)	(0.153)			(0.256)	(0.207)
Candidate Strong	–	–	0.087	-0.0004	0.043	0.044
			(0.165)	(0.173)	(0.230)	(0.216)
Male × Strong	–	–	–	–	0.087	-0.093
					(0.333)	(0.274)
Constant	3.630***	3.759***	3.522***	3.454***	3.609***	3.498***
	(0.114)	(0.169)	(0.127)	(0.204)	(0.151)	(0.210)
Covariates (Z)	–	✓	–	✓	–	✓
R-squared	0.007	0.343	0.003	0.339	0.011	0.344
Observations	92	92	92	92	92	92

Notes: Robust standard errors in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table 3: OLS Regression – Dependent Variable: Hireability Score

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Candidate Male	-0.326*	-0.276*	–	–	-0.609*	-0.487*
	(0.160)	(0.163)			(0.256)	(0.236)
Candidate Strong	–	–	0.109	0.038	-0.174	-0.145
			(0.164)	(0.189)	(0.196)	(0.210)
Male × Strong	–	–	–	–	0.565*	0.404
					(0.318)	(0.284)
Constant	3.783***	3.825***	3.565***	3.634***	3.870***	3.880***
	(0.098)	(0.215)	(0.134)	(0.280)	(0.130)	(0.273)
Covariates (Z)	–	✓	–	✓	–	✓
R-squared	0.044	0.250	0.005	0.223	0.082	0.265
Observations	92	92	92	92	92	92

Notes: Robust standard errors in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Salary Recommendations

Table 5 presents the results of OLR regression models analyzing salary expectations for the presented candidates. In models 1 and 3, neither candidate gender nor candidate qualification has a statistically significant effect on the recommended salary category. However, the coefficient for a strong candidate is positive ($\beta = 0.530$) in Model 3, indicating a tendency towards higher salary suggestions, though this effect is not statistically significant. When controlling for demographic covariates, strong candidates are significantly more likely to receive higher salary recommendations ($\beta = 0.849$). This effect persists in the full model, although it is no longer statistically significant. In Model 5, we observe a positive, but non-significant interaction between candidate gender and qualifications. Later, the interaction becomes marginally significant in the full model, suggesting that male candidates with strong qualifications may receive higher salary recommendations than their female counterparts. These findings suggest that while gender and qualifications alone do not strongly predict salary recommendations, their interaction and the characteristics of the evaluators themselves may play an influential role. An overview of these findings is depicted in Figures B6–B8.

Influence of Participant Demographics

To explore whether certain participant characteristics influenced how candidates were evaluated, I ran an additional set of models using only demographic variables. The results are presented in Tables A 7 and A 8 in the Appendix A. Participants aged 23–27 ($\beta = 0.557$) and especially those over 33 ($\beta = 0.916$) gave significantly higher competence ratings compared to the youngest group. The same age groups also showed more willingness to offer professional support, with the strongest effect observed among those over 33 ($\beta = 1.605$). Age also had a positive effect on

Table 4: OLS Regression – Dependent Variable: Professional Support Score

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Candidate Male	-0.109	-0.147	–	–	0.217	0.231
	(0.202)	(0.227)			(0.266)	(0.329)
Candidate Strong	–	–	0.109	0.080	0.435*	0.417
			(0.202)	(0.235)	(0.273)	(0.295)
Male × Strong	–	–	–	–	-0.652*	-0.728
					(0.402)	(0.452)
Constant	4.000***	3.637***	3.891***	3.486***	3.783***	3.411***
	(0.139)	(0.307)	(0.133)	(0.332)	(0.198)	(0.368)
Covariates (Z)	–	✓	–	✓	–	✓
R-squared	0.003	0.214	0.003	0.210	0.035	0.245
Observations	92	92	92	92	92	92

Notes: Robust standard errors in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table 5: Ordinal Logistic Regression – Salary Recommendations

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Candidate Male	-0.019	0.216	–	–	-0.453	-0.472
	(0.378)	(0.414)			(0.531)	(0.607)
Candidate Strong	–	–	0.531	0.849*	0.093	0.224
			(0.383)	(0.468)	(0.534)	(0.605)
Male × Strong	–	–	–	–	0.892	1.435*
					(0.762)	(0.866)
Covariates (Z)	–	✓	–	✓	–	✓
AIC	259.96	266.55	258.03	263.48	260.65	264.31
Observations	92	92	92	92	92	92

Notes: Standard errors in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$. *AIC = Akaike Information Criterion; lower values indicate better model fit.*

hireability ratings, although the result for participants over 33 was only marginally significant (β = 0.924). Regional origin also played a role in shaping evaluation behavior. Respondents from Eastern Europe consistently rated candidates lower in terms of competence (β = -0.732) and hireability (β=-0.454). A negative trend was also observed among participants from Asia, particularly in competence and support ratings. On the other hand, students from Asia and the Americas were more likely to assign higher salary categories, indicating potential regional differences in perceived salary expectations. Work experience was especially relevant for salary evaluations. Those with five to seven years (β = -2.941) and more than seven years (β = -2.318) of experience tended to suggest lower starting salaries, which may reflect more cautious or realistic expectations. Other demographic variables, including participant gender and academic background, did not show a consistent or statistically significant influence on candidate evaluations.

5.3 Robustness Check

To make sure that the main results are reliable, I repeated the full model (Model 6) using only participants with a STEM academic background (n = 53). This group is particularly relevant because the study focuses on perceptions of hiring in STEM fields. The goal was to test whether the patterns found in the overall sample also appear among those most likely to enter STEM careers. The results are summarized in Table A 9.

Across all outcome variables, the main effects of candidate gender, qualification, and their interaction were not statistically significant. For competence ratings, however, demographic characteristics continued to play a role: older participants (33+) gave higher competence scores, while participants from Eastern Europe and Asia gave lower scores. A similar pattern was observed in hireability ratings, where no significant treatment effects emerged, although participants from Eastern Europe tended to assign slightly lower scores. For professional support, the only notable result was that older participants were more willing to offer help to strug-

gling candidates. Finally, salary recommendations showed a marginally significant interaction effect ($\beta = 2.55$), indicating a possible tendency to favor highly qualified male candidates. Age and origin again played a role, with older participants and those from Eastern Europe, Asia, and the Americas suggesting higher starting salaries.

Overall, the robustness check using the STEM-only sample confirms that the direction of effects remains similar to the full sample. While most treatment effects are not statistically significant in this subgroup, the influence of participant demographics, particularly age and region of origin, remains evident. These findings support the stability of the results and suggest that background characteristics continue to shape perceptions, even within a more homogeneous STEM population.

6 Discussion and Conclusion

To investigate how university students perceive gender in hiring contexts, this study used an experimental design with manipulated CVs to explore how they evaluate applicants based on standard hiring criteria. The analysis considered both the direct effects of candidate gender and interaction effects with qualification strength, as well as the influence of respondent demographic characteristics. Despite expectations based on generally accepted findings in labor economics and psychology, the study did not reveal a consistent preference for male candidates based on assessment results. In simpler models, male candidates were rated lower on hireability.

The presented results raise an interesting question about how this shift may have occurred among students. One plausible explanation is that the academic and cultural environment of modern universities, where diversity and inclusion are actively promoted, especially in STEM fields, is beginning to create a more egalitarian mentality[8] in students. Perhaps the impact of inclusive discourse has influenced their values, making them less likely to apply traditional gender stereotypes in their evaluations. Social desirability bias may still have influenced responses, as students could have unconsciously adjusted their answers to match internalized social norms about fairness and equality, even without knowing the study's true focus (Fisher, 1993; Krumpal, 2013). While the direct effects of candidate gender were limited, interaction terms revealed more nuanced dynamics. Strong male candidates were marginally more likely to receive higher salary recommendations but less likely to receive professional support compared to other groups. This may reflect typical assumptions in male-dominated fields: strong men are expected to be competent and independent and, therefore, not need help. This pattern aligns with the shifting standards framework (Biernat & Kobrynowicz, 1997), which suggests that people do not use a single, fixed standard when evaluating others. Instead, they adjust their expectations based on social group stereotypes, for example, assuming women in STEM are less common, they might be judged as "exceptional" for meeting the same criteria expected of men. As a result, men might receive less support, not because of bias but because of underlying expectations that they should manage on their own in these roles.

Another key insight is that the evaluators' background, especially their age and region of origin, consistently influenced how they rated candidates across all outcomes. Regional differences, especially among students from Eastern Europe and Asia, may reflect cultural norms around hierarchy, competition, or gender roles. For example, stricter academic standards in some regions may influence how competence is rated, while conservative gender norms might shape attitudes toward female candidates. Moreover, students with more work experience recommended lower starting salaries, suggesting a more labor-market-aware perspective. This indicates that participants' views on fairness and value may have been shaped not only by bias but also by their practical experience and knowledge of how the job market works. In addition to behavioral measures, the study explored students' perceptions of gender-based discrimination. Perceptions of self-related discrimination were strongest among early-career students, suggesting that perceived vulnerability to bias may decline with experience.

While the study contributes valuable insights, several limitations must be noted. First, the sample size (n = 92) limits the ability to detect minor effects and fully explore subgroup variation. Second, as in most online experiments, participants faced no real consequences for their decisions. This may reduce ecological validity[9], as real-life hiring involves risks, incentives, and institutional norms that shape behavior in ways experiments cannot fully simulate. Third, although the average survey completion time was within an acceptable range (8–10 minutes), some participants completed it in under 7 minutes, raising concerns about attentiveness and the depth of engagement with the evaluation task. Fourth, while the resumes were designed to reflect a neutral, entry-level electrical engineering position, participants from non-engineering backgrounds (e.g., business and social sciences) may have lacked the context to accurately interpret technical qualifications despite efforts to make the content broadly understandable. Finally, the study focused solely on gender and qualification strength.

These limitations open several promising directions for future research. One important avenue is to explore intersectional dimensions – such as ethnicity, age of the applicant, or motherhood/fatherhood signals – to better understand how multiple identities may compound bias in hiring evaluations. Additionally, longitudinal research could examine how students' attitudes evolve as they transition into professional roles, revealing whether the relatively egalitarian

[8] According to the Oxford English Dictionary, egalitarianism is "the belief in or advocacy of the principle that all people are equal and deserve equal rights and opportunities."

[9] Ecological validity refers to how well the results of a study can be applied to real-life settings and behaviors outside the experimental environment.

patterns observed in academic settings persist or shift under real-world workplace pressures. Cross-cultural studies could also provide valuable insights by comparing results across different countries or academic systems, helping to identify which institutional features most influence bias-related behavior. Finally, future work should investigate the potential of educational interventions to reduce gender bias and promote more equitable evaluation practices. Can targeted training in bias awareness meaningfully change how people evaluate others? And are such effects durable over time, particularly once individuals enter professional roles and face more complex decision-making environments?

Taken together, the findings suggest that, while students are not immune to bias, their judgments reflect a shift toward greater awareness and caution. Compared to previous generations, today's students seem to be more sensitive to issues of equity and inclusion. However, awareness alone may not eliminate deeper social expectations or implicit assumptions. As bias appears increasingly context-sensitive and less predictable, efforts to combat it must also evolve through education, reflection, and policies that take into account the complex ways in which bias operates in practice.

References

Alan, S., Ertac, S., Kubilay, E., & Loranth, G. (2019). Understanding Gender Differences in Leadership. *The Economic Journal*, *130*(626), 263–289. https://doi.org/10.1093/ej/uez050

Babcock, L., Laschever, S., Gelfand, M., & Small, D. (2003). Nice Girls Don't Ask. *Harvard Business Review*, *81*(10), 14–14.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, *94*(4), 991–1013.

Biernat, M., & Kobrynowicz, D. (1997). Gender- and Race-Based Standards of Competence: Lower Minimum Standards but Higher Ability Standards for Devalued Groups. *Journal of Personality and Social Psychology*, *72*(3), 544.

Correll, S. J. (2001). Gender and the Career Choice Process: The Role of Biased Self-Assessments. *American Journal of Sociology*, *106*(6), 1691–1730. Retrieved March 21, 2025, from http://www.jstor.org/stable/10.1086/321299

Cuddy, A. J., Wolf, E. B., Glick, P., Crotty, S., Chong, J., & Norton, M. I. (2015). Men as Cultural Ideals: Cultural Values Moderate Gender Stereotype Content. *Journal of Personality and Social Psychology*, *109*(4), 622.

Evagorou, M., Puig, B., Bayram, D., Janeckova, H., & European Commission and Directorate-General for Education, Youth, Sport and Culture. (2024). *Addressing the Gender Gap in STEM Education across Educational Levels – Analytical Report*. Publications Office of the European Union. https://doi.org/10.2766/260477

Fisher, R. J. (1993). Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research*, *20*(2), 303–315.

Friedmann, E., & Efrat-Treister, D. (2023). Gender Bias in STEM Hiring: Implicit In-Group Gender Favoritism Among Men Managers. *Gender & Society*, *37*(1), 32–64. https://doi.org/10.1177/08912432221137910

Funk, C., & Parker, K. (2018). Women and Men in STEM Often at Odds over Workplace Equity [https://www.pewresearch.org/social-trends/2018/01/09/women-and-men-in-stem/].

Gerhart, B., & Rynes, S. (2003). *Compensation: Theory, Evidence, and Strategic Implications*. Sage.

Goldin, C., & Rouse, C. (2000). Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. *American Economic Review*, *90*(4), 715–741.

Hardy III, J. H., Tey, K. S., Cyrus-Lai, W., Martell, R. F., Olstad, A., & Uhlmann, E. L. (2022). Bias in Context: Small Biases in Hiring Evaluations Have Big Consequences. *Journal of Management*, *48*(3), 657–692. http://www.jstor.org/stable/10.1086/321299

Heilman, M. E. (2012). Gender Stereotypes and Workplace Bias. *Research in Organizational Behavior*, *32*, 113–135.

Holman, L., Stuart-Fox, D., & Hauser, C. E. (2018). The Gender Gap in Science: How Long until Women Are Equally Represented? *PLoS Biology*, *16*(4), e2004956.

Kost, R. G., & da Rosa, J. C. (2018). Impact of Survey Length and Compensation on Validity, Reliability, and Sample Characteristics for Ultrashort-, Short-, and Long-Research Participant Perception Surveys. *Journal of Clinical and Translational Science*, *2*(1), 31–37. https://doi.org/10.1017/cts.2018.18

Krumpal, I. (2013). Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review. *Quality & Quantity*, *47*(4), 2025–2047.

Kung, F. Y., Kwok, N., & Brown, D. J. (2018). Are Attention Check Questions a Threat to Scale Validity? *Applied Psychology*, *67*(2), 264–283. https://doi.org/10.1111/apps.12108

Meagher, K. D., & Shu, X. (2019). Trends in US Gender Attitudes, 1977 to 2018: Gender and Educational Disparities. *Socius*, *5*, 2378023119851692.

Milkman, K. L., Akinola, M., & Chugh, D. (2015). What Happens Before? A Field Experiment Exploring How Pay and Representation Differentially Shape Bias on the Pathway into Organizations. *Journal of Applied Psychology*, *100*(6), 1678.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science Faculty's Subtle Gender Biases Favor Male Students. *Proceedings of the National Academy of Sciences*, *109*(41), 16474–16479.

Muszyński, M. (2023). Attention Checks and How to Use Them: Review and Practical Recommendations. *Research and Methods*, *32*(1), 3–38. https://doi.org/10.18061/ask.v32i1.0001

Neumark, D. (2018). Experimental Research on Labor Market Discrimination. *Journal of Economic Literature*, *56*(3), 799–866.

Paloniemi, S. (2006). Experience, Competence and Workplace Learning. *Journal of Workplace Learning*, *18*(7/8), 439–450. https://doi.org/10.1108/13665620610693006

Quadlin, N. (2018). The Mark of a Woman's Record: Gender and Academic Performance in Hiring. *American Sociological Review*, *83*(2), 331–360.

Rich, J. (2014). *What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000* (tech. rep.). IZA Discussion Papers.

Rudman, L. A., & Glick, P. (2001). Prescriptive Gender Stereotypes and Backlash toward Agentic Women. *Journal of Social Issues*, *57*(4), 743–762. https://doi.org/10.1111/0022-4537.00239

Sipe, S., Johnson, C. D., & Fisher, D. K. (2009). University Students' Perceptions of Gender Discrimination in the Workplace: Reality Versus Fiction. *Journal of Education for Business*, *84*(6), 339–349. https://doi.org/10.3200/JOEB.84.6.339-349

Stanculescu, M., Marinescu, S., & Cheregi, G. (2009). MATLAB in Electrical Engineering. *Journal of Electrical and Electronics Engineering*, *2*(10).

United Nations Development Programme. (2023). *Gender social norms index (gsni)* [Accessed: March 22, 2025].

Wilkinson, A., & Barry, M. (2020). *The Future of Work and Employment*. Edward Elgar Publishing.

Yigit, Y., & Ali, Y. (2024). Generational Differences in Attitudes towards Gender Roles and Violence against Women. *International Journal of Caring Sciences*, *17*(1), 198–208.

Junior Management Science

www.jums.academy
ISSN: 2942-1861

The Influence of Perceived Undervaluation on the Duration of Share Repurchase Programmes

Lars Kiehne

Humboldt-Universität zu Berlin

Abstract

This thesis examines whether undervaluation—measured using eight market-based proxies—leads, ceteris paribus, to shorter share repurchase programme announcements and a higher probability of tender offers over open market repurchases. Using ordinary least squares and logistic regressions on 329 German repurchase announcements from 2016 to 2025, the analysis finds no significant link between undervaluation and either programme duration or repurchase method. Instead, repurchase size and trading volume emerge as the key drivers of programme design. These results suggest that, in Germany, repurchase structures are driven more by executional factors than by signalling undervaluation.

Keywords: share repurchases; undervaluation; open market repurchases; tender offers

1 Introduction

Since their effective legalisation in 1998, share repurchases have become a recurring component of corporate financial strategy in Germany. Between July 3, 2016, and March 20, 2025, listed firms announced buyback programmes totalling over 104 billion EUR, indicating a consistently active use of this mechanism in recent years. Several motives may underlie such programmes: buybacks can offer a tax-efficient method of returning capital to shareholders, help offset dilution from stock option plans or merger and acquisition (M&A) activity, serve as a defence against hostile takeovers, or be used to optimise a firm's capital structure.

Another widely discussed motivation for buybacks—and the primary focus of this thesis—is the belief that the company's stock is undervalued. According to the market timing hypothesis, supported by Chan et al. (2007) and Ben-Rephael et al. (2014), managers can use private assessments of intrinsic value to generate shareholder gains. By repurchasing undervalued shares, the ownership stake of remaining shareholders increases, and intrinsic value per share rises, as selling shareholders transact below true value (D. Ikenberry & Vermaelen, 1996).

Signalling theory offers a complementary explanation. As proposed by Dharmawan and Mitchell (2007), firms may use repurchases to signal undervaluation to the market. Empirical evidence supports this, with studies documenting abnormal positive returns after repurchase announcements. In jurisdictions like Australia, where disclosure of motives is required, firms citing undervaluation experience larger abnormal returns (Akyol & Foo, 2013). In contrast, most countries—including Germany—do not require such disclosures. Investors must thus infer motivations from public information.

This thesis addresses that challenge by investigating whether the announced duration of a repurchase programme can serve as a signal of perceived undervaluation. Under European Union (EU) regulations introduced on July 3, 2016, firms must disclose the intended time frame of their buyback authorisation (The European Commission, 2016, Article 2, 1. Regulation (EU) 2016/1052). In line with signalling theory, this could form a communication channel: shorter durations may signal urgency and confidence in undervaluation, while longer durations might reflect opportunistic flexibility or other underlying motivations.

That firms consider market conditions when determining the duration of buyback programmes is evident from corporate statements:

> "The Executive Board of Deutsche Börse AG informs, that the period, in which the adopted share buyback program with a volume of around EUR 200 million will be implemented, can extend until the end of the first half-year 2018. An extended implementation period as decided by the Executive Board today enables the company, to react more flexibly towards current

I would like to express my sincere gratitude to my supervisors, Christian Bernard and Ulf Brüggemann, for their valuable guidance and support throughout this thesis. I am particularly grateful for their critical feedback and constructive suggestions, which significantly contributed to improving the quality of this work.

DOI: https://doi.org/10.5282/jums/v11i3pp611-641

© The Author(s) 2026. Published by Junior Management Science.
This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.

> developments and market conditions" (Deutsche Börse AG, 2017, p. 1).

Regarding buyback duration, Thomas Toepfer, CFO of Covestro AG, stated in a quarterly call:

> "We always said that we would do this in an opportunistic and anti-cyclical way. So we do feel that the two year horizon gives us the flexibility to do exactly this. [...] Therefore, we do feel that both from a company, but more specifically from a shareholder perspective, this is the right approach to give us the time to really act flexibly and, as I said, opportunistically with respect to the share buyback" (Covestro AG, 2022, p. 12).

These statements indicate that programme duration is a strategic choice, tailored to anticipated market conditions rather than set arbitrarily.

A study by Gould (2019), based on Australian data, finds that firms may indeed use shorter repurchase durations to signal undervaluation. However, this finding stems from a regulatory environment where firms are obliged to disclose their motives—an advantage absent in most other markets.

This thesis extends the analysis to such a context. I construct eight different measures of perceived undervaluation: four based on deviations of valuation multiples from historical averages, three on peer-based comparisons, and one on prior share price performance. Given the difficulty of proving whether a stock is truly undervalued, the thesis focuses on "perceived undervaluation"—that is, management's belief that the firm is undervalued. This belief, rather than objective mispricing, is what shapes managerial decisions and market signals.

Using data from 299 open market repurchase (OMR) programmes and 30 tender offers, this thesis tests two hypotheses:

> Hypothesis 1: Perceived undervaluation leads ceteris paribus to a shorter announced open market repurchase period.
>
> Hypothesis 2: Companies that are perceived as undervalued are more likely to use a tender offer than a open market repurchase programme.

To test the first hypothesis, I employ ordinary least squares (OLS) regressions with repurchase duration as the dependent variable. For the second, I estimate logistic regressions, with the dependent variable equal to one for a tender offer and zero for an open market repurchase.

The empirical results show that none of the eight valuation proxies significantly influence either repurchase duration nor repurchase type. Consequently, both hypotheses are rejected. These findings suggest that programme length is not used as a credible signal of undervaluation, nor does perceived undervaluation influence the choice between tender offers and OMRs.

This study contributes to the literature on repurchase motivations by offering new evidence on the informational content—or lack thereof—of programme characteristics. The results imply that investors cannot rely on repurchase duration as a signal of management's valuation beliefs. These insights are relevant for both investors and regulators considering disclosure rules for corporate buybacks.

The thesis begins with an overview of institutional background and literature. This is followed by methodology and descriptive statistics. Chapter 5 presents the main empirical analysis, followed by robustness checks. The thesis concludes with limitations, future research directions, and final remarks.

2 Institutional Background

This chapter provides the institutional context for corporate share repurchases in Europe. It begins with an overview of EU regulations governing buyback programmes, including disclosure requirements and trading limitations. It then traces the development of German repurchase law and its key provisions. Finally, the chapter introduces the primary repurchase methods—open market repurchases, tender offers, and accelerated share repurchases—laying the foundation for the detailed analysis that follows.

2.1 European Union Repurchase Regulations

Under EU law, companies may repurchase their own shares without facing market manipulation charges, provided the buybacks serve one of the following purposes: "(a) to reduce the capital of an issuer; (b) to meet obligations arising from debt financial instruments that are exchangeable into equity instruments; or (c) to meet obligations arising from share option programmes, or other allocations of shares, to employees or to members of the administrative, management or supervisory bodies of the issuer or of an associate company" (European Parliament and the Council of the European Union, 2014, Article 5, 2. Regulation (EU) No 596/2014). Before initiating a repurchase programme, companies are required to disclose its maximum monetary amount, the maximum number of shares to be acquired, and the duration of the authorisation (The European Commission, 2016, Article 2(1), Regulation (EU) 2016/1052). The same paragraph further specifies that these conditions may subsequently be modified by the companies, provided that any changes are properly and publicly disclosed.

Buyback activity is also subject to trading restrictions. Specifically, companies are prohibited from repurchasing more than 25% of the average daily trading volume and from paying a price exceeding the last independent trade (The European Commission, 2016, Article 3, 2. & 3. Regulation (EU) 2016/1052).

All companies incorporated in the EU must comply with both EU regulations and relevant national laws governing repurchase activity.

2.2 Repurchase Regulations in Germany and Their Development over Time

Since 1998 German companies are able to repurchase their own shares without naming the motivation for such a programme (Pistor et al., 2003). For this, a German company needs the authorisation from its general shareholder meeting (Aktiengesetz §71, 1998, (1) 8.). This authorisation has a maximum validity of five years, during which the management board is granted the right—but not the obligation—to initiate a repurchase programme (Andres et al., 2016). Consequently, the five-year authorisation also constrains the duration of any individual buyback programme (Aktiengesetz §71, 1998, (1) 8.).

Some exceptions to this rule, already in place since 1965, include instances where the firm faces immediate danger or aims to incentivise employees (Pistor et al., 2003). In such cases, board approval suffices, and no resolution by the general meeting is required. However, in all cases, German companies are not permitted to repurchase more than 10% of their outstanding shares within a five-year period (Aktiengesetz §71, 1998, (2)).

As explained by Andres et al. (2016, p. 421), firms have two primary options for handling repurchased shares: "First, it can treat them as an asset on the asset side of the balance sheet. They can then be used to cover outstanding convertible bonds or executive stock options. The maximum number of shares a firm can hold on its balance sheet is 10% of the shares outstanding. Alternatively, the firm can retire the shares. In this case the firm's book equity is reduced accordingly."

2.3 Different Types of Repurchases

2.3.1 Open Market Repurchases

Open market repurchases are the most common form of share buyback programmes. This method provides companies with the flexibility to repurchase up to the announced number of shares—or none at all—on the stock exchange, allowing them to control both the timing and volume of the repurchases (Oded, 2011). However, acquiring a substantial volume of shares through the open market typically requires a longer time frame. The advantages and disadvantages of this method are discussed in detail in Section 3.4 of this thesis.

2.3.2 Tender Offers

Several forms of tender offers exist. The most common are fixed-price self-tender offers, in which a company announces the number of shares it intends to buy and the price it is willing to pay per share. Shareholders may then choose whether, and how many, shares to tender. If the total number of shares tendered exceeds the amount sought, the allocation is usually made on a pro rata basis (Kamma et al., 1992). To make the offer attractive, the company typically needs to offer a premium over the current share price (Oded, 2011).

Alternatively, companies may conduct what is known as a Dutch auction tender offer. In this case, shareholders may submit bids within a pre-specified price range. When the offer concludes, the company pays a single clearing price to all shareholders who submitted offers at or below that price, allowing it to acquire the targeted number of shares (Kamma et al., 1992).

2.3.3 Accelerated Share Repurchases

Accelerated share repurchase (ASR) programmes allow companies to buy back large quantities of their shares almost immediately through a contract with a financial intermediary, who provides the agreed number of shares in exchange for payment (Bargeron et al., 2011). "The intermediary obtains the shares that it delivers to the repurchasing firm by borrowing them, typically from institutions. The intermediary then covers its short position by purchasing shares in the market over a specified time period, normally several months" (Bargeron et al., 2011, p. 69).

Although ASR programmes are common in jurisdictions such as the United States, they have not been implemented by European firms. Nevertheless, Schulz (2024) finds that such programmes are likely permitted under current EU and German legal frameworks.

3 Literature Review

This chapter reviews the existing literature on corporate share repurchases, with a focus on market reactions, underlying motivations, programme duration, and repurchase methods. It begins with empirical studies on abnormal returns following buyback announcements, then examines theoretical and empirical research on drivers such as undervaluation, managerial incentives, and capital structure considerations. Finally, common methodological approaches are discussed, and the key research gap this thesis addresses is identified.

3.1 Market Reaction to Repurchase Announcements

Numerous studies have examined both short-term and long-term market reactions to share repurchase announcements. An influential paper by D. Ikenberry et al. (1995) reports an abnormal four-year buy-and-hold return of 12.1% following the initial announcement of share buybacks in the US.

These results are supported by a comprehensive paper by Manconi et al. (2019), which looks at abnormal returns of 31 non-U.S. countries. In the vast majority of these countries, buyback announcements are followed by both short-term and long-term positive abnormal share price performance. The authors find a 4 year abnormal return of 20.79% for German repurchasing companies, significant at the 5% level.

Short-term abnormal returns are also evident in Germany, with announcement effects ranging from 2.32% (Andriosopoulos & Lasfer, 2015) to 3.58% (Lee et al., 2010).

However, the long-term positive market reactions seem not to occur gradually but rather stepwise, mainly driven by completion announcements, further announced buyback programmes and takeover attempts (Bargeron et al., 2017).

3.2 Motivations for Repurchase Programmes

There are many reasons why companies choose to repurchase their own shares. To understand why the market generally reacts positively to buyback announcements, it's important to distinguish between the different motives and their respective impact on shareholder value.

3.2.1 Profit Distribution

Share repurchases are the main alternative to dividends when it comes to distributing profits (Grullon & Michaely, 2002). While selling shareholders receive cash, remaining shareholders see their ownership stake increase. Brav et al. (2005) discuss how buybacks offer greater flexibility to companies than dividends, as reducing buybacks when cash is scarce causes less negative market reaction than cutting dividends. In addition, Grullon and Ikenberry (2005) notes that buybacks generally have a tax advantage over dividends at the individual shareholder level, because the remaining shareholders do not have to pay any immediate taxes for their ownership increase.

3.2.2 Employee Share Options

Some companies use buybacks to counter dilution from employee share options that would otherwise decrease the earnings per share (EPS) (Bens et al., 2003). However, the market recognizes this motivation, which leads to a lower announcement return for firms with many outstanding employee share options (Kahle, 2002).

3.2.3 Takeover Defence

Repurchases can serve as a defence against hostile takeovers by reducing the number of shares available to acquirers (Pearce & Robinson, 2004). In a comprehensive empirical analysis Billett and Xue (2007) provide empirical support, showing that firms facing takeover threats are more likely to initiate buybacks.

3.2.4 Misleading Investors

Chan et al. (2010) suggest that firms with poor earnings quality and aggressive accounting practices may use buybacks to mislead investors, benefiting from short-term abnormal returns that dissipate over time.

3.2.5 Capital Structure Adjustments

In addition, debt-financed buybacks can add shareholder value by optimizing the company's leverage to add a tax-shield and reduce agency costs of free cash flow (Lei & Zhang, 2016). That said, Bonaimé et al. (2014, p. 197) report "that the extent to which adjusting capital structure through a share repurchase creates value depends on the undervaluation of the firm." Further, Chan et al. (2004) find no evidence that the market rewards these capital structure adjustments over the long term.

3.2.6 Mergers and Acquisitions

Companies sometimes repurchase their shares to use them as acquisition currency (Andersson et al., 2007). This motivation for share repurchases is barely researched. However, the economic intuition is that companies buy back their shares either before or subsequent to the acquisition to offset the dilutive effect on their shareholders. Faccio and Masulis (2005) show that roughly 20% of acquisitions in Europe are at least partly financed by shares, which often has tax advantages for the seller and aligns the interest between the two parties.

3.2.7 Managerial Performance-Related Compensation

Several studies link repurchases to executive compensation. Cheng et al. (2015) show that managers with bonuses tied to EPS are more likely to initiate buybacks, especially when EPS is close to bonus thresholds. Cook and Zhang (2022) find that such incentives can lead to reduced investment in research and development and human capital.

3.2.8 Undervaluation

The positive short- and long-term price movements discussed at the beginning of this chapter appear to be partly driven by a motivation not yet examined: share repurchases initiated due to a belief that the firm's shares are undervalued. I find this motive particularly compelling and focus my research on it, as it provides management with an opportunity to leverage their internal valuation estimate of the company to enhance shareholder value. When a company repurchases undervalued shares, not only does each remaining shareholder's ownership percentage increase, but so does the intrinsic value of their holdings. This occurs because those who sell their shares to the firm do so at a price below what management believes to be their true value (D. Ikenberry & Vermaelen, 1996).

That undervaluation is indeed a key driver of positive abnormal returns following an announcement is consistent with findings from Australia, where companies have to state their motivation for a repurchase. Both Otchere and Ross (2002) and Akyol and Foo (2013) find that the market reacts especially positively to buybacks when undervaluation is cited as the reason for the buyback.

These market reactions appear justified, as firms that repurchase undervalued stock often outperform in the years following the announcement. Zhang (2005, p. 1887) reports from Hong Kong that "the market responds the most favorably to repurchases that are made by small and value (high book-to-market value) firms. Over a long horizon, there is some evidence that repurchases made by value firms show superior performance. The three-year buy-and-hold abnormal return, which is measured against a portfolio of control firms that are matched by size and book-to-market value ratios, is over 20%."

Similarly D. Ikenberry et al. (1995, p. 206) finds that "firms ranked in the top book-to-market quintile have four-year abnormal performance of 45.3% following the repurchase announcement. This occurs using a benchmark that

explicitly controls for size and book-to-market effects in stock returns."

Further support for this undervaluation hypothesis comes from evidence that managers are capable of timing the market. Dittmar and Field (2015) demonstrate that infrequent repurchasers—those presumed to be acting opportunistically—manage to repurchase shares at prices 6% to 8% lower than those paid by a naïve investor. This aligns with the findings of Stephens and Weisbach (1998), who show that firms tend to increase repurchase activity following declines in their share price, suggesting that management responds strategically to perceived market mispricing.

3.3 Share Repurchase Programme Duration

Gould (2019) finds that shorter repurchase programmes tend to have higher completion rates and higher announcement related stock returns, because they are seen as more credible. Furthermore, the results indicate that "if firms wish to signal the undervaluation of their shares then [...] investors evaluate this signal based on the intended program length indicated in an announcement" (Gould, 2019, pp. 242–243). The effect of the programme duration on the market reaction is therefore especially strong in cases where companies state undervaluation as their repurchase motive. However, in this Australian study undervalued is a "dummy variable which takes on the value of 1 if the program announcement indicates 'undervaluation' or 'good investment' as a motive." (Gould, 2019, p. 246) and not a market based valuation variable. This limits the application in most jurisdictions, but shows, that the duration on a programme is chosen consciously by the management. My research extends these findings, making them more applicable to other countries.

That the programme length matters is further supported by Ota et al. (2022), who show that to increase the credibility of an OMR programme, "managers of repurchasing firms increase the repurchase plan size and shorten the repurchase plan period with the magnitude of the simultaneously announced bad news" (Ota et al., 2022, p. 17). Furthermore, the authors find, that "the market reacts positively to the signal strength adjustments of both repurchase plan size and period" (Ota et al., 2022, p. 11).

Few other studies have specifically examined the duration of share repurchase programmes. However, Mietzner (2017) find that shorter programmes are more likely to be cancelled, and Akyol and Foo (2013) report that firms with shorter durations are more likely to repurchase no shares at all. These findings partially contradict the signalling hypothesis, which suggests that a shorter duration should reflect stronger managerial conviction about undervaluation and thus enhance the credibility of the signal.

3.4 Tender Offers Versus Open Market Repurchases

This section outlines the factors influencing the choice between tender offers and OMRs.

Banyi et al. (2008) show that between 1984 and 2004, OMRs outnumbered tender offers by nearly 13 to 1 in the US. This preference is likely due to the flexibility OMRs provide. While tender offers commit the firm to repurchase a specified number of shares at a fixed price, OMRs give management discretion to buy back fewer—or even no—shares (Oded, 2011).

However, this flexibility comes at a cost. Acquiring shares via the open market is typically slower, particularly for large repurchase volumes. Tender offers, by contrast, enable rapid execution, allowing firms to efficiently deploy idle cash (Oded, 2011) and act quickly on perceived undervaluation (Vermaelen, 1981). Because they are binding and often involve a premium, tender offers are also viewed as a stronger and more credible signal of undervaluation (Vermaelen, 1981).

Empirical evidence supports this signalling interpretation. Vafeas (1997) find that firms experiencing negative abnormal returns in the months leading up to a repurchase announcement are more likely to opt for an OMR, possibly to avoid the costs associated with a tender offer. Indeed, due to asymmetric information, tender offers often require a premium to induce shareholders to sell. The greater the information asymmetry, the higher this premium, and thus the greater the cost to remaining shareholders (Oded, 2011).

The role of ownership structure remains debated. Oded (2011) argue that firms with concentrated insider ownership may prefer OMRs, as high information asymmetry in such cases makes the premium for tender offers particularly costly. On the other hand, Vermaelen (1981) suggests that if insiders refrain from tendering their shares, it sends a credible undervaluation signal by aligning insider and outsider interests. Empirical findings favour the latter view: higher insider ownership is associated with a greater probability of using tender offers (Vafeas, 1997).

3.5 Research Gap

As outlined above, existing literature offers limited insights into determinants of announced share repurchase durations. More specifically, to the best of my knowledge, no prior study examines whether publicly available valuation metrics—used as proxies for perceived undervaluation—are systematically related to programme length. This remains an under-explored area in the broader discourse on corporate payout policy and signalling.

While most jurisdictions do not require firms to disclose their motives for initiating buybacks, signalling theory suggests that managers may convey private information—such as internal assessments of undervaluation—through observable actions (Vermaelen, 1981). One such indirect signal may lie in the design of the repurchase programme, including its duration and method.

A shorter announced duration may signal to investors strong confidence in current undervaluation and an intent to act swiftly before markets adjust. Conversely, longer durations can reflect a more flexible strategy or other underlying motivations.

Similarly, the choice of repurchase method may have signalling implications. Tender offers—requiring greater commitment and enabling the repurchase of a larger share volume over a shorter period—are often perceived as more credible signals of undervaluation than open market repurchases, which are more flexible and less binding (Comment & Jarrell, 1991).

These theoretical considerations motivate the central hypotheses tested in this study:

> Hypothesis 1: Perceived undervaluation leads ceteris paribus to a shorter announced open market repurchase period.
>
> Hypothesis 2: Companies that are perceived as undervalued are more likely to use a tender offer than a open market repurchase programme.

This thesis contributes to the literature by examining how structural features of repurchase programmes may act as valuation signals, and by developing an empirical framework linking observable valuation indicators to programme design.

4 Methodology

This chapter presents the empirical framework used to assess how various undervaluation metrics influence the length of share repurchase programmes among German-listed firms. The initial (pre-cleaning) sample includes all repurchases announced between July 3, 2016 and March 20, 2025. Following a description of the model specification and variable construction, I outline the data sources, cleaning procedures, and key descriptive statistics that underpin the analysis.

4.1 Model Specification

To test whether perceived undervaluation leads to a shorter announced repurchase period, I estimate the following OLS regression, where the dependent variable *Duration* represents the announced programme length in days:

$$\begin{aligned} Duration = \alpha &+ \beta_1 * \ln(Valuation) + \beta_2 * \ln(Size) \\ &+ \beta_3 * \ln(Volume) + \beta_4 * \ln(Liquidity) \\ &+ \beta_5 * \ln(EquityRatio) + \beta_6 * Dilution \\ &+ \beta_7 * \ln(1 + Dividends) \end{aligned}$$

To test the second hypothesis—that companies perceived as undervalued are more likely to implement a tender offer rather than an open market repurchase—I use a logistic regression model. The dependent variable $TenderOffer^P$ indicates the probability that the repurchase programme is a tender offer. For greater interpretability, I also report the marginal effects. The logistic model is specified as follows:

$$\begin{aligned} TenderOffer^P = \Lambda * (\alpha &+ \beta_1 * \ln(Valuation) \\ &+ \beta_2 * \ln(Size) + \beta_3 * \ln(Volume) \\ &+ \beta_4 * \ln(Liquidity) \\ &+ \beta_5 * \ln(EquityRatio) \\ &+ \beta_6 * Dilution \\ &+ \beta_7 * \ln(1 + Dividends) \\ &+ \beta_8 * Ownership) \end{aligned}$$

4.2 Choice of Independent Variables

1. *Valuation*: To capture firms' valuation levels at the time of a share repurchase announcement, I construct eight distinct variables, all based on publicly available data that may plausibly influence managerial perceptions of undervaluation (Jenter, 2005).

The first four variables assess each company's current valuation relative to its own historical average. Specifically, these are calculated as the ratio of the firm's market capitalisation at the announcement date to one of the following financial fundamentals: Earnings Before Interest, Tax, Depreciation and Amortization (EBITDA), net income, cash flow from operations (CFO)—each measured over the last twelve months (LTM) prior to the announcement—and book value of equity. These current valuation multiples are then benchmarked against their historical counterparts, computed as the average of the same multiples measured at each quarter-end over the preceding three years.

The next three variables measure relative valuation by comparing the announcing firm's current LTM multiples—based on EBITDA, net income, and CFO—to the contemporaneous average multiples of a selected peer group. These peer-relative ratios aim to identify valuation deviations from industry norms.

Finally, I include another market-based valuation proxy, following Stephens and Weisbach (1998) and Jagannathan et al. (2000): the firm's share price performance over the 365 days preceding the repurchase announcement. This backward-looking indicator captures market sentiment and performance trajectory, which may influence managerial perceptions of misvaluation as well.

The underlying assumption across these valuation measures is that lower current multiples—whether relative to historical norms, peer firms, or prior share price performance—signal a perception of undervaluation. Consistent with this, I hypothesise that firms will aim to repurchase shares more aggressively when they believe the stock is trading below its intrinsic value. Specifically, I expect firms to announce shorter buyback periods under such circumstances, reflecting a desire to execute repurchases while the stock remains undervalued.

For the second hypothesis, I anticipate a negative association between a firm's overall valuation and its probability of adopting a tender offer. Specifically, lower valuations are expected to increase the probability of a firm choosing a tender

offer, as this structure facilitates a swift and concentrated repurchase, signalling managerial confidence and capitalising on perceived market mispricing.

2. *Size*: The size of the repurchase programme in EUR relative to the market capitalisation at the time of announcement. I expect that the larger the proportion of shares to be repurchased, the longer the duration of the programme and the more likely the use of a tender offer.

3. *Volume*: This variable is defined as the ratio of the announced number of shares to be repurchased to the average daily trading volume on the relevant stock exchange during the 90 days prior to the announcement. A higher ratio indicates that the targeted repurchase volume represents a larger share of market activity, potentially making it more difficult to complete the programme quickly via open market transactions. As a result, I expect a higher *Volume* ratio to be associated with longer programme durations and to increase the likelihood of a tender offer, which facilitates faster execution despite lower market liquidity.

4. *Liquidity*: The quick ratio of the company, defined as (Cash + Accounts Receivable) divided by Current Liabilities. Higher liquidity should lead to a shorter programme because the company already has the cash available for repurchases. Accordingly, I expect a higher quick ratio to increase the likelihood of a tender offer rather than an open market repurchase.

5. *Equity Ratio*: This variable captures capital structure adjustments as a potential motivation. It is calculated as book equity divided by total assets, using the latest available data prior to the announcement. Bonaimé et al. (2014) demonstrate that both book and market values (market capitalisation as equity) are appropriate measures of equity ratio in such models. I expect a higher *Equity Ratio* to be associated with a shorter programme duration, ceteris paribus, and to increase the probability of a tender offer, as management might value the stronger signalling effect of such a programme.

6. *Dilution*: Shareholder dilution resulting from option plans or M&A activities, measured as the percentage increase in outstanding shares between the last annual report and the most recent one preceding the announcement. I expect that higher levels of *Dilution* will be associated with longer buyback periods and a preference for open market repurchases over tender offers. This is because such dilution is generally independent of the current share price, as is the intention to offset it—thereby reducing the urgency to repurchase shares swiftly or to signal undervaluation to the market.

7. *Dividends*: The average dividend payout yield over the past three years, calculated as total dividends paid over a calendar year divided by the firm's market capitalisation as of December 31st of the same year. *Dividends* indicate that the company has excess cash that it cannot invest at reasonable returns. I expect a higher dividend payout ratio to correlate with a longer buyback period, as profit distribution depends less on share price and more on cash generation. If distribution is the main motivation, companies are likely to prefer an OMR to avoid paying the tender premium.

8. *Ownership*: For the second hypothesis, I include this variable to control for ownership concentration, as discussed in Section 3.4. Based on Vafeas (1997), I expect larger insider ownership to decrease the probability of a tender offer.

4.3 Data Collection

To compile the underlying dataset on share repurchase announcements, I utilised the S&P Capital IQ Transactions database. Initially, this database identified 326 announcements within the specified period from July 3, 2016 (the effective date of Regulation (EU) 2016/1052) to March 20, 2025—the date on which data collection commenced. However, key information such as the programme start and end dates, as well as programme size, were not consistently populated. Consequently, I manually reviewed each transaction to extract the following details: announcement date, programme start and end dates, number of shares targeted, maximum monetary value, and the stated reason for the repurchase. In cases where multiple programme announcements were listed under a single transaction entry, each programme was separated and recorded individually. Following this step, the pre-cleaning dataset comprised 374 observations. Where specific data fields were missing, I referred directly to the original announcements published on companies' investor relations websites to complete the dataset.

For financial statement information, I used Capital IQ's Companies database, retrieving data for all firms incorporated and listed in Germany. The following financial variables were collected on a calendar-quarter basis: cash flow from operations, net income, EBITDA, total assets, book value of equity, current liabilities, accounts receivable, cash holdings, and ownership data. Additionally, using the same database and filters, I obtained annual data on shares outstanding, revenues, and dividends paid. Industry classifications used for constructing peer valuation benchmarks are based on Capital IQ's primary industry classification.

Lastly, for stock market data, I utilised LSEG's Datastream. From this source, I retrieved daily stock prices, trading volumes, and market capitalisation figures for all German companies identified by their ISINs beginning with "DE," as well as index data for the CDAX from 2015 onwards.

4.4 Data Preparation

To prepare the raw dataset for regression analysis, I employ custom Python routines that implement the variable constructions described in Section 4.2, along with several additional adjustments and assumptions, which I explain in this section.

For firms absent from the Datastream market-capitalisation series, I estimate daily market capitalisation by multiplying shares outstanding at the most recent calendar year-end by the closing price on the relevant trading day.

When repurchase announcements specify only the maximum number of shares to be acquired—without indicating a corresponding EUR value—I approximate the programme's monetary size by multiplying the announced share count by

the share price on the trading day preceding the announcement.

Additionally, because negative valuation multiples are not economically meaningful and extremely high positive multiples—often caused by small denominators—can unduly influence estimates, I exclude observations with negative multiples and winsorise the remaining values at the 95th percentile within each quarter.

To identify peer companies, each sample firm is matched with all other firms sharing the same primary industry classification, regardless of whether they conduct repurchases. For each peer group, I calculate the same financial metrics on the same dates as for the sample firm. The peer multiple is then defined as the average of these values across the peer group, with group sizes varying by firm.

To calculate the abnormal return of a stock following the share repurchase announcement, I subtract the absolute return of the CDAX index in the exact same time period from the total return of the company's stock. The CDAX index represents all German listed companies and is market capitalisation-weighted.

To enhance the interpretability of the descriptive statistics and ensure consistency across models, I exclude all announcements with at least one missing value for any control variable. This step reduces the dataset from 374 to 329 observations but does not materially affect the sample composition.

Finally, as part of the estimation procedure, I employ robust standard errors to improve the reliability of the regression results. Specifically, HC1 heteroskedasticity-consistent standard errors are used for the OLS regressions, while HC3 standard errors are applied for the logistic regressions. This adjustment mitigates potential biases caused by heteroskedasticity and small-sample distortions.

4.5 Repurchases in Germany: A Sector and Size Analysis

During the sample period, 117 German companies announce at least one share repurchase programme. Table 1 presents the industry distribution of these companies. The range of sectors is notably broad. Firms categorised under *Application Software* constitute the largest single group, accounting for nearly 6% of all repurchasing companies.

A comparison between repurchasing and non-repurchasing firms reveals modest but observable industry differences. Notably, only 3.4% of the repurchasing companies operate as asset managers or custody banks, whereas this group comprises approximately 9.3% of the non-repurchasing firms. Conversely, healthcare equipment firms and industrial conglomerates each represent 3.4% of the repurchasing sample, but less than 1% of the non-repurchasing group. While these differences suggest some variation in industry participation, no clear structural pattern emerges in the overall industry distribution.

In contrast, the financial characteristics presented in Table 2 highlight more pronounced distinctions between the two groups. A substantial number of holding companies with minimal operating activities skew the mean figures—particularly for the non-repurchasing firms—as evidenced by the high skewness values and large standard deviations reported in the table. As each firm is treated as a separate observation in the empirical analysis that follows, I place greater emphasis on median values when interpreting financial indicators.

Repurchasing firms tend to be significantly larger than their non-repurchasing counterparts, with median *Revenues* and *Market Capitalisation* exceeding the latter group by more than a factor of ten. Furthermore, repurchasing firms exhibit superior operational performance: both *Operating Income Margin* and *Return on Assets* are approximately twice as high as those of firms that do not repurchase shares. Taken together, these findings suggest that repurchasing companies are, on average, larger and more profitable than non-repurchasing firms.

4.6 Monthly and Annual Dynamics of Repurchase Activity

Figure 1 illustrates the distribution of share repurchase announcements across the sample period. The figures for 2016 and 2025 appear artificially low, as these years are only partially included in the observation window. Among the fully covered years, 2017 records the fewest announcements.

Between 2017 and 2020, repurchase activity increases gradually but steadily. This trend accelerates in 2021, with announcements rising from 35 in 2020 to 49 in 2021, and peaking at 53 in 2022. Thereafter, the number of announcements declines slightly, falling to 40 in 2024.

Interestingly, 2022—the year with the highest number of repurchase announcements—coincides with a year of negative returns in the German stock market. By contrast, other years with elevated repurchase activity, notably 2021, 2023, and 2024, correspond to periods in which the CDAX index registers positive annual returns (STOXX Ltd., 2025).

A forthcoming study by Dittmann et al. (2022) finds that share repurchase initiations are more frequent in the second, fifth, eighth, and eleventh months of the fiscal year—typically coinciding with earnings announcements—because trading restrictions such as blackout periods are least stringent during these intervals. The alignment of relaxed trading windows with the release of updated financial information facilitates increased repurchase activity and enables executive equity sales shortly after earnings disclosures.

In Germany, listed firms are required to publish annual earnings within 90 days of the fiscal year-end and to report quarterly or half-year financials within 45 days of the respective reporting period (Regierungskommission Deutscher Corporate Governance Kodex, 2022). Since the fiscal year coincides with the calendar year for most firms, annual earnings are usually disclosed in February or March, while quarterly results are published in May, August, and November.

Figure 2 displays the monthly distribution of repurchase announcement dates for the firms in my sample.

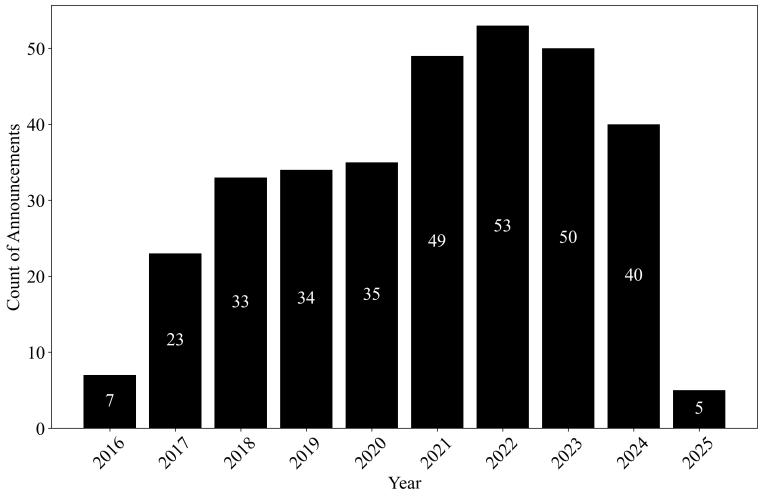


Figure 1: Distribution of Repurchase Announcements Over the Years

Source: Data sourced from Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: The figure illustrates the annual distribution of share repurchase announcements made by German companies between July 3, 2016, and March 20, 2025. The X-axis represents calendar years within the sample period, while the Y-axis indicates the total number of repurchase announcements recorded each year. It is important to note that the figures for 2016 and 2025 are artificially low, as these years are only partially covered by the sample window—beginning mid-2016 and ending in early 2025—resulting in incomplete data for those periods.

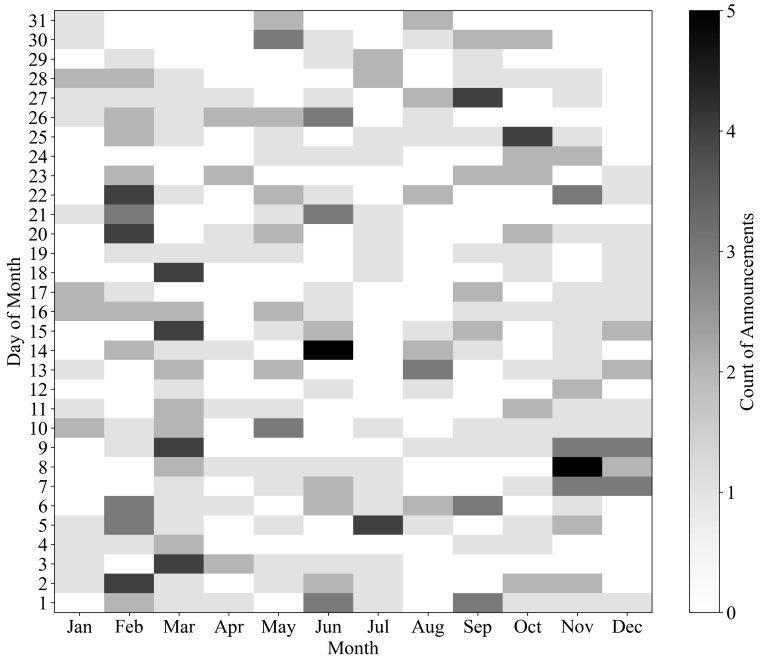


Figure 2: Calendar Year Distribution of Repurchase Announcements

Source: Data sourced from Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: The figure displays the distribution of share repurchase announcements across the calendar year using a heatmap format. The X-axis represents the months from January to December, while the Y-axis denotes the days of each month. Each cell indicates the number of repurchase announcements made on a given day, with shading intensity corresponding to the count of announcements, as shown in the color bar on the right. Darker shades represent higher frequencies of announcements.

There is a noticeable concentration of announcements around 21 February and mid-March. In total, more than 12% of programmes are announced in both February and March each, and over 10% occur in November. These figures exceed the 8.3% that would be expected if announcements were evenly distributed throughout the year. By comparison, May accounts for 8.8% of announcements, while August records only 6.1%.

Table 1: Distribution of German Companies by Primary Industry and Share Repurchase Activity

Primary Industry	Count		Percentage	
	(Rep.)	(Non-Rep.)	(Rep.)	(Non-Rep.)
Application Software	7	20	5.98	4.89
Industrial Machinery and Supplies and Components	5	26	4.27	6.36
IT Consulting and Other Services	5	15	4.27	3.67
Asset Management and Custody Banks	4	38	3.42	9.29
Electronic Equipment and Instruments	4	8	3.42	1.96
Health Care Equipment	4	4	3.42	0.98
Industrial Conglomerates	4	1	3.42	0.24
Construction Machinery and Heavy Transportation Equipment	4	9	3.42	2.20
Real Estate Operating Companies	4	22	3.42	5.38
Specialty Chemicals	3	7	2.56	1.71
Heavy Electrical Equipment	3	3	2.56	0.73
Multi-line Insurance	2	3	1.71	0.73
Systems Software	2	2	1.71	0.49
Specialized Finance	2	3	1.71	0.73
Semiconductors	2	2	1.71	0.49
Office Services and Supplies	2	2	1.71	0.49
Automobile Manufacturers	2	5	1.71	1.22
Investment Banking and Brokerage	2	9	1.71	2.20
Integrated Telecommunication Services	2	2	1.71	0.49
Apparel, Accessories and Luxury Goods	2	2	1.71	0.49
Health Care Technology	2	3	1.71	0.73
Diversified Support Services	2	2	1.71	0.49
Diversified Chemicals	2	0	1.71	0.00
Aerospace and Defense	2	3	1.71	0.73
Other	44	218	37.61	53.30
Total	117	409	100.00	100.00

Source: The presented data is obtained from Capital IQ. For more details, see Section 4.3.
Notes: This table presents the distribution of German companies across primary industries, categorized by whether they announced a share repurchase programme ("Rep.") or not ("Non-Rep.") between July 3, 2016, and March 20, 2025. Industry classifications follow Capital IQ's March 2025 taxonomy. Percentages indicate the share of companies within each repurchase status group that fall under the respective industry. Industries with fewer than two repurchasing companies are grouped into the *Other* category.

Table 2: Financial Analysis of German Repurchasing and Non-Repurchasing Companies

	Mean	Std. Dev.	Min.	25%	Median	75%	Max.	Skew
Companies With at Least One Announced Repurchase Programme (n = 117)								
Revenues (m EUR)	12,354.89	2,977.92	2.82	258.30	1,140.20	10,143.10	155,498.00	3.51
Operating Income Margin (%)	9.22	32.23	-167.92	3.45	9.40	13.78	81.42	-4.07
Return on Assets (%)	5.50	1.10	-10.82	2.16	4.89	7.97	29.11	0.57
Market Capitalisation (m EUR)	11,559.63	1,618.92	10.47	291.24	1,205.38	7,683.10	171,499.30	3.67
Companies Without Announced Repurchase Programme (n = 409)								
Revenues (m EUR)	27,446.71	18,340.40	0.00	18.20	103.40	613.70	322,284.00	13.96
Operating Income Margin (%)	22.27	3,540.02	-19,466.67	-3.22	4.21	10.83	66,796.12	15.93
Return on Assets (%)	5.84	12.97	-112.50	-1.25	2.60	6.83	46.48	-2.94
Market Capitalisation (m EUR)	26,996.58	6,261.94	0.90	23.71	102.17	486.88	72,898.19	7.54

Source: The presented data is obtained from Capital IQ and LSEG Datastream. For more details, see Section 4.3.
Notes: This table presents descriptive statistics for the financial profiles of German companies, categorized based on whether they announced at least one share repurchase programme between July 3, 2016, and March 20, 2025, or not. For each variable, the mean, standard deviation, minimum, 25th percentile, median, 75th percentile, maximum and skewness are reported. *Revenues* are reported in millions of EUR for the calendar year 2023. *Operating Income Margin* is calculated as operating income divided by revenues, expressed as a percentage. *Return on Assets* is defined as operating earnings divided by average total assets (2022–2023), also expressed as a percentage. *Market Capitalisation* is stated in millions of EUR as of December 31, 2023. Financial data from 2023 is used due to its greater availability at the time of writing compared to more recent figures.

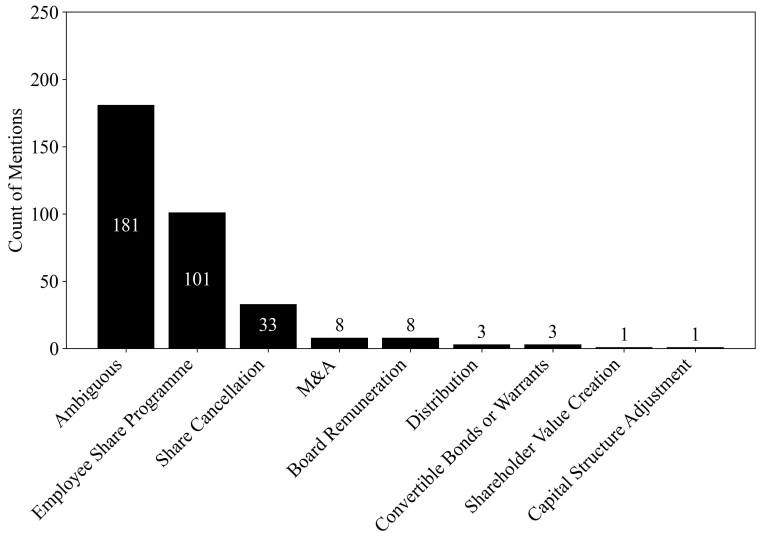


Figure 3: Stated Reasons for the Repurchase Programme

Source: The presented data is obtained from Capital IQ.
Notes: This chart presents the reasons companies state for their repurchase programmes announced between July 3, 2016 and March 20, 2025. The *Ambiguous* bar contains companies which either name no reason or state more than two reasons, since naming too many reasons makes it impossible for observers to identify the real underlying motivation behind a repurchase programme. Some companies publish two reasons, each of which is counted separately. The total number of reasons is therefore larger than the number of repurchase announcements. To improve the interpretability I combine reasons with the same underlying motivation but slightly different wording into one category.

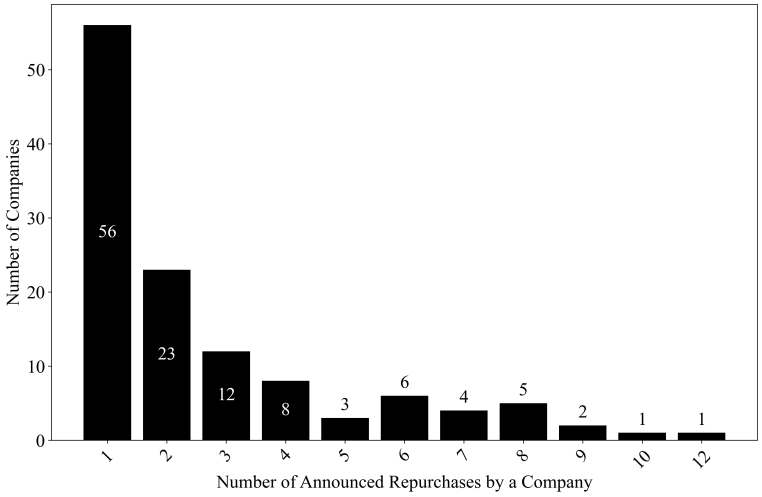


Figure 4: Number of Announced Repurchase Programmes by Company

Source: Data sourced from Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: The figure illustrates the distribution of firms based on the number of share repurchase programmes they have announced between July 3, 2016, and March 20, 2025. The X-axis denotes the number of distinct repurchase announcements made by individual companies within the sample period. The Y-axis represents the corresponding number of companies that fall into each category. The data indicates that the majority of firms (56) announced only one repurchase programme, with progressively fewer firms engaging in multiple repurchase announcements.

Overall, the data suggest that firms are more likely to announce share repurchase programmes around the time of their annual earnings disclosures, but not necessarily in conjunction with quarterly announcements.

4.7 Repurchase Reasons and Number of Announced Repurchases by Company

Some companies in my sample voluntarily disclose the reasons for their share repurchase announcements, despite not being legally obliged to do so. Figure 3 presents the distribution of these stated motivations. The *Ambiguous* category includes firms that either provide no rationale or cite more than two distinct reasons, as listing multiple motives can obscure the primary driver behind the repurchase. In cases where two reasons are given, each is recorded separately. Consequently, the total number of reported reasons exceeds the number of individual repurchase announcements.

To facilitate interpretation, I group semantically similar justifications—albeit phrased differently—into broader categories based on their underlying intent. The majority of announcements fall into the *Ambiguous* category, highlighting the analytical challenges involved in inferring managerial motives from public disclosures alone.

Notably, 101 announcements explicitly mention the use of repurchased shares for employee share programmes, making this the most frequently cited reason. The second most common rationale is share cancellation. While this may signal managerial perceptions of undervaluation, it can also reflect a variety of strategic objectives, thereby limiting its informativeness from an investor's perspective.

Less frequently cited motivations include M&A activity, capital distribution, and adjustments to capital structure. However, each of these categories is mentioned in eight announcements or fewer, indicating that they are not primary drivers for most repurchase decisions.

Figure 4 illustrates the distribution of repurchase activity by firm over the sample period. A total of 56 companies announce only one programme, while 23 firms initiate two. In aggregate, 274 repurchase programmes are launched by companies that conduct multiple buybacks, and 228 announcements originate from firms that repurchase shares on three or more occasions over the nearly nine-year window.

At the upper end of the distribution, Allianz SE announces share repurchase programmes ten times, while IVU Traffic Technologies AG does so on twelve occasions—demonstrating a recurring use of repurchases as part of their corporate financial strategy.

4.8 Descriptive Statistics

4.8.1 Open Market Repurchase Programmes

Table 3 presents summary statistics for the 299 open market share repurchase programmes announced between July 3, 2016 and March 20, 2025. On average, firms announce programmes with a potential duration of 236 days, although this figure conceals substantial variation. The standard deviation is high (298.5 days), and the distribution is heavily right-skewed, ranging from zero to 1,815 days. The median duration of 150 days indicates that most programmes are relatively short, with only a few long-duration announcements extending the upper tail.

The descriptive statistics for the valuation multiples suggest that, at the time of announcement, firms typically trade slightly below their historical norms. The median values for the *EBITDA*, *Net Income*, *CFO*, and *Price to Book Value Multiples* are all below one—ranging from 0.89 to 0.96—indicating mild undervaluation on a historical basis. However, the distribution is broad: minimum values fall well below 0.5, and the upper quartile values exceed 1.1 in all cases. This suggests that while a typical firm may be modestly undervalued, there exists substantial heterogeneity, with a non-trivial share of firms trading at significantly depressed levels relative to their historical averages.

Valuation ratios relative to peer firms also display considerable variation. The median peer-relative valuation multiples are 0.80 for *EBITDA*, 0.78 for *Net Income*, and 0.76 for *CFO*, reinforcing the impression that many firms are undervalued not only in absolute terms but also compared to their industry peers. The interquartile ranges for these metrics are wide, especially for *CFO Peer Valuation*, which ranges from 0.42 at the 25th percentile to 1.51 at the 75th. These findings point to significant dispersion in perceived relative value across firms at the time of repurchase announcements.

Figures 5 and 6 illustrate the relationship between programme duration and valuation ratios, but no clear pattern emerges. While some firms with high valuation multiples—relative to historical averages or peers—appear as outliers, the data do not consistently associate low valuations with shorter programme durations. Notably, there is a concentration of announcements with a duration of exactly 365 days, which likely reflects a standard one-year time frame adopted for reasons of administrative convenience rather than strategic responses to valuation signals.

The average *Prior Share Performance* over the 365 days preceding the announcement is modestly positive at 4.5%, but the distribution is highly dispersed. The standard deviation exceeds 37 percentage points, with returns ranging from –80% to +207%. The median is slightly negative (–0.02%), and the 25th percentile stands at –18.1%. Figure 7 illustrates the relationship between *Prior Share Performance* and programme duration; again, no clear trend emerges, suggesting that *Prior Share Performance* does not systematically influence the length of repurchase programmes.

Regarding control variables, the *Size* of the announced programme relative to market capitalisation is generally modest, with a mean of 3.2% and a median of 1%. Most firms aim to repurchase a relatively small proportion of their outstanding shares. The *Volume* variable—defined as the number of shares targeted relative to average daily trading volume—shows high variability, with a mean of 23.02 and a standard deviation of 45.36.

Liquidity, proxied by the quick ratio, has a median value of 0.82, indicating that more than half of the firms lack sufficient liquid assets to meet short-term liabilities. A median *Equity Ratio* of 38.3% suggests that firms are predominantly debt-financed. *Dilution* is negligible, with a median of zero, implying that recent share count changes have been limited. *Dividend* yields are modest but common, with a median of 1.6%. Finally, *Ownership*—measuring the proportion of shares held by publicly disclosed investors—is relatively

Table 3: Descriptive Statistics – Market Repurchase Programmes (n = 299)

	Obs	Mean	Std. Dev.	Min.	25%	Median	75%	Max.
Announced Programme Duration	299	236.237	298.516	0.000	29.500	150.000	355.000	1,815.000
EBITDA Multiple	284	0.950	0.322	0.291	0.756	0.909	1.132	2.773
Net Income Multiple	276	0.975	0.461	0.088	0.712	0.911	1.162	3.798
CFO Multiple	273	0.976	0.496	0.131	0.689	0.890	1.156	4.282
Price to Book Value Multiple	296	0.963	0.242	0.372	0.796	0.958	1.135	2.175
EBITDA Peer Valuation	253	1.050	1.068	0.055	0.491	0.803	1.272	10.972
Net Income Peer Valuation	241	1.028	0.903	0.032	0.531	0.779	1.149	6.449
CFO Peer Valuation	236	1.312	1.815	0.018	0.420	0.760	1.507	15.047
Prior Share Performance	298	4.511	37.187	-79.933	-18.065	-0.019	23.225	206.542
Size	299	0.032	0.095	0.000	0.003	0.010	0.031	1.000
Volume	299	23.017	45.355	0.002	2.580	9.926	23.169	446.230
Liquidity	299	5.070	33.794	0.010	0.599	0.819	1.285	429.545
Equity Ratio	299	0.369	0.198	-0.080	0.224	0.383	0.500	0.971
Dilution	299	0.007	0.071	-0.188	-0.005	0.000	0.000	0.866
Dividends	299	0.022	0.018	0.000	0.009	0.016	0.035	0.111
Ownership	299	0.625	0.203	0.003	0.514	0.658	0.786	1.000

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: This table provides descriptive statistics for companies that announced at least one share market repurchase programme within the period from July 3, 2016, to March 20, 2025. For each variable, the table presents the number of observations, mean, standard deviation, minimum, 25th percentile, median, 75th percentile, and maximum. *Announced Programme Duration* measures the number of days between the announced starting date of the repurchase and the latest possible programme end date. The three variables—*EBITDA* (Earnings Before Interest, Tax, Depreciation and Amortization) *Multiple*, *Net Income Multiple*, and *CFO* (Cash-Flow from Operations) *Multiple*—represent the ratio of the current LTM valuation multiple to the average LTM valuation multiple over the past three years. *Price to Book Value* (PB) *Multiple* is a ratio of the current PB multiple divided by the average PB multiple of the last three years. *EBITDA Peer Valuation*, *Net Income Peer Valuation*, and *CFO Peer Valuation* are defined as the ratio of a firm's current LTM valuation multiple to the average current multiple of its peer group. *Prior Share Performance* is the company's share price performance over the last 365 days before the announcement multiplied by 100. The control variables are defined as follows: *Size* measures the maximum monetary value of the repurchase programme relative to the company's market capitalisation; *Volume* represents the ratio of the number of shares sought to the average daily trading volume; *Liquidity* corresponds to the quick ratio; *Equity Ratio* is calculated as book equity divided by total assets; *Dilution* captures the change in outstanding shares between the two most recent annual reports; and *Dividends* reflects the company's average dividend yield over the three years prior to the repurchase announcement. *Ownership* denotes the percentage of publicly known owners of the company. For more detailed definitions of the variables, see Section 4.2.

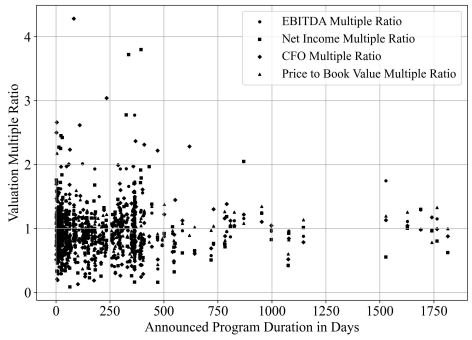


Figure 5: Valuation Multiple Ratio vs. Announced Programme Duration

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3. *Notes:* The X-axis shows the *Announced Programme Duration in Days* measured by the number of days between the announced starting date of the repurchase and the latest possible programme end date. The Y-axis shows four variables. Three of the four main variables—*EBITDA* (Earnings Before Interest, Tax, Depreciation and Amortization) *Multiple*, *Net Income Multiple*, and *CFO* (Cash-Flow from Operations) *Multiple*—represent the ratio of the current LTM valuation multiple to the average LTM valuation multiple over the past three years. Similarly, the *Price to Book Value Multiple* is defined as the ratio of the current market capitalization to the current book value of equity, divided by the average of this multiple over the past three years.

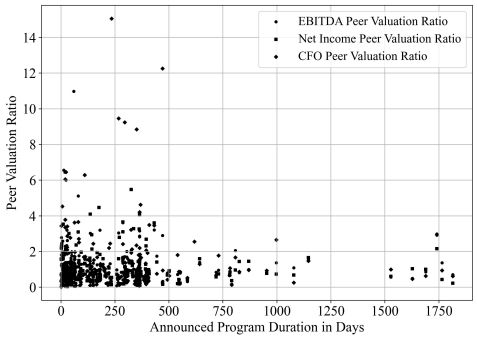


Figure 6: Peer Valuation Ratio vs. Announced Programme Duration

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3. *Notes:* The X-axis shows the *Announced Program Duration in Days* measured by the number of days between the announced starting date of the repurchase and the latest possible programme end date. The Y-axis shows three variables—*EBITDA* (Earnings Before Interest, Tax, Depreciation and Amortization) *Peer Valuation*, *Net Income Peer Valuation*, and *CFO* (Cash-Flow from Operations) *Peer Valuation*, which represent the ratio of a firm's current LTM valuation multiple to the average current multiple of its peer group.

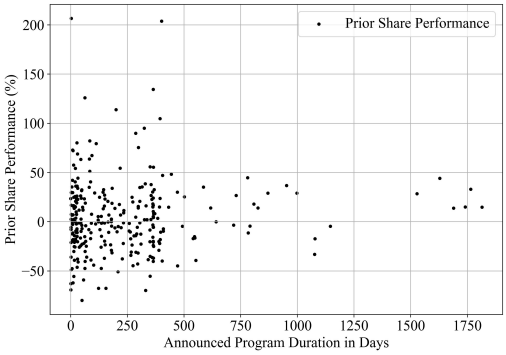


Figure 7: Prior Share Performance vs. Announced Programme Duration

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3. *Notes:* The X-axis shows the *Announced Programme Duration in Days* measured by the number of days between the announced starting date of the repurchase and the latest possible programme end date. The Y-axis shows the variable *Prior Share Performance*, which is the company's share price performance over the last 365 days before the announcement multiplied by 100.

high, with a mean of 62.5%.

4.8.2 Tender Offer Programmes

Table 4 presents descriptive statistics for the 30 firms in the sample that announced at least one tender offer pro-

gramme between July 3, 2016 and March 20, 2025. On average, firms conducting tender offers exhibited valuation multiples that were below their historical norms. The mean *EBITDA Multiple* relative to the three-year historical average is 0.895, with a median of 0.800. Similarly, the average *Net Income Multiple* is 0.844 (median: 0.775), indicating that firms tended to be trading at a discount based on their earnings metrics at the time of the announcement. The picture is less clear for the *CFO Multiple* where the mean is 1.131 but the median is below 1 at 0.859.

The average *Price to Book Value Multiple* relative to its three-year historical average is 0.836, with a median of 0.789, again pointing to a valuation below historical benchmarks.

Relative to peer firms, valuation ratios are also generally lower. The mean *EBITDA Peer Valuation* is 0.776, with a median of 0.468, and the mean *Net Income Peer Valuation* is 0.617 (median: 0.353), further reinforcing the perception of relative undervaluation. Notably, the *CFO Peer Valuation* shows a higher mean of 1.557, though accompanied by substantial dispersion (standard deviation: 2.276); the median stands at 0.644, indicating considerable skewness and variation across firms in the sample.

The mean *Prior Share Performance* of firms announcing tender offers is -10.5%, with a median of -14.3% and a wide range from -79.7% to +47.1%. The negative average and median returns suggest that many firms experienced notable price declines in the year preceding the announcement. This aligns with the broader theme of tender offer announcements being associated with perceived undervaluation in both absolute and relative terms.

Regarding the control variables, the median *Size* is 5.3%, with a median of 4.5%, meaning that companies on average plan to repurchase around 5% of their outstanding shares via a tender offer.

The average *Volume* is 380, with a median of 77, indicating that tender offers often represent significant interventions in the firm's trading activity compared to companies that use OMRs. The average *Liquidity*, measured by the quick ratio with a median of 1.38, suggests that these firms generally maintain relatively healthy short-term solvency positions.

The median *Equity Ratio* stands at 0.566, indicating a generally conservative capital structure. *Dilution* is close to zero for most firms, with a mean of 0.024 and a median of 0.000, implying that firms engaging in tender offers did not typically expand their share base prior to announcement. *Dividend* yields are modest, with a median of 1.3%. *Ownership* has a mean of 56.2% and a median of 68.4%, suggesting a lower average but higher median ownership concentration than for OMRs.

4.8.3 Skewness

Table 5 reports the skewness of the variables used in this thesis. Skewness is a measure of distributional asymmetry. A distribution is said to be right-skewed (positively skewed) when the majority of values are concentrated near the lower end, with a long tail extending towards higher values. This pattern typically arises when a few extreme observations inflate the mean. Conversely, a left-skewed (negatively skewed) distribution is characterised by a concentration of higher values and a tail that stretches towards the lower end.

Only four variables in the dataset display approximate symmetry, with skewness values less than 0.5. One variable—*Ownership*—is moderately left-skewed, with a skewness of –0.846, indicating that publicly disclosed ownership levels are generally high, but a minority of firms exhibit notably lower levels. Three additional variables are moderately right-skewed (skewness between 1 and 2), while the remaining 13 variables exhibit severe right skewness, with values well above 2. Notable examples include *Size*, *Volume*, and *Liquidity*, with skewness values of 8.698, 11.406, and 10.315, respectively. Similarly, several valuation metrics demonstrate substantial right skewness, indicating the presence of extreme observations that are far removed from the majority of the sample, as can be seen in Figures 5, 6, and 7.

To address these distributional characteristics and better satisfy the assumptions of classical linear regression, I apply logarithmic transformations to most of the skewed variables. However, I retain some variables in their raw form. Specifically, *Prior Share Performance* and *Dilution* are not transformed, as their values may be either negative or positive, making log transformation inappropriate. Likewise, all abnormal return variables are kept in their original form for the same reason. *Ownership* is also retained in its raw form, as a logarithmic transformation of its moderate left skewness would exacerbate the results. The variable *Dividends*, which can take on zero values, is transformed using a log-plus-one specification in order to retain all observations while reducing skewness.

All other continuous variables, including valuation multiples, peer-relative valuations, and financial ratios such as *Size*, *Volume*, *Liquidity*, and the *Equity Ratio*, were log-transformed to mitigate the effects of extreme values and improve the interpretability of the regression results. This approach allows for more robust statistical analysis in the empirical sections that follow, as discussed in the next chapter.

5 Analysis Results

This chapter first presents the results for the two hypotheses. It then offers a comprehensive discussion, highlighting that operational factors—not alternative motives—are the primary drivers of buyback programme design.

5.1 Empirical Analysis

5.1.1 Hypothesis 1: Influence of Undervaluation on Programme Duration

Table 6 and 7 present the results of testing Hypothesis 1: that perceived undervaluation leads, ceteris paribus, to shorter announced repurchase durations. I conduct eight

Table 4: Descriptive Statistics – Tender Offer Programmes (n = 30)

	Obs	Mean	Std. Dev.	Min.	25%	Median	75%	Max.
EBITDA Multiple	28	0.895	0.422	0.364	0.654	0.800	1.028	2.472
Net Income Multiple	27	0.844	0.458	0.072	0.596	0.775	1.005	2.247
CFO Multiple	25	1.131	1.081	0.218	0.651	0.859	1.005	5.699
Price to Book Value Multiple	30	0.836	0.254	0.199	0.716	0.789	0.952	1.449
EBITDA Peer Valuation	24	0.776	0.771	0.214	0.359	0.468	0.995	3.907
Net Income Peer Valuation	23	0.617	1.045	0.036	0.279	0.353	0.530	5.258
CFO Peer Valuation	22	1.557	2.276	0.097	0.384	0.644	1.584	10.400
Prior Share Performance	29	-10.534	25.611	-79.721	-25.926	-14.286	7.207	47.059
Size	30	0.053	0.035	0.008	0.023	0.045	0.081	0.136
Volume	30	379.526	782.326	11.355	40.901	77.021	289.157	3,850.544
Liquidity	30	1.750	1.239	0.014	0.999	1.377	2.014	5.386
Equity Ratio	30	0.580	0.175	0.317	0.446	0.566	0.732	0.899
Dilution	30	0.024	0.150	-0.067	0.000	0.000	0.000	0.813
Dividends	30	0.014	0.014	0.000	0.000	0.013	0.019	0.048
Ownership	30	0.562	0.325	0.009	0.381	0.684	0.773	1.000

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: This table provides descriptive statistics for companies that announced at least one share tender offer within the period from July 3, 2016, to March 20, 2025. For each variable, the table presents the number of observations, mean, standard deviation, minimum, 25th percentile, median, 75th percentile, and maximum. The three variables—*EBITDA* (Earnings Before Interest, Tax, Depreciation and Amortization) *Multiple*, *Net Income Multiple*, and *CFO* (Cash-Flow from Operations) *Multiple*—represent the ratio of the current LTM valuation multiple to the average LTM valuation multiple over the past three years. *Price to Book Value* (PB) *Multiple* is a ratio of the current PB multiple divided by the average PB multiple of the last three years. *EBITDA Peer Valuation*, *Net Income Peer Valuation*, and *CFO Peer Valuation* are defined as the ratio of a firm's current LTM valuation multiple to the average current multiple of its peer group. *Prior Share Performance* is the company's share price performance over the last 365 days before the announcement multiplied by 100. The control variables are defined as follows: *Size* measures the maximum monetary value of the offer relative to the company's market capitalisation; *Volume* represents the ratio of the number of shares sought to the average daily trading volume; *Liquidity* corresponds to the quick ratio; *Equity Ratio* is calculated as book equity divided by total assets; *Dilution* captures the change in outstanding shares between the two most recent annual reports; *Dividends* reflects the company's average dividend yield over the three years prior to the tender offer announcement; and *Ownership* denotes the percentage of publicly known owners of the company. For more detailed definitions of the variables, see Section 4.2.

ordinary least squares regressions to examine this assumption. Table 6 contains results based on the four historic multiple approaches described in Section 4.2, while Table 7 covers the peer valuation and prior share price performance approaches.

None of the main independent variables used to measure perceived undervaluation are statistically significant. Their p-values range from $p = 0.182$ for *Prior Share Performance* to $p = 0.819$ for *ln(Net Income Multiple)*. Four of the main variables have positive coefficients, and four have negative coefficients. These results provide no empirical support for Hypothesis 1.

The R^2 values of the regressions range from 0.603 to 0.630, indicating that the variables in the models explain roughly 60% of the variance in the dependent variable. This explanatory power stems mainly from the two control variables, *ln(Size)* and *ln(Volume)*, which are statistically significant at the $p < 0.001$ level in all eight regressions. Both variables have positive coefficients.

To illustrate the interpretation of these coefficients, consider *ln(Size)* in the first regression of Table 6: the *Announced Programme Duration in Days* is, ceteris paribus, 0.4018 percent higher on average for observations with one percent higher repurchase *size* ratios. This kind of interpretation applies to all log-transformed variables in these regressions.

Raw variables, such as *Prior Share Performance*, can be interpreted differently. For example, in the fourth regression of Table 7: the *Announced Programme Duration in Days* is, ceteris paribus, 0.22 percent higher on average for observations with a one percentage point increase in the *Prior Share Performance*. However, due to the lack of statistically significance, the raw coefficient values should not be used to assert influence over the dependent variable.

Interestingly, the variable *ln(Equity Ratio)* is significant at the $p < 0.05$ level in regressions one, three, and four of Table 6, and regressions two and four of Table 7. It is also significant at the $p < 0.01$ level in models one and three of Table 7. These results suggest that firms with higher equity ratios, ceteris paribus, tend to announce shorter repurchase durations, as indicated by the negative coefficients.

5.1.2 Hypothesis 2: Tender Offer Versus Open Market Repurchase

Tables 8 to 11 present the results for Hypothesis 2, which posits that perceived undervaluation increases the likelihood of a tender offer relative to an open market repurchase programme. However, the data does not provide strong support for this hypothesis. Of the eight regression models, only one undervaluation proxy—*ln(EBITDA Peer Valuation)*—is statistically significant, and only marginally so ($p = 0.040$). The remaining proxies yield p-values ranging from $p = 0.081$ for *Prior Share Performance* to $p = 0.975$ for *ln(CFO Multiple)*, offering no compelling evidence that undervaluation meaningfully influences the choice of repurchase method.

Table 5: Skewness of Variables

Variable	Skewness
Announced Programme Duration	2.928
EBITDA Multiple	1.279
Net Income Multiple	2.118
CFO Multiple	3.205
Price to Book Value Multiple	0.319
EBITDA Peer Valuation	4.647
Net Income Peer Valuation	2.674
CFO Peer Valuation	4.039
Prior Share Performance	1.431
Size	8.698
Volume	11.406
Liquidity	10.315
Equity Ratio	0.146
Dilution	7.390
Dividends	1.142
Ownership	–0.846
Abnormal Announcement Return (3 Days)	0.436
Abnormal Return (30 Days)	0.336
Abnormal Return (180 Days)	13.200
Abnormal Return (365 Days)	2.825
Abnormal Return (730 Days)	2.082
Abnormal Return (1095 Days)	2.479

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: This table reports the skewness of the variables used in this thesis. Four variables are relatively symmetric ($|skew| < 0.5$). *Ownership* is moderately left skewed, meaning that a larger proportion of observations are concentrated on the higher end of the scale, with a long tail toward lower ownership values. Three more variables are highly right skewed ($1 < |skew| < 2$). The remaining 13 variables are severely right skewed ($|skew| > 2$), meaning that most observations are clustered near the lower end of the distribution, with a small number of extreme values pulling the mean upward. These distributions require transformation to satisfy normality assumptions in statistical analyses. For detailed definitions of the variables, see Section 4.2 or Tables 3 and 15.

As in previous analyses, most of the explanatory power stems from *ln(Volume)*, which is highly significant across all models. The variable *ln(Equity Ratio)* also shows consistent significance at the $p < 0.01$ level in all regressions. Additionally, *ln(1 + Dividends)* is statistically significant in three models ($p < 0.05$), while *ln(Liquidity)* reaches significance in one model.

Given the use of logistic models to test this hypothesis, the coefficients in Tables 8 and 10 indicate only the direction of effects and not their magnitude. The marginal effects reported in Tables 9 and 11 allow for quantitative interpretation. For example, in the first model of Table 9, a 1% increase in the volume ratio is associated with an average increase of approximately 0.0567 percentage points in the likelihood of a tender offer, holding other factors constant.

Overall, the models exhibit good fit, as indicated by their Pseudo R^2 and log-likelihood values. However, this is largely attributable to *ln(Volume)*, with limited contribution from the undervaluation proxies. Other variables with modest significance will be discussed further in the subsequent sections.

5.2 Discussion

5.2.1 Hypothesis 1: Influence of Undervaluation on Programme Duration

Hypothesis 1 predicts that perceived undervaluation shortens the announced duration of share repurchase programmes. However, the empirical analysis provides no support for this proposition. Across eight model specifications, undervaluation—measured using a range of market-based proxies—does not significantly influence programme duration. Consequently, Hypothesis 1 is rejected: perceived undervaluation, as defined in this study, does not appear to affect the length of announced repurchase programmes.

These findings contrast with those of Gould (2019) and Ota et al. (2022), as discussed in Section 3.3. One plausible explanation is that German managers do not use programme duration as a signalling mechanism for undervaluation, in contrast to their counterparts in other jurisdictions. Alternatively, differences in research design may account for the discrepancy. While the present study employs continuous, market-based valuation metrics, the aforementioned studies rely on dummy variables based on firms' stated repurchase motives.

Notably, the only motivation-related variable that exhibits a consistent influence on programme duration is the equity ratio. Given that the models already control for liquidity, the most plausible explanation is that firms with higher leverage face covenant-based constraints on cash outflows, including share repurchases. In contrast, firms with stronger equity positions are less restricted and can execute buybacks more swiftly.

The role of the equity ratio is particularly striking, considering that only one firm explicitly cites capital structure optimisation as the primary motive for their repurchase activity (see Figure 3). Beyond this, programme duration appears to be largely determined by operational factors, as reflected in the significance of the control variables *ln(Size)* and *ln(Volume)*. These capture logistical constraints in executing large repurchase volumes within the open market.

In sum, the announced duration of repurchase programmes in Germany conveys limited information about managerial perceptions of undervaluation, and appears to reflect practical implementation considerations rather than strategic signalling.

5.2.2 Hypothesis 2: Tender Offer Versus Open Market Repurchase

A similar pattern emerges for Hypothesis 2. Although the variable *ln(EBITDA Peer Valuation)* is statistically significant at the $p < 0.05$ level in one model specification, undervaluation does not consistently influence the choice between tender offers and open market repurchase programmes. Accordingly, Hypothesis 2 is rejected.

Instead, the primary determinant of repurchase method appears to be the *Volume* ratio. This result is intuitive, as

Table 6: Hypothesis 1 - Historic Multiple Ratio Approach

Dependent Variable:	ln(Announced Programme Duration in Days)			
	(1)	(2)	(3)	(4)
ln(EBITDA Multiple)	-0.0819			
	(0.172)			
ln(Net Income Multiple)		0.0314		
		(0.137)		
ln(CFO Multiple)			0.0550	
			(0.103)	
ln(Price to Book Value Multiple)				0.1802
				(0.260)
ln(Size)	0.4018***	0.4126***	0.3996***	0.3372***
	(0.069)	(0.067)	(0.074)	(0.086)
ln(Volume)	0.2523***	0.2594***	0.2649***	0.3075***
	(0.061)	(0.060)	(0.068)	(0.073)
ln(Liquidity)	-0.0355	-0.0268	-0.0491	-0.0312
	(0.040)	(0.038)	(0.049)	(0.039)
ln(Equity Ratio)	-0.1972*	-0.1441	-0.2286*	-0.1997*
	(0.085)	(0.083)	(0.091)	(0.084)
Dilution	0.0032	-0.3410	0.0775	-1.6315
	(1.474)	(1.455)	(1.410)	(1.788)
ln(1+Dividends)	-0.5330	-0.1211	-1.2328	-0.7567
	(3.426)	(3.386)	(3.345)	(3.197)
Constant	5.8939***	5.9948***	5.8530***	5.4793***
	(0.430)	(0.407)	(0.452)	(0.553)
Estimator	OLS	OLS	OLS	OLS
Observations	284	276	272	296
R^2	0.624	0.622	0.627	0.620
Adjusted R^2	0.614	0.612	0.617	0.610

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: This table presents the results of four OLS regressions, with the natural logarithm of the announced programme duration as the dependent variable and valuation multiples and control variables as independent variables. *Announced Programme Duration in Days* measures the number of days between the announced starting date of the repurchase and the latest possible programme end date. Log transformations are applied to all independent variables except *Dilution*. The natural log (ln) is used to reduce skewness. Coefficients on log-transformed variables represent approximate percentage changes in the dependent variable for a 1% increase in the independent variable. *Dilution* is included in its original scale. *Dividends* is transformed as *ln(1 + Dividends)* to accommodate zero values. Three of the four main variables—*EBITDA* (Earnings Before Interest, Tax, Depreciation and Amortization) *Multiple*, *Net Income Multiple*, and *CFO* (Cash-Flow from Operations) *Multiple*—represent the ratio of the current LTM valuation multiple to the average LTM valuation multiple over the past three years. Similarly, the *Price to Book Value Multiple* is defined as the ratio of the current market capitalisation to the current book value of equity, divided by the average of this multiple over the past three years. The control variables are defined as follows: *Size* measures the maximum monetary value of the repurchase programme relative to the company's market capitalisation; *Volume* represents the ratio of the number of shares sought to the average daily trading volume; *Liquidity* corresponds to the quick ratio; *Equity Ratio* is calculated as book equity divided by total assets; *Dilution* captures the change in outstanding shares between the two most recent annual reports; and *Dividends* reflects the company's average dividend yield over the three years prior to the repurchase announcement. For more detailed definitions of the variables, see Section 4.2. Robust standard errors of the type HC1 are used. Standard errors are shown in parentheses. The asterisks indicate two-sided significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The main variables for these OLS regressions have the following p-values: *ln(EBITDA Multiple)*: $p = 0.634$, *ln(Net Income Multiple)*: $p = 0.819$, *ln(CFO Multiple)*: $p = 0.592$, *ln(Price to Book Value Multiple)*: $p = 0.489$.

companies are generally restricted to repurchasing no more than 25% of the average daily trading volume (The European Commission, 2016, Article 3, Paragraphs 2–3, Regulation (EU) 2016/1052). Firms constrained by this limit may resort to a tender offer when seeking to repurchase a substantial number of shares within a short time frame.

As with programme duration, a higher *Equity Ratio* is positively associated with the likelihood of opting for a tender offer. This supports the earlier interpretation: firms with lower leverage face fewer constraints, giving them greater flexibility to carry out repurchases via tender offers.

Additionally, the significance of the *Dividends* variable in three out of eight model specifications suggests that payout policy may influence the choice of repurchase method. Specifically, firms with higher dividend distributions are more likely to favour OMR programmes over tender offers. This finding aligns with expectations, as dividends typically indicate a commitment to regular, ongoing shareholder returns—an approach that complements the incremental nature of open market repurchases.

6 Robustness Checks

This chapter validates the models and conclusions from earlier chapters using robustness checks. It assesses multicollinearity, then re-examines the drivers of tender offers versus short-term open market repurchases, confirming that undervaluation is not a primary factor. Finally, it explores the link between repurchase duration and subsequent share

Table 7: Hypothesis 1 - Peer Valuation & Share Price Performance Approaches

Dependent Variable:	ln(Announced Programme Duration in Days)			
	(1)	(2)	(3)	(4)
ln(EBITDA Peer Valuation)	-0.1022			
	(0.084)			
ln(Net Income Peer Valuation)		-0.0929		
		(0.112)		
ln(CFO Peer Valuation)			-0.0215	
			(0.060)	
Prior Share Performance				0.0022
				(0.002)
ln(Size)	0.4239***	0.4228***	0.4502***	0.3414***
	(0.069)	(0.067)	(0.070)	(0.083)
ln(Volume)	0.2962***	0.2779***	0.2966***	0.3041***
	(0.063)	(0.061)	(0.061)	(0.071)
ln(Liquidity)	-0.0586	-0.0672	-0.0668	-0.0329
	(0.041)	(0.042)	(0.053)	(0.038)
ln(Equity Ratio)	-0.2698**	-0.2474*	-0.3028**	-0.1988*
	(0.102)	(0.098)	(0.103)	(0.083)
Dilution	-0.0384	0.0331	0.3741	-1.5247
	(1.402)	(1.491)	(1.361)	(1.813)
ln(1+Dividends)	-3.6175	-2.1342	-4.8385	-0.4498
	(3.304)	(3.497)	(3.193)	(3.150)
Constant	5.8502***	5.8956***	5.9789***	5.4768***
	(0.415)	(0.407)	(0.425)	(0.529)
Estimator	OLS	OLS	OLS	OLS
Observations	253	241	235	295
R^2	0.628	0.603	0.630	0.621
Adjusted R^2	0.617	0.591	0.619	0.612

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: This table presents the results of four OLS regressions, with the natural logarithm of the announced programme duration as the dependent variable and measures of undervaluation and control variables as independent variables. *Announced Programme Duration in Days* measures the number of days between the announced starting date of the repurchase and the latest possible programme end date. Log transformations are applied to all independent variables except *Dilution* and *Prior Share Performance*. The natural log (ln) is used to reduce skewness. Coefficients on log-transformed variables represent approximate percentage changes in the dependent variable for a 1% increase in the independent variable. *Dilution* and *Prior Share Performance* are included in their original scale. *Dividends* is transformed as *ln(1 + Dividends)* to accommodate zero values. Three of the four main variables—*EBITDA* (Earnings Before Interest, Tax, Depreciation and Amortization) *Peer Valuation*, *Net Income Peer Valuation*, and *CFO* (Cash-Flow from Operations) *Peer Valuation*—represent the ratio of a firm's current LTM valuation multiple to the average current multiple of its peer group. The fourth main independent variable: *Prior Share Performance* is the company's share price performance over the last 365 days before the announcement multiplied by 100. The control variables are defined as follows: *Size* measures the maximum monetary value of the repurchase programme relative to the company's market capitalisation; *Volume* represents the ratio of the number of shares sought to the average daily trading volume; *Liquidity* corresponds to the quick ratio; *Equity Ratio* is calculated as book equity divided by total assets; *Dilution* captures the change in outstanding shares between the two most recent annual reports; and *Dividends* reflects the company's average dividend yield over the three years prior to the repurchase announcement. For more detailed definitions of the variables, see Section 4.2. Robust standard errors of the type HC1 are used. Standard errors are shown in parentheses. The asterisks indicate two-sided significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The main variables for these OLS regressions have the following p-values: *ln(EBITDA Peer Valuation)*: $p = 0.226$, *ln(Net Income Peer Valuation)*: $p = 0.405$, *ln(CFO Peer Valuation)*: $p = 0.721$, *Prior Share Performance*: $p = 0.182$.

price performance and whether announcement returns differ for initial versus repeated repurchase programmes.

6.1 Multicollinearity

To assess the validity of my models, I test them for multicollinearity. Table 12 presents the Variance Inflation Factors (VIFs), which measure how much the variance of a regression coefficient is inflated due to multicollinearity among the independent variables. Most variables exhibit VIFs slightly above 1, indicating no multicollinearity concerns. However, *ln(Size)* and *ln(Volume)* show VIFs ranging from 3.375 to 5.226. These values approach or slightly exceed the commonly used threshold of 5, which suggests moderate multicollinearity.

The observed correlation between these two variables is expected, as *Size* reflects the total number of shares (or market capitalisation) targeted in the repurchase, while *Volume* is defined as the ratio of the same number of shares to the average daily trading volume (see Section 4.2). Thus, both variables share the same numerator, inherently linking them.

Despite this moderate multicollinearity, both variables are retained in the models, as they capture distinct dimensions of repurchase programmes. Specifically, *Size* reflects the overall scale of the intended buyback, whereas *Volume* provides a measure of its executional feasibility within the constraints of market liquidity. Their inclusion thus enhances the explanatory power of the models without compromising their interpretability.

Table 8: Hypothesis 2 - Historic Multiple Ratio Approach (Coefficients)

Dependent Variable:	Tender Offer Probability			
	(1)	(2)	(3)	(4)
ln(EBITDA Multiple)	-0.4802			
	(0.895)			
ln(Net Income Multiple)		-0.5098		
		(0.334)		
ln(CFO Multiple)			0.0216	
			(0.681)	
ln(Price to Book Value Multiple)				-1.1793
				(0.872)
ln(Size)	0.1404	0.1239	0.0292	0.0965
	(0.266)	(0.273)	(0.253)	(0.234)
ln(Volume)	1.1474***	1.1585***	1.2061***	1.1334***
	(0.233)	(0.211)	(0.243)	(0.222)
ln(Liquidity)	0.2675	0.3416	-0.0076	0.2010
	(0.302)	(0.282)	(0.382)	(0.310)
ln(Equity Ratio)	2.0025**	1.9740**	1.8252**	2.0089**
	(0.585)	(0.612)	(0.585)	(0.594)
Dilution	-0.4811	2.4601	-2.7592	-0.1235
	(3.716)	(2.085)	(3.853)	(1.795)
ln(1 + Dividends)	-27.2398	-24.6936	-33.7171*	-36.5734
	(15.903)	(17.103)	(15.494)	(17.730)
Ownership	-0.2272	-0.5473	0.0354	-0.2334
	(1.077)	(1.100)	(1.089)	(1.005)
Constant	-3.6533*	-3.7610*	-4.3370*	-3.7470*
	(1.846)	(1.610)	(1.792)	(1.657)
Estimator	Logit	Logit	Logit	Logit
Observations	312	303	297	326
Log-Likelihood	-52.257	-47.799	-49.449	-54.790
Pseudo R^2	0.4453	0.4750	0.4236	0.4529

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3. *Notes:* This table presents the results of four logistic regressions, with the probability of a tender offer as the dependent variable and valuation multiples alongside control variables as independent variables. *Tender Offer Probability* is a binary variable equal to one if the firm announces a tender offer share repurchase, and zero if it announced a market repurchase programme. Log transformations are applied to all independent variables except *Dilution* and *Ownership*. The natural log (ln) is used to reduce skewness. Coefficients on log-transformed variables represent changes in the log-odds of the outcome associated with a 1% increase in the independent variable. *Dilution* and *Ownership* are included in their original scale. *Dividends* is transformed as *ln(1 + Dividends)* to accommodate zero values. Three of the four main variables—*EBITDA* (Earnings Before Interest, Tax, Depreciation and Amortization) *Multiple*, *Net Income Multiple*, and *CFO* (Cash-Flow from Operations) *Multiple*—represent the ratio of the current LTM valuation multiple to the average LTM valuation multiple over the past three years. Similarly, the *Price to Book Value Multiple* is defined as the ratio of the current market capitalisation to the current book value of equity, divided by the average of this multiple over the past three years. The control variables are defined as follows: *Size measures* the maximum announced repurchase volume relative to the company's market capitalisation; *Volume* captures the number of shares sought relative to the average daily trading volume; *Liquidity* is measured by the quick ratio; *Equity Ratio* is calculated as book equity divided by total assets; *Dilution* reflects the change in outstanding shares between the two most recent annual reports; and *Dividends* represents the company's average dividend yield over the three years preceding the repurchase announcement. *Ownership* denotes the percentage of publicly known owners of the company. For more detailed definitions of the variables, see Section 4.2. Robust standard errors of the type HC3 are used. Standard errors are shown in parentheses. The asterisks indicate two-sided significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The main variables for these logistic regressions have the following p-values: *ln(EBITDA Multiple)*: $p = 0.592$, *ln(Net Income Multiple)*: $p = 0.127$, *ln(CFO Multiple)*: $p = 0.975$, and *ln(Price to Book Value Multiple)*: $p = 0.176$.

6.2 Tender Offers Versus Short-Term Open Market Repurchases

Tables 13 and 14 further test the second hypothesis—that perceived undervaluation increases the probability of a tender offer—using the same model specifications as in the main analysis. In this robustness check, however, only OMR programmes with an announced duration of 90 days or less are included. The rationale for this restriction is that, under conditions of perceived undervaluation, firms may prefer to repurchase shares quickly—before the market corrects the mispricing—rather than commit to a longer programme, which may be less effective if the valuation gap is expected to close over time. Including longer OMR programmes could therefore distort the analysis by incorporating cases unlikely to reflect undervaluation-driven repurchase decisions.

Despite this adjustment, the results are broadly consistent with those from the full sample (Tables 8 to 11). None of the undervaluation proxies reaches conventional levels of statistical significance, with p-values ranging from 0.058 to 0.985. While *ln(Volume)* shows slightly reduced significance compared to the full sample, it remains the primary explanatory variable. In addition, *ln(Size)* is statistically significant at the $p < 0.05$ level in two models, and *ln(Equity Ratio)* is significant in one. These findings reinforce the earlier conclusion that the choice between tender offers and OMR programmes is not primarily driven by undervaluation, but rather by struc-

Table 9: Hypothesis 2 - Historic Multiple Ratio Approach (Marginal Effects)

Dependent Variable:	Tender Offer Probability			
	(1)	(2)	(3)	(4)
ln(EBITDA Multiple)	-0.0237			
	(0.043)			
ln(Net Income Multiple)		-0.0234		
		(0.015)		
ln(CFO Multiple)			0.0011	
			(0.033)	
ln(Price to Book Value Multiple)				-0.0589
				(0.042)
ln(Size)	0.0069	0.0057	0.0014	0.0048
	(0.013)	(0.013)	(0.012)	(0.012)
ln(Volume)	0.0567***	0.0532***	0.0585***	0.0567***
	(0.009)	(0.008)	(0.009)	(0.009)
ln(Liquidity)	0.0132	0.0157	-0.0004	0.0100
	(0.015)	(0.013)	(0.019)	(0.015)
ln(Equity Ratio)	0.0989**	0.0906**	0.0886**	0.1004**
	(0.030)	(0.030)	(0.032)	(0.030)
Dilution	-0.0238	0.1129	-0.1339	-0.0062
	(0.183)	(0.098)	(0.182)	(0.090)
ln(1 + Dividends)	-1.3456	-1.1333	-1.6367*	-1.8281
	(0.762)	(0.764)	(0.731)	(0.829)
Ownership	-0.0112	-0.0251	0.0017	-0.0117
	(0.053)	(0.050)	(0.053)	(0.050)
Estimator	Logit (ME)	Logit (ME)	Logit (ME)	Logit (ME)
Observations	312	303	297	326

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3. *Notes:* This table presents the average marginal effects from four logistic regressions, with the probability of a tender offer as the dependent variable and valuation multiples alongside control variables as independent variables. *Tender Offer Probability* is a binary variable equal to one if the firm announces a tender offer share repurchase, and zero if it announced a market repurchase programme. Log transformations are applied to all independent variables except *Dilution* and *Ownership*. The natural log (ln) is used to reduce skewness. The marginal difference for an independent variable is the average difference in the probability of the programme being a tender offer, that corresponds to a one unit difference in the independent variable. *Dividends* is transformed as *ln(1 + Dividends)* to accommodate zero values. Three of the four main variables—*EBITDA* (Earnings Before Interest, Tax, Depreciation and Amortization) *Multiple*, *Net Income Multiple*, and *CFO* (Cash-Flow from Operations) *Multiple*—represent the ratio of the current LTM valuation multiple to the average LTM valuation multiple over the past three years. Similarly, the *Price to Book Value Multiple* is defined as the ratio of the current market capitalisation to the current book value of equity, divided by the average of this multiple over the past three years. The control variables are defined as follows: *Size* measures the maximum announced repurchase volume relative to the company's market capitalisation; *Volume* captures the number of shares sought relative to the average daily trading volume; *Liquidity* is measured by the quick ratio; *Equity Ratio* is calculated as book equity divided by total assets; *Dilution* reflects the change in outstanding shares between the two most recent annual reports; and *Dividends* represents the company's average dividend yield over the three years preceding the repurchase announcement. *Ownership* denotes the percentage of publicly known owners of the company. Robust standard errors of the type HC3 are used. Standard errors are shown in parentheses. The asterisks indicate two-sided significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

tural and operational considerations.

6.3 Repurchase Duration and Subsequent Share Price Performance

If undervaluation were to lead to shorter repurchase programmes, firms announcing shorter open market repurchase programmes should exhibit abnormally positive share price performance as markets close the valuation gap. Table 15 presents descriptive statistics of these abnormal returns across various time horizons from 30 to 1,095 days (three years).

Abnormal returns are calculated as the difference between a stock's total return and the performance of the broad German CDAX index over each horizon, starting on the announcement date. To ensure adequate data coverage, only repurchase announcements made sufficiently prior to the dataset cut-off date of March 20, 2025 are included. As a result, observation counts decline with longer horizons.

The mean abnormal return is mostly positive across the windows, but the median is only positive for the 30-day OMR window and turns negative thereafter. This divergence reflects outliers with strong positive performance, evident in high maximum values. These are partly due to the post-COVID-19 rally, during which some stocks rebounded sharply from depressed levels.

Figure 8 illustrates share price developments following repurchase announcements over a 365-day period. The effect of outliers is especially evident in the mean abnormal return line, notably around the 150-day mark. Overall, the figure does not reveal a clear or consistent pattern in abnormal returns over time.

Table 16 reports five OLS regressions, with abnormal return as the dependent variable and *ln(Program Duration)* as

Table 10: Hypothesis 2 - Peer Valuation & Share Price Performance Approaches (Coefficients)

Dependent Variable:	Tender Offer Probability			
	(1)	(2)	(3)	(4)
ln(EBITDA Peer Valuation)	-0.8034*			
	(0.391)			
ln(Net Income Peer Valuation)		-0.7760		
		(0.695)		
ln(CFO Peer Valuation)			-0.0404	
			(0.306)	
Prior Share Performance				-0.0158
				(0.009)
ln(Size)	0.2826	0.2283	0.1787	0.1107
	(0.290)	(0.272)	(0.256)	(0.208)
ln(Volume)	0.9230**	1.0231***	0.8766***	1.1746***
	(0.265)	(0.254)	(0.246)	(0.228)
ln(Liquidity)	0.5184*	0.4334	0.2427	0.0981
	(0.261)	(0.272)	(0.364)	(0.318)
ln(Equity Ratio)	2.3531**	2.0481**	1.9214**	2.0315**
	(0.692)	(0.671)	(0.670)	(0.608)
Dilution	0.0423	1.9659	1.1867	0.2132
	(3.689)	(2.139)	(1.566)	(1.624)
ln(1 + Dividends)	-42.3661	-41.8897	-40.8611*	-42.8331*
	(22.263)	(22.133)	(20.614)	(18.145)
Ownership	-0.9093	-0.6418	-0.6962	-0.0625
	(1.154)	(1.077)	(0.999)	(1.011)
Constant	-0.6938	-1.7397	-2.0791	-3.7837*
	(2.207)	(2.057)	(1.945)	(1.483)
Estimator	Logit	Logit	Logit	Logit
Observations	277	264	257	324
Log-Likelihood	-45.640	-40.943	-46.175	-52.806
Pseudo R^2	0.4409	0.4758	0.3852	0.4592

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: This table presents the results of four logistic regressions, with the probability of a tender offer as the dependent variable and valuation multiples alongside control variables as independent variables. *Tender Offer Probability* is a binary variable equal to one if the firm announces a tender offer share repurchase, and zero if it announced a market repurchase programme. All independent predictors except *Dilution*, *Ownership* and *Prior Share Performance* are natural-log transformed to reduce skewness. Coefficients on log-transformed variables represent changes in the log-odds of the outcome associated with a 1% increase in the independent variable. *Dividends* is transformed as *ln(1 + Dividends)* to accommodate zero values. The main variables—*EBITDA* (Earnings Before Interest, Tax, Depreciation and Amortization) *Peer Valuation*, *Net Income Peer Valuation* and *CFO* (Cash-Flow from Operations) *Peer Valuation*—are the ratios of a firm's current LTM multiple to the peer-group average, while *Prior Share Performance* is the percentage change in share price over the 365 days before the announcement. The control variables are defined as follows: *Size* measures the maximum announced repurchase volume relative to the company's market capitalisation; *Volume* captures the number of shares sought relative to the average daily trading volume; *Liquidity* is measured by the quick ratio; *Equity Ratio* is calculated as book equity divided by total assets; *Dilution* reflects the change in outstanding shares between the two most recent annual reports; and *Dividends* represents the company's average dividend yield over the three years preceding the repurchase announcement. *Ownership* denotes the percentage of publicly known owners of the company. For more detailed definitions of the variables, see Section 4.2. Robust standard errors of the type HC3 are used. Standard errors are shown in parentheses. The asterisks indicate two-sided significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The main variables for these logistic regressions have the following p-values: *ln(EBITDA Peer Valuation)*: $p = 0.040$, *ln(Net Income Peer Valuation)*: $p = 0.264$, *ln(CFO Peer Valuation)*: $p = 0.895$, and *Prior Share Performance*: $p = 0.081$.

the main independent variable, plus three control variables. Only *ln(Liquidity)* is significant at the $p < 0.05$ level in one model. Additionally, the models exhibit very low explanatory power, as reflected in the low R^2 values.

These findings suggest that abnormal returns following repurchase announcements are not systematically related to programme duration. This aligns with the conclusions drawn in Chapter 5, which indicate that undervaluation does not result in shorter repurchase programmes. Furthermore, the control variables intended to capture the conditions under which firms announce repurchases do not exhibit a significant effect on abnormal returns, consistent with their role as non-value-adding logistical controls.

6.4 Announcement Return: Initial vs. Repeated Repurchase

This section investigates whether credibility potentially earned through prior repurchase programmes positively influences the abnormal return of new repurchase announcements.

The first row for both repurchase types in Table 15 reports *Abnormal Announcement Returns*, calculated over a three-trading-day event window. These returns are based exclusively on repurchase announcements made on or after July 3, 2018. Due to this 2 year shorter period, this subsample contains fewer observations than the one used for the 30-day abnormal return analysis.

To classify repurchase announcements as either initial or

Table 11: Hypothesis 2 - Peer Valuation & Share Price Performance Approaches (Marginal Effects)

Dependent Variable:	Tender Offer Probability			
	(1)	(2)	(3)	(4)
ln(EBITDA Peer Valuation)	-0.0389*			
	(0.019)			
ln(Net Income Peer Valuation)		-0.0343		
		(0.029)		
ln(CFO Peer Valuation)			-0.0021	
			(0.016)	
Prior Share Performance				-0.0008
				(0.000)
ln(Size)	0.0137	0.0101	0.0093	0.0054
	(0.015)	(0.012)	(0.014)	(0.010)
ln(Volume)	0.0447**	0.0452***	0.0456***	0.0571***
	(0.011)	(0.012)	(0.012)	(0.008)
ln(Liquidity)	0.0251*	0.0191	0.0126	0.0048
	(0.012)	(0.012)	(0.019)	(0.015)
ln(Equity Ratio)	0.1138**	0.0905**	0.1000**	0.0988**
	(0.034)	(0.030)	(0.038)	(0.030)
Dilution	0.0020	0.0868	0.0617	0.0104
	(0.179)	(0.099)	(0.082)	(0.079)
ln(1 + Dividends)	-2.0497	-1.8501	-2.1259*	-2.0832*
	(1.023)	(0.918)	(1.002)	(0.793)
Ownership	-0.0440	-0.0283	-0.0362	-0.0030
	(0.056)	(0.047)	(0.051)	(0.049)
Estimator	Logit (ME)	Logit (ME)	Logit (ME)	Logit (ME)
Observations	277	264	257	324

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3. *Notes:* This table presents the average marginal effects from four logistic regressions, with the probability of a tender offer as the dependent variable and valuation multiples alongside control variables as independent variables. *Tender Offer Probability* is a binary variable equal to one if the firm announces a tender offer share repurchase, and zero if it announced a market repurchase programme. Log transformations are applied to all independent variables except *Dilution*, *Ownership* and *Prior Share Performance*. The natural log (ln) is used to reduce skewness. The marginal difference for an independent variable is the average difference in the probability of the programme being a tender offer, that corresponds to a one unit difference in the independent variable. *Dividends* is transformed as *ln(1 + Dividends)* to accommodate zero values. Three of the four main variables—*EBITDA* (Earnings Before Interest, Tax, Depreciation and Amortization) *Peer Valuation*, *Net Income Peer Valuation*, and *CFO* (Cash-Flow from Operations) *Peer Valuation*—represent the ratio of a firm's current LTM valuation multiple to the average current multiple of its peer group. The fourth main independent variable: *Prior Share Performance* is the company's share price performance over the last 365 days before the announcement multiplied by 100. The control variables are defined as follows: *Size* measures the maximum announced repurchase volume relative to the company's market capitalisation; *Volume* captures the number of shares sought relative to the average daily trading volume; *Liquidity* is measured by the quick ratio; *Equity Ratio* is calculated as book equity divided by total assets; *Dilution* reflects the change in outstanding shares between the two most recent annual reports; and *Dividends* represents the company's average dividend yield over the three years preceding the repurchase announcement. *Ownership* denotes the percentage of publicly known owners of the company. Robust standard errors of the type HC3 are used. Standard errors are shown in parentheses. The asterisks indicate two-sided significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

repeated, a two-year look-back window from July 3, 2016 to July 3, 2018 is employed. Announcements made within this window are excluded to reduce the probability of misclassifying firms with regularly recurring repurchase activity as first-time repurchasers due to incomplete observable history. This adjustment helps mitigate classification bias, albeit at the expense of a reduced sample size.

However, this approach does not fully eliminate the risk of misclassification. Specifically, firms that conducted repurchase programmes prior to July 2016 and resumed them after July 3, 2018 are still categorised as initial repurchasers, owing to the absence of earlier data. This limitation arises from left-censoring in the dataset. While this constraint cannot be entirely resolved given the available data, the two-year look-back window represents a conservative compromise aimed at improving classification accuracy.

Table 17 tests whether the abnormal announcement return differs significantly depending on whether the repurchase programme is classified as repeated, as indicated by the binary variable *Repeated Repurchase*. The regression model also includes the standard set of control variables discussed previously.

The coefficient for *Repeated Repurchase* is statistically insignificant, with a p-value of 0.368, indicating no discernible effect on the abnormal announcement return. Similarly, none of the included control variables are statistically significant. Moreover, the model exhibits a very low R^2, reinforcing the conclusion that it possesses limited explanatory power in predicting abnormal announcement returns. These results suggest that markets do not appear to reward firms for a history of repurchase activity when a new programme is announced, implying that the signalling effect of credibility gained through past repurchases may be limited or already priced in.

Table 12: Multicollinearity Across OLS and Logit Models

Variable:	EBITDA Multiple	Net Income Multiple	CFO Multiple	Price to Book Value Multiple	EBITDA Peer Valuation	Net Income Peer Valuation	CFO Peer Valuation	Prior Share Performance
OLS Models								
ln(EBITDA Multiple)	1.016							
ln(Net Income Multiple)		1.023						
ln(CFO Multiple)			1.010					
ln(Price to Book Multiple)				1.047				
ln(EBITDA Peer Valuation)					1.106			
ln(Net Income Peer Valuation)						1.078		
ln(CFO Peer Valuation)							1.143	
Prior Share Performance								1.021
ln(Size)	5.044	4.091	5.173	4.097	4.756	3.835	3.808	4.088
ln(Volume)	4.979	3.982	5.226	4.124	4.653	3.733	3.750	4.131
ln(Liquidity)	1.177	1.137	1.241	1.153	1.171	1.111	1.241	1.153
ln(Equity Ratio)	1.276	1.245	1.299	1.304	1.281	1.216	1.243	1.292
Dilution	1.088	1.094	1.050	1.033	1.098	1.105	1.047	1.031
ln(1 + Dividends)	1.181	1.170	1.152	1.176	1.249	1.248	1.185	1.178
Logit Models								
ln(EBITDA Multiple)	1.016							
ln(Net Income Multiple)		1.032						
ln(CFO Multiple)			1.011					
ln(Price to Book Value Multiple)				1.118				
ln(EBITDA Peer Valuation)					1.103			
ln(Net Income Peer Valuation)						1.062		
ln(CFO Peer Valuation)							1.131	
Prior Share Performance								1.037
ln(Size)	4.394	3.672	4.608	3.835	4.205	3.375	3.537	3.814
ln(Volume)	4.292	3.594	4.651	3.813	4.076	3.414	3.544	3.819
ln(Liquidity)	1.155	1.129	1.260	1.150	1.134	1.084	1.238	1.144
ln(Equity Ratio)	1.309	1.277	1.348	1.362	1.313	1.263	1.302	1.339
Dilution	1.087	1.046	1.049	1.079	1.097	1.053	1.036	1.044
ln(1 + Dividends)	1.200	1.192	1.181	1.222	1.271	1.251	1.223	1.228
Ownership	1.067	1.086	1.053	1.096	1.063	1.063	1.059	1.088

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: The table reports Variance Inflation Factors (VIFs) for all independent variables across the 16 model specifications. All VIF values are below the threshold of 6, suggesting that multicollinearity is not a serious concern in any of the models. The intercept term is excluded from the table for clarity. For detailed definitions of the variables, see Table 4 or Section 4.2.

Table 13: Tender Offer vs. Short Term Open Market Repurchase - Historic Multiple Ratio Approach (Coefficients)

Dependent Variable:	Tender Offer Probability			
	(1)	(2)	(3)	(4)
ln(EBITDA Multiple)	-0.1473			
	(1.575)			
ln(Net Income Multiple)		-0.2005		
		(0.881)		
ln(CFO Multiple)			0.2317	
			(0.990)	
ln(Price to Book Value Multiple)				-3.3219
				(2.029)
ln(Size)	1.0861*	0.8088	0.9662	0.6277
	(0.529)	(0.602)	(0.564)	(0.389)
ln(Volume)	1.0061***	1.1478**	1.0529**	1.3563**
	(0.279)	(0.343)	(0.347)	(0.393)
ln(Liquidity)	-0.5304	-0.2595	-0.7015	-0.2801
	(0.860)	(0.769)	(0.931)	(0.794)
ln(Equity Ratio)	3.1254	3.0370	3.0294	1.7709
	(1.648)	(1.679)	(1.841)	(1.185)
Dilution	14.7811	17.2954	15.5674	1.0362
	(8.032)	(9.345)	(10.175)	(11.402)
ln(1+ Dividends)	-24.9041	-10.6636	-22.3816	-11.2204
	(29.037)	(37.114)	(33.867)	(29.568)
Ownership	-1.4744	-1.0595	-1.6821	-1.5190
	(1.510)	(1.586)	(1.876)	(1.259)
Constant	4.1643	1.7475	3.6194	-0.6197
	(3.408)	(4.318)	(3.432)	(3.305)
Estimator	Logit	Logit	Logit	Logit
Observations	140	134	134	149
Log-Likelihood	-20.210	-18.867	-19.680	-20.629
Pseudo R^2	0.6923	0.7005	0.6711	0.7074

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: This table presents the results of four logistic regressions, with the probability of a tender offer as the dependent variable and valuation multiples alongside control variables as independent variables. *Tender Offer Probability* is a binary variable equal to one if the firm announces a tender offer share repurchase, and zero if it announced a market repurchase programme. Open market repurchase programmes are only included if their duration is shorter than or equal to 90 days. Log transformations are applied to all independent variables except *Dilution* and *Ownership*. The natural log (ln) is used to reduce skewness. Coefficients on log-transformed variables represent changes in the log-odds of the outcome associated with a 1% increase in the independent variable. *Dilution* and *Ownership* are included in their original scale. *Dividends* is transformed as *ln(1 + Dividends)* to accommodate zero values. For detailed definitions of the variables, see Table 4 or Section 4.2. Robust standard errors of the type HC3 are used. Standard errors are shown in parentheses. The asterisks indicate two-sided significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The main variables for these logistic regressions have the following p-values: *ln(EBITDA Multiple)*: $p = 0.925$, *ln(Net Income Multiple)*: $p = 0.820$, *ln(CFO Multiple)*: $p = 0.815$, *ln(Price to Book Value Multiple)*: $p = 0.102$.

7 Limitations and Future Research

This chapter discusses the key limitations of the thesis and their impact on the results. Additionally, it offers suggestions for future research.

7.1 Limitations

1. Choice of variables: The primary limitation of this research concerns the validity of the main independent variables intended to measure undervaluation. In this study, publicly available information is used to infer whether management perceives the firm's shares to be undervalued at the time of the repurchase announcement. However, the multiple-based and peer valuation ratios may be outdated at the point of the repurchase decision—up to six months old for firms reporting semi-annually, or approximately 90 days old for quarterly reporters. By contrast, management has access to current internal data and superior foresight regarding future business performance. As a result, all key valuation metrics employed here are inherently backward-looking, despite the fact that forward-looking expectations—especially regarding future earnings and cash flows—are more relevant in assessing true valuation. Additionally, the historic multiple variables may be misleading in scenarios where, for instance, a firm's growth rate has declined, making a lower multiple justifiable compared to historical levels. Similarly, the peer valuation approach is based solely on primary industry classification, which overlooks other crucial comparative factors such as size, growth trajectory, margins, and specific business models. While the industry classifications used are de-

Table 14: Tender Offer vs. Short Term Open Market Repurchase - Peer Valuation & Share Price Performance Approaches (Coefficients)

Dependent Variable:	Tender Offer Probability			
	(1)	(2)	(3)	(4)
ln(EBITDA Peer Valuation)	-1.6458			
	(0.867)			
ln(Net Income Peer Valuation)		-0.6424		
		(0.616)		
ln(CFO Peer Valuation)			-0.0111	
			(0.607)	
Prior Share Performance				-0.0251
				(0.016)
ln(Size)	1.3315	0.8118	0.9304	0.6483*
	(0.704)	(0.618)	(0.546)	(0.318)
ln(Volume)	0.8121*	1.0839**	0.9122**	1.3095***
	(0.321)	(0.350)	(0.306)	(0.300)
ln(Liquidity)	-0.3838	-0.0972	-0.2416	-0.3162
	(0.882)	(0.932)	(0.782)	(0.753)
ln(Equity Ratio)	3.4638*	2.4671	2.4807	1.8260
	(1.533)	(1.619)	(2.079)	(1.325)
Dilution	19.1307	19.9882	13.6754	-3.2008
	(10.573)	(12.066)	(9.619)	(2.843)
ln(1+ Dividends)	-53.0128	-19.9751	-27.3065	-15.9716
	(35.736)	(34.271)	(33.306)	(31.099)
Ownership	-1.5951	-1.7113	-1.7937	-1.2171
	(1.434)	(1.604)	(1.876)	(1.535)
Constant	5.8742	1.5620	3.4764	-0.1306
	(4.207)	(3.916)	(3.750)	(2.245)
Estimator	Logit	Logit	Logit	Logit
Observations	122	111	109	147
Log-Likelihood	-17.151	-17.287	-18.914	-20.635
Pseudo R^2	0.7021	0.6789	0.6360	0.6992

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: This table presents the results of four logistic regressions, with the probability of a tender offer as the dependent variable and valuation multiples alongside control variables as independent variables. *Tender Offer Probability* is a binary variable equal to one if the firm announces a tender offer share repurchase, and zero if it announced a market repurchase programme. Open market repurchase programmes are only included if their duration is shorter than or equal to 90 days. Log transformations are applied to all independent variables except *Dilution*, *Ownership* and *Prior Share Performance*. The natural log (ln) is used to reduce skewness. Coefficients on log-transformed variables represent changes in the log-odds of the outcome associated with a 1% increase in the independent variable. *Dilution*, *Ownership* and *Prior Share Performance* are included in their original scale. *Dividends* is transformed as *ln(1 + Dividends)* to accommodate zero values. For detailed definitions of the variables, see Table 4 or Section 4.2. Robust standard errors of the type HC3 are used. Standard errors are shown in parentheses. The asterisks indicate two-sided significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The main variables for these logistic regressions have the following p-values: *ln(EBITDA Peer Valuation)*: $p = 0.058$, *ln(Net Income Peer Valuation)*: $p = 0.297$, *ln(CFO Peer Valuation)*: $p = 0.985$, *Prior Share Performance*: $p = 0.119$.

fined narrowly to ensure firms operate in similar domains, this does not guarantee true comparability from the perspective of investors.

The variable *Prior Share Performance* presents further challenges. Macroeconomic events—such as the COVID-19 pandemic—may have significantly altered firms' business prospects in ways accurately reflected in their stock prices. As a result, price movements may coincide with management's internal valuation estimates, potentially obscuring signs of undervaluation. Additionally, firms may experience substantial improvements in intrinsic value without corresponding increases in share price, leading to undervaluation not captured by this variable.

Due to these challenges, accurately measuring either true undervaluation or management's perception thereof is inherently difficult. However, the use of the aforementioned variables represents standard academic practice. Each of these proxies has individual limitations, but in combination they likely provide a realistic approximation of undervaluation. This is particularly relevant given the uniform findings across regression analyses, which consistently show that undervaluation does not influence either the duration of repurchase programmes or the choice between open market repurchases and tender offers.

Some control variables in this study are designed to capture alternative motivations for repurchases. However, due to measurement limitations, they may only approximate their intended effects. For instance, publicly disclosed ownership

Table 15: Descriptive Statistics of Abnormal Returns Following the Announcement

	Obs	Mean	Std. Dev	Min.	25%	Median	75%	Max.
Open Market Repurchase (n = 296)								
Abnormal Announcement Return (%)	257	0.495	4.026	-17.189	-1.470	0.367	2.393	21.274
Abnormal Return Over 30 Days (%)	296	1.139	9.559	-46.715	-3.631	0.938	5.349	38.869
Abnormal Return Over 180 Days (%)	285	3.248	60.162	-89.920	-11.857	-2.476	10.223	923.787
Abnormal Return Over 365 Days (%)	273	2.060	42.158	-106.549	-20.126	-2.322	17.404	336.166
Abnormal Return Over 730 Days (%)	227	0.239	56.094	-93.402	-35.371	-9.601	22.451	306.881
Abnormal Return Over 1095 Days (%)	184	1.093	75.896	-107.888	-42.184	-10.806	25.946	493.664
Tender Offer (n = 28)								
Abnormal Announcement Return (%)	25	0.931	3.761	-6.349	-1.238	0.433	1.896	10.431
Abnormal Return Over 30 Days (%)	28	1.667	9.192	-12.395	-5.375	-0.231	7.757	25.785
Abnormal Return Over 180 Days (%)	28	-5.635	20.726	-48.307	-17.814	-6.109	5.703	30.996
Abnormal Return Over 365 Days (%)	23	3.707	42.896	-51.463	-27.845	-3.458	20.926	146.346
Abnormal Return Over 730 Days (%)	18	17.039	88.013	-85.973	-30.379	-13.031	46.043	279.958
Abnormal Return Over 1095 Days (%)	13	-2.042	75.740	-88.582	-48.205	-13.099	4.496	198.181

Source: The presented data is obtained from Capital IQ, LSEG Datastream and is partly hand-collected. For more details, see Section 4.3.
Notes: This table presents descriptive statistics for abnormal share price performance following share repurchase announcements. All return variables are rescaled to percentage-point units by multiplying the raw values by 100. Abnormal returns are computed as the difference between a stock's total return and the corresponding performance of the CDAX index over the specified time horizon, beginning on the announcement date. To ensure sufficient data coverage, abnormal return measures include only those announcements made sufficiently prior to the dataset cut-off date of March 20, 2025. The *Abnormal Announcement Return* (3-day window) is calculated using only announcements dated July 3, 2018 or later. This restriction ensures sufficient data to determine whether a repurchase programme was repeated and aligns with the focus of Table 17 on repeated repurchases. Announcements from the two-year period between July 3, 2016 and July 3, 2018 are excluded for this reason.

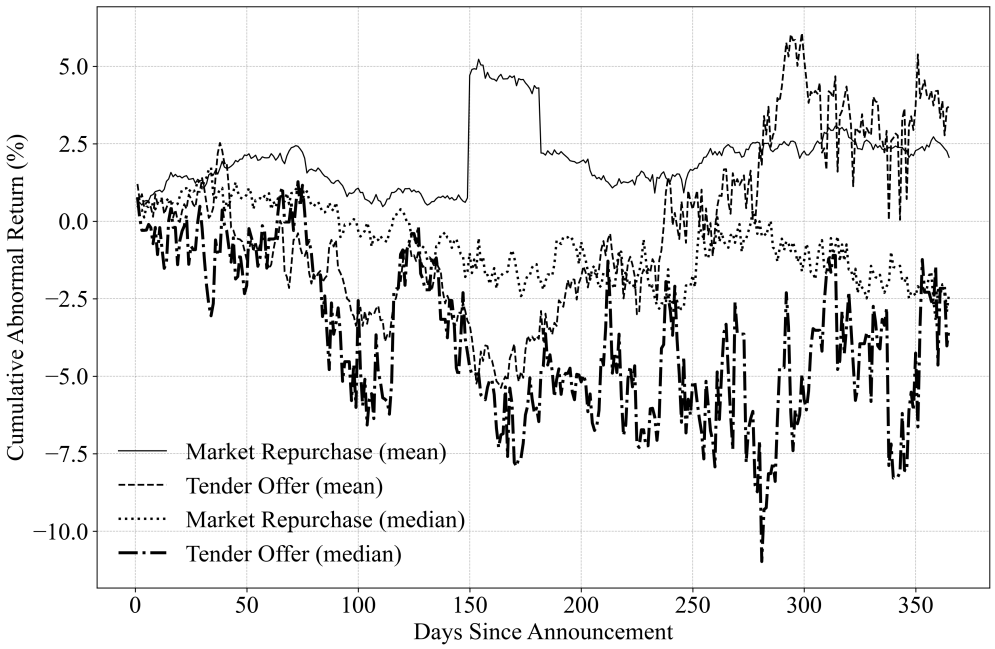


Figure 8: Abnormal Returns for 365 Days following the Repurchase Announcement

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: The figure presents the cumulative abnormal returns of companies' stock prices over a 365-day period following the announcement of share repurchase programmes, categorized by repurchase method. The X-axis indicates the number of days since the repurchase announcement, while the Y-axis shows the cumulative abnormal return in percentage terms. The graph compares the mean and median cumulative abnormal returns for two types of repurchase mechanisms: market repurchases and tender offers. Mean returns are represented by solid (market repurchase) and dashed (tender offer) lines, while median returns are shown using dotted (market repurchase) and dash-dotted (tender offer) lines.

data may not perfectly reflect true ownership concentration. Similarly, the *Dilution* variable attempts to capture a combination of factors—including stock option schemes and the use of equity in M&A—but also includes the offsetting effect

Table 16: Abnormal Return by Holding Period Dependent on the Programme Duration

Dependent Variable:	Abnormal Return (%)				
	(30 Days)	(180 Days)	(365 Days)	(730 Days)	(1095 Days)
	(1)	(2)	(3)	(4)	(5)
ln(Programme Duration)	-0.5208	-4.7893	-1.6003	-4.2396	-5.0443
	(0.535)	(6.944)	(2.820)	(5.067)	(7.373)
ln(Size)	0.4296	1.5354	-0.4718	5.4350	7.0984
	(0.611)	(3.062)	(3.140)	(5.835)	(6.693)
ln(Volume)	0.3719	0.0501	3.4057	1.7624	3.5712
	(0.554)	(1.752)	(2.957)	(4.648)	(5.376)
ln(Liquidity)	0.6442	-2.4020	-4.3556*	-4.1785	-2.4801
	(0.341)	(2.942)	(2.004)	(3.875)	(4.475)
Constant	5.1202	32.1211	0.5700	42.4668	52.8289
	(4.833)	(49.360)	(24.692)	(48.660)	(60.048)
Estimator	OLS	OLS	OLS	OLS	OLS
Observations	295	284	272	226	183
R^2	0.019	0.011	0.021	0.032	0.045
Adjusted R^2	0.006	-0.003	0.006	0.014	0.023

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: This table presents the results of five OLS regressions. They differ in the time span for which the dependent variable was observed. The abnormal return is calculated by subtracting the overall performance of the CDAX index from the overall share price performance of the stock in the relevant time period starting immediately after the announcement multiplied by 100. Log transformations are applied to all independent variables. The natural log (ln) is used to reduce skewness. Coefficients on log-transformed independent variables indicate that the dependent variable is approximately β / 100 units higher, on average, for a 1% increase in the independent variable. The main independent variable of the models is the variable *Programme Duration* which is the amount of days between the announced start of the repurchase programme and its announced end date. The control variables are defined as follows: *Size* measures the maximum monetary value of the repurchase programme relative to the company's market capitalisation; *Volume* represents the ratio of the number of shares sought to the average daily trading volume; *Liquidity* corresponds to the quick ratio. For more detailed definitions of the variables, see Section 4.2. Robust standard errors of the type HC1 are used. Standard errors are shown in parentheses. The asterisks indicate two-sided significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The main variable *ln(Programme Duration)* for these OLS regressions has the following p-values: (1): $p = 0.331$, (2): $p = 0.490$, (3): $p = 0.570$, (4): $p = 0.403$,(5): $p = 0.494$

of past repurchases, potentially complicating interpretation.

2. Abnormal returns: To calculate abnormal returns, the CDAX index is used as a benchmark. This index is the most appropriate choice as it encompasses all publicly listed German companies. However, it is market capitalisation-weighted, which results in an under-representation of small-cap firms—many of which feature prominently in the sample. This may bias abnormal return calculations, given that larger firms outperformed smaller ones during the sample period (Deutsche Börse AG, 2025a, 2025b).

3. Research setting: The restriction of this study to German firms limits the generalisability of the findings to other jurisdictions, particularly those with different regulatory frameworks or managerial cultures. Nonetheless, this focus contributes to a relatively homogenous and unbiased research setting, reducing confounding variation and increasing internal validity.

4. Subsequent announcements: The analysis does not account for official increases to repurchase programmes that may be announced subsequent to the initial programme declaration. These follow-up announcements can influence market perceptions and, in particular, may affect the abnormal returns following the initial announcement.

5. Calendar year: Unlike standard academic practice, this study relies on Capital IQ's calendar year classification rather than fiscal years. This decision was necessitated by frequent inconsistencies or unavailability of fiscal year data for smaller German firms. This choice is unlikely to materially affect the results for two reasons. First, the fiscal and calendar years align for the majority of firms in the sample. Second, Capital IQ's classification reallocates fiscal year data to the appropriate calendar quarter if the fiscal year ends within that quarter, thereby preserving temporal accuracy. More broadly, using calendar-based data was essential for aligning financial and stock market data, which is critical for producing consistent and interpretable results.

6. Sample Concentration in frequent repurchasers: As illustrated in Figure 4, a significant share of repurchase announcements originate from a relatively small number of firms. These firms appear multiple times in the sample, meaning their specific characteristics exert a disproportionate influence on the findings. This may be especially problematic if their repurchase conditions remain constant across announcements and differ from the sample median. Furthermore, repeated announcements by the same firm on the same date each year may have contributed to the clusters in the distributional pattern observed in Figure 2.

7. Class imbalance in logistic models: In the logistic regressions testing the second hypothesis (Tables 8–11), a notable class imbalance exists: OMRs occur approximately ten

Table 17: Announcement Return: First or Repeated Repurchase

Dependent Variable:	Abnormal Announcement Return (3 Days, %)
	(1)
Repeated Repurchase	0.4467
	(0.496)
ln(Size)	0.1722
	(0.299)
ln(Volume)	–0.0450
	(0.241)
ln(Liquidity)	–0.1754
	(0.197)
Constant	1.2246
	(2.017)
Estimator	OLS
Observations	257
R^2	0.008
Adj. R^2	–0.007

Source: Data sourced from LSEG Datastream, Capital IQ and supplemented with hand-collected information. For further details, refer to Section 4.3.
Notes: This table presents the results of one OLS regression. The *Abnormal Announcement Return* is calculated by subtracting the overall performance of the CDAX index from the overall share price performance of the stock in the in the three trading days immediately after the announcement multiplied by 100. Log transformations are applied the independent variables *Size*, *Volume* and *Liquidity*. The natural log (ln) is used to reduce skewness. Coefficients on log-transformed independent variables indicate that the dependent variable is approximately β / 100 units higher, on average, for a 1% increase in the independent variable. The variables *Abnormal Announcement Return* and *Repeated Repurchase* are included in their original form. The main independent variable in the models is the binary indicator *Repeated Repurchase*, which equals one if the company has previously announced repurchase programs and zero otherwise. Its effect can be interpreted as follows: a repeated repurchase, on average and holding other factors constant, is associated with a 0.4467 percentage point higher announcement return compared to an initial repurchase. However, the *Repeated Repurchase* has the following p-value: (1): $p = 0.368$ and is therefore not statistically significant. The sample period runs from July 3, 2018 through the end of the dataset; any repurchase announcements before July 3, 2018 are used solely to identify whether a firm is a "repeat" repurchaser and are then excluded from the estimation sample. Among the 257 repurchase announcements analysed in this regression, 89 are classified as repeated repurchases, while 168 are initial repurchases. The control variables are defined as follows: *Size* measures the maximum monetary value of the repurchase programme relative to the company's market capitalisation; *Volume* represents the ratio of the number of shares sought to the average daily trading volume; *Liquidity* corresponds to the quick ratio. For more detailed definitions of the variables, see Section 4.2. Robust standard errors of the type HC1 are used. Standard errors are shown in parentheses. The asterisks indicate two-sided significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

times more frequently than tender offers. This raises concerns about potential bias in model estimation, favouring the majority class. To assess the robustness of these results, comparisons are made with additional logistic models presented in Tables 13 and 14, where the imbalance is less pronounced (a ratio of roughly four to one). As the substantive findings are consistent across both sets of models, it appears unlikely that the class imbalance significantly distorts the results.

7.2 Future Research

Several areas offer promising avenues for further investigation. One notable gap is the lack of a comprehensive overview of global repurchase regulations and their influence on the use and structure of such programmes. A systematic cross-country comparison of legal frameworks would enhance understanding of share repurchases as a global phenomenon and improve the design of future research settings.

Further work on the duration of repurchase programmes could also be valuable. Studies in other countries or cultural contexts may clarify the factors influencing programme length. Additionally, using alternative proxies for undervaluation or incorporating qualitative approaches such as management interviews could help address some limitations highlighted in this study—especially those linked to the indirect nature of valuation metrics.

The relationship between capital structure and repurchase duration merits closer attention, as it was the only motivation-related variable to show statistical significance in this analysis. A more targeted investigation into how repurchase strategy aligns with capital structure goals could yield useful insights.

Lastly, future research should explore the regulation and adoption of accelerated share repurchase programmes outside the United States. ASRs appear to shift the dynamics of timing and signalling in share buybacks. Understanding their use and regulatory treatment across countries would help build a more complete picture of global repurchase strategies.

8 Conclusion

Drawing on signalling theory and evidence from disclosure-based regulatory environments, this thesis investigates whether German firms use the duration of announced share repurchase programmes to signal undervaluation. I construct eight market-based undervaluation proxies. These

include four measures comparing current valuation multiples to historical averages, three relative to industry peers, and one based on prior share price performance. The analysis is using in a dataset of 329 repurchase announcements between July 3, 2016 and March 20, 2025, covering 299 open market repurchases and 30 tender offers.

Despite the theoretical plausibility of a link between perceived undervaluation and repurchase programme design, the empirical evidence from my dataset does not support this relationship. Across eight OLS regressions, none of the undervaluation proxies show statistical significance. Their p-values remain far from conventional thresholds, and the direction of coefficients is inconsistent. This suggests that firms do not shorten programme duration in response to undervaluation, contrary to the signalling interpretation.

Instead, the duration of buyback programmes appears primarily influenced by two operational factors: *Size* and *Volume*. Both variables are highly statistically significant across all OLS regressions. These results indicate that logistical constraints and execution feasibility play a more substantial role than valuation considerations in determining programme length.

The second hypothesis, which posits that undervaluation increases the probability of a tender offer relative to an OMR, is also not supported by the data. Only one undervaluation proxy—peer-relative *EBITDA Valuation*—shows marginal significance in a single logistic regression. The remaining measures do not meaningfully affect the choice of repurchase method. As with programme duration, *Volume* again emerges as the most important predictor: companies are more likely to use tender offers when the intended number of shares to be repurchased is high in comparison to daily trading volumes, making market-based execution impractical.

Robustness checks further support the validity of the main findings. Restricting the sample to tender offers and short-term open market repurchases ($\leq$ 90 days) does not alter the core results. Additional analyses examining post-announcement stock performance—over horizons ranging from 30 days to three years—reveal no evidence that programme duration influences subsequent share price development. Furthermore, the data does not support the notion that repeated repurchase announcements yield higher abnormal announcement returns, which would be expected if market participants viewed such programmes as more credible.

In summary, this study finds no evidence that firms in Germany use either repurchase duration or method as a signal of perceived undervaluation. Instead, programme design reflects operational and financial constraints—particularly the scale of the intended buyback and prevailing market liquidity. These insights challenge the assumption that repurchase announcements inherently convey informational value about firm valuation and suggest that, in the German context, repurchase structure is driven more by implementation practicality than by signalling strategy.

References

Aktiengesetz §71. (1998). §71 Erwerb Eigener Aktien. https://www.gesetze-im-internet.de/aktg/__71.html

Akyol, A. C., & Foo, C. C. (2013). Share Repurchase Reasons and the Market Reaction to Actual Share Repurchases: Evidence from Australia. *International Review of Finance*, *13*(1), 1–37. https://doi.org/https://doi.org/10.1111/j.1468-2443.2012.01159.x

Andersson, T., Haslam, C., Lee, E., & Tsitisianis, N. (2007). Financialized Accounts: Share Buy-Backs, Mark to Market and Holding the Financial Line in the S&P 500. *Accounting Forum*, *31*(2), 165–178. https://doi.org/https://doi.org/10.1016/j.accfor.2006.11.002

Andres, C., Betzer, A., Doumet, M., & Theissen, E. (2016). Open Market Share Repurchases in Germany: A Conditional Event Study Approach. *Abacus*, *54*(4), 417–444. https://doi.org/https://doi.org/10.1111/abac.12094

Andriosopoulos, D., & Lasfer, M. (2015). The Market Valuation of Share Repurchases in Europe [Global Governance and Financial Stability]. *Journal of Banking and Finance*, *55*, 327–339. https://doi.org/https://doi.org/10.1016/j.jbankfin.2014.04.017

Banyi, M. L., Dyl, E. A., & Kahle, K. M. (2008). Errors in Estimating Share Repurchases. *Journal of Corporate Finance*, *14*(4), 460–474. https://doi.org/https://doi.org/10.1016/j.jcorpfin.2008.06.004

Bargeron, L., Bonaime, A., & Thomas, S. (2017). The Timing and Source of Long-Run Returns Following Repurchases. *The Journal of Financial and Quantitative Analysis*, *52*(2), 491–517. https://doi.org/https://www.jstor.org/stable/26164607

Bargeron, L., Kulchania, M., & Thomas, S. (2011). Accelerated Share Repurchases. *Journal of Financial Economics*, *101*(1), 69–89. https://doi.org/https://doi.org/10.1016/j.jfineco.2011.02.004

Ben-Rephael, A., Oded, J., & Wohl, A. (2014). Do Firms Buy Their Stock at Bargain Prices? Evidence from Actual Stock Repurchase Disclosures. *Review of Finance*, *18*(4), 1299–1340. https://doi.org/https://doi.org/10.1093/rof/rft028

Bens, D. A., Nagar, V., Skinner, D. J., & Wong, M. H. F. (2003). Employee Stock Options, EPS Dilution, and Stock Repurchases. *Journal of Accounting and Economics*, *36*(1-3), 51–90. https://doi.org/https://doi.org/10.1016/j.jacceco.2003.10.006

Billett, M. T., & Xue, H. (2007). The Takeover Deterrent Effect of Open Market Share Repurchases. *The Journal of Finance*, *62*, 1827–1850. https://doi.org/https://doi.org/10.1111/j.1540-6261.2007.01258.x

Bonaimé, A. A., Öztekin, Ö., & Warr, R. S. (2014). Capital Structure, Equity Mispricing, and Stock Repurchases. *Journal of Corporate Finance*, *26*, 182–200. https://doi.org/https://doi.org/10.1016/j.jcorpfin.2014.03.007

Brav, A., Graham, J. R., Harvey, C. R., & Michaely, R. (2005). Payout Policy in the 21st Century. *Journal of Financial Economics*, *77*(3), 483–527. https://doi.org/https://doi.org/10.1016/j.jfineco.2004.07.004

Chan, K., Ikenberry, D., & Lee, I. (2004). Economic Sources of Gain in Stock Repurchases. *Journal of Financial and Quantitative Analysis*, *39*(3), 461–479. https://doi.org/https://doi.org/10.1017/S0022109000003987

Chan, K., Ikenberry, D. L., & Lee, I. (2007). Do Managers Time the Market? Evidence from Open-Market Share Repurchases. *Journal of Banking & Finance*, *31*(9), 2673–2694. https://doi.org/https://doi.org/10.1016/j.jbankfin.2006.09.017

Chan, K., Ikenberry, D. L., Lee, I., & Wang, Y. (2010). Share Repurchases as a Potential Tool to Mislead Investors. *Journal of Corporate Finance*, *16*(2), 137–158. https://doi.org/https://doi.org/10.1016/j.jcorpfin.2009.10.003

Cheng, Y., Harford, J., & Zhang, T. (2015). Bonus-Driven Repurchases. *Journal of Financial and Quantitative Analysis*, *50*(3), 447–475. https://doi.org/https://doi.org/10.1017/S0022109015000149

Comment, R., & Jarrell, G. A. (1991). The Relative Signalling Power of Dutch-Auction and Fixed-Price Self-Tender Offers and Open-Market Share Repurchases. *The Journal of Finance*, *46*(4), 1243–1271. https://doi.org/https://doi.org/10.1111/j.1540-6261.1991.tb04617.x

Cook, D. O., & Zhang, W. (2022). CEO Option Incentives and Corporate Share Repurchases. *International Review of Economics & Finance*, *78*, 355–376. https://doi.org/https://doi.org/10.1016/j.iref.2021.12.002

Covestro AG. (2022). Transcript Q4/FY 2021 Earnings Call. Retrieved May 10, 2025, from https://www.covestro.com/en/investors/-/media/E5037DB2A68C4E858B8E867434C35C1B.ashx

Deutsche Börse AG. (2017). Deutsche Börse AG: Implementation Period of Share Buyback Program Can Extend Until End of 1st Half-Year 2018. Retrieved May 10, 2025, from https://www.deutsche-boerse.com/dbg-en/investor-relations/announcements-and-services/ad-hoc-announcements/Deutsche-B-rse-AG-Implementation-period-of-share-buyback-program-can-extend-until-end-of-1st-half-year-2018-4036650

Deutsche Börse AG. (2025a). CDAX-Performance Price-History. Retrieved June 7, 2025, from https://www.boerse-frankfurt.de/indices/cdax-performance/price-history/historical-prices-and-volumes

Deutsche Börse AG. (2025b). DAX Price-History. Retrieved June 7, 2025, from https://www.boerse-frankfurt.de/indices/dax/price-history/historical-prices-and-volumes

Dharmawan, G. V., & Mitchell, J. D. (2007). Incentives for On-Market Buy-Backs: Evidence from a Transparent Buy-Back Regime. *Journal of Corporate Finance*, *13*(1), 146–169. https://doi.org/https://doi.org/10.1016/j.jcorpfin.2006.12.002

Dittmann, I., Li, A. Y., Obernberger, S., & Zheng, J. (2022). *Equity-Based Compensation and the Timing of Share Repurchases: The Role of the Corporate Calendar*, Tinbergen Institute Discussion Paper 21-101/IV, Forthcoming. https://doi.org/http://dx.doi.org/10.2139/ssrn.4084794

Dittmar, A., & Field, L. C. (2015). Can Managers Time the Market? Evidence Using Repurchase Price Data. *Journal of Financial Economics*, *115*(2), 261–282. https://doi.org/https://doi.org/10.1016/j.jfineco.2014.09.007

European Parliament and the Council of the European Union. (2014). REGULATION (EU) No 596/2014 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. https://eur-lex.europa.eu/eli/reg/2014/596/oj/eng

Faccio, M., & Masulis, R. W. (2005). The Choice of Payment Method in European Mergers and Acquisitions. *The Journal of Finance*, *60*, 1345–1388. https://doi.org/https://doi.org/10.1111/j.1540-6261.2005.00764.x

Gould, G. P. (2019). Repurchases and Intended Program Length. *Pacific-Basin Finance Journal*, *56*, 234–247. https://doi.org/https://doi.org/10.1016/j.pacfin.2019.05.011

Grullon, G., & Michaely, R. (2002). Dividends, Share Repurchases, and the Substitution Hypothesis. *The Journal of Finance*, *57*(4), 1649–1684. https://doi.org/http://www.jstor.org/stable/3094520

Grullon, G., & Ikenberry, D. L. (2005). What Do We Know About Stock Repurchases? *Journal of Applied Corporate Finance*, *13*(1), 1078–1196. https://doi.org/https://doi.org/10.1111/j.1745-6622.2000.tb00040.x

Ikenberry, D., Lakonishok, J., & Vermaelen, T. (1995). Market Underreaction to Open Market Share Repurchases. *Journal of Financial Economics*, *39*(2), 181–208. https://doi.org/https://doi.org/10.1016/0304-405X(95)00826-Z

Ikenberry, D., & Vermaelen, T. (1996). The Option to Repurchase Stock. *Financial Management*, *25*(4), 9–24. https://www.jstor.org/stable/3665586

Jagannathan, M., Stephens, C. P., & Weisbach, M. S. (2000). Financial Flexibility and the Choice Between Dividends and Stock Repurchases. *Journal of Financial Economics*, *57*(3), 355–384. https://doi.org/https://doi.org/10.1016/S0304-405X(00)00061-1

Jenter, D. (2005). Market Timing and Managerial Portfolio Decisions. *The Journal of Finance*, *60*(4), 1903–1949. https://doi.org/https://doi.org/10.1111/j.1540-6261.2005.00783.x

Kahle, K. M. (2002). When a Buyback Isn't a Buyback: Open Market Repurchases and Employee Options. *Journal of Financial Economics*, *63*, 235–261. https://doi.org/https://doi.org/10.1016/S0304-405X(01)00095-2

Kamma, S., Kanatas, G., & Raymar, S. (1992). Dutch Auction Versus Fixed-Price Self-Tender Offers for Common Stock. *Journal of Financial Intermediation*, *2*(3), 277–307. https://doi.org/https://doi.org/10.1016/1042-9573(92)90003-V

Lee, C. I., Ejara, D. D., & Gleason, K. C. (2010). An Empirical Analysis of European Stock Repurchases. *Journal of Multinational Financial Management*, *20*(2-3), 114–125. https://doi.org/https://doi.org/10.1016/j.mulfin.2010.07.002

Lei, Z., & Zhang, C. (2016). Leveraged Buybacks. *Journal of Corporate Finance*, *39*, 242–262. https://doi.org/https://doi.org/10.1016/j.jcorpfin.2016.04.004

Manconi, A., Peyer, U., & Vermaelen, T. (2019). Are Buybacks Good for Long-Term Shareholder Value? Evidence from Buybacks Around the World. *Journal of Financial and Quantitative Analysis*, *54*(5), 1899–1935. https://doi.org/https://doi.org/10.1017/S0022109018000984

Mietzner, M. (2017). Why Do Firms Decide to Stop Their Share Repurchase Programs? *Review of Managerial Science*, *11*, 815–855. https://doi.org/https://doi.org/10.1007/s11846-016-0206-z

Oded, J. (2011). Stock Repurchases: How Firms Choose Between a Self Tender Offer and an Open-Market Program. *Journal of Banking & Finance*, *35*(12), 3174–3187. https://doi.org/https://doi.org/10.1016/j.jbankfin.2011.05.006

Ota, K., Lau, D., & Kawase, H. (2022). Signal Strength Adjustment Behavior: Evidence from Share Repurchases. *Journal of Banking & Finance*, *143*, 106545. https://doi.org/https://doi.org/10.1016/j.jbankfin.2022.106545

Otchere, I., & Ross, M. (2002). Do Share Buy Back Announcements Convey Firm-Specific or Industry-Wide Information? A Test of the Undervaluation Hypothesis. *International Review of Financial Analysis*, *11*(4), 511–531. https://doi.org/https://doi.org/10.1016/S1057-5219(02)00068-6

Pearce, J. A., & Robinson, R. B. (2004). Hostile Takeover Defenses That Maximize Shareholder Wealth. *Business Horizons*, *47*(5), 15–24. https://doi.org/https://doi.org/10.1016/j.bushor.2004.07.004

Pistor, K., Keinan, Y., Kleinheisterkamp, J., & West, M. D. (2003). Innovation in Corporate Law. *Journal of Comparative Economics*, *31*(4), 676–694. https://doi.org/https://doi.org/10.1016/j.jce.2003.09.004

Regierungskommission Deutscher Corporate Governance Kodex. (2022). Deutscher Corporate Governance Kodex F.2. https://www.dcgk.de//files/dcgk/usercontent/de/download/kodex/220627_Deutscher_Corporate_Governance_Kodex_2022.pdf

Schulz, B. (2024). *Accelerated Share Repurchase, Der beschleunigte Rückkauf eigener Aktien* (1st). Nomos Verlagsgesellschaft mbH & Co. KG.

Stephens, C. P., & Weisbach, M. S. (1998). Actual Share Reacquisitions in Open-Market Repurchase Programs. *The Journal of Finance*, *53*(1), 313–333. https://doi.org/https://doi.org/10.1111/0022-1082.115194

STOXX Ltd. (2025). Blue Chip Indices CDAX Index Description. Retrieved May 18, 2025, from https://stoxx.com/index/CDAX/?factsheet=true

The European Commission. (2016). COMMISSION DELEGATED REGULATION (EU) 2016/1052. https://eur-lex.europa.eu/eli/reg_del/2016/1052/eng

Vafeas, N. (1997). Determinants of the Choice Between Alternative Share Repurchase Methods. *Journal of Accounting, Auditing & Finance*, *12*(2), 101–124. https://doi.org/https://doi.org/10.1177/0148558X9701200201

Vermaelen, T. (1981). Common Stock Repurchases and Market Signalling: An Empirical Study. *Journal of Financial Economics*, *9*(2), 139–183. https://doi.org/https://doi.org/10.1016/0304-405X(81)90011-8

Zhang, H. (2005). Share Price Performance Following Actual Share Repurchases. *Journal of Banking & Finance*, *29*(7), 1887–1901. https://doi.org/https://doi.org/10.1016/j.jbankfin.2004.06.038

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Firm-wide Effects of Earnings Stripping Rules

Christoph Rehrl

Katholische Universität Eichstätt-Ingolstadt

Abstract

This paper estimates the impact of the earnings stripping rule reform by the ATAD Directive on the firm-wide external debt ratio of large multinational firms in the European Union in 2019. The analysis is based on the consolidated financial statements of STOXX Europe 600 constituents in 12 different Member States between 2014 and 2022. Firms that were not subject to a limitation of total interest expenses before the reform reduced their external debt ratio by 2.1 percentage points on average, which indicates an 8.3 percent decrease based on the pre-reform average external debt ratio of treated firms. The effect is lower at 1.3 percentage points but equally significant upon variations in the control group. The estimation results indicate that firms restricted in only internal interest expenses before the reform, firms in sink tax havens, and firms in industries with naturally higher net-interest-to-EBITDA ratios were particularly strongly affected by the reform. The findings suggest that the earnings stripping rule reform in the European Union significantly affected the external tax-related debt bias of large multinational firms.

Keywords: capital structure; earnings stripping rule; interest limitation; tax avoidance

1 Introduction

In most countries, interest expenses resulting from a loan are usually deductible from the taxable income at the corporate level, whereas dividends must be paid out of after-tax corporate income. Therefore, debt has systematic tax advantages in comparison to equity in most corporate taxation systems, incentivizing companies to increase their leverage (DeAngelo and Masulis, 1980, p. 21; Modigliani and Miller, 1958, p. 293; Modigliani and Miller, 1963, p. 434). For multinational firms, the incentive to manipulate their capital structure is increased due to their ability to exploit international tax differentials by giving internal loans from affiliates in countries with a relatively low tax burden to affiliates in countries with a relatively high tax burden (Desai et al., 2004, p. 2484; Huizinga et al., 2008, p. 1165).

International debt shifting is one of the main channels, besides transfer pricing and the allocation of intangibles, of corporate tax avoidance (Heckemeyer and Overesch, 2017, p. 986; Tørsløv et al., 2023, p. 1518). Due to the mobility of capital, debt-shifting between controlled entities is arguably one of the simplest methods of international profit-shifting (OECD, 2015a, p. 11). One prominent legislative answer to prevent excessive interest deductions that could erode tax revenues is interest limitation rules (ILR). Interest limitation rules aim to protect the domestic tax base by denying interest expense deductibility for tax purposes above a certain threshold.

In 1971, Canada was the first OECD country to implement an interest limitation rule, followed by Australia (1987) and the U. S. (1989) (Buettner et al., 2012, p. 930). Similar regulations became increasingly popular, and by 2005, around two-thirds of OECD members and more than half of EU Member States had implemented some form of interest limitation rule (Blouin et al., 2014, pp. 23 f. Buettner et al., 2012, p. 931; Ruf and Schindler, 2015, p. 19). This "first generation" of interest limitation rules is called thin capitalization rules (TCR) and limits interest expenses, usually only on internal loans from foreign affiliates, if the entity exceeds a certain debt-to-equity ratio (Blouin et al., 2014, p. 8; Buettner et al., 2012, p. 931).

The empirical studies by Blouin et al. (2014) and Buettner et al. (2012, pp. 935 f.) suggest that thin capitalization rules are successful at reducing the internal leverage of affected affiliates, which generally implies their effectiveness against debt-shifting (see also: Overesch and Wamser, 2010, p. 572; Weichenrieder and Windischbauer, 2008, p. 29). However, theoretical literature expresses concerns that thin capitalization rules may adversely affect investment due to an increase in the cost of capital (Mardan, 2017, pp. 3 ff.) and fuel tax competition (Haufler & Runkel, 2012, p. 1094). Both Buettner et al. (2018, p. 16) and De Mooij and Liu (2021, p. 5) find adverse effects on investment upon the implementation of thin capitalization rules. The

First, I would like to thank Prof. Dr. Reinald Koch for his inspiration and continuous support throughout this thesis. I would also like to thank Fabian Eigner for patiently proofreading a text far outside his field of interest.

DOI: https://doi.org/10.5282/jums/v11i3pp642-670

© The Author(s) 2026. Published by Junior Management Science.
This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.

results by Buettner et al. (2012, p. 935), Blouin et al. (2014, p. 16) and Wamser (2014, p. 18) indicate that firms partly substitute external debt for internal debt as an answer to a restriction on only internal interest expenses. Additionally, a simple debt-to-equity ratio is susceptible to manipulation, as corporations may specifically increase the equity of affiliates to avoid an interest limitation (Goyvaerts and Roggeman, 2020, p. 43; OECD, 2015a, p. 97; Weichenrieder and Windischbauer, 2008, p. 14).

The weak points mentioned, as well as the increased awareness of the risk of excessive corporate debt levels through the global financial crisis in 2008, contributed to the popularity of the "second generation" of interest limitation rules (Bachmann et al., 2015, p. 2; IMF, 2009, p. 12). These so-called earnings stripping rules (ESR) generally[1] limit internal and external interest expenses if the net interest expenses of an entity exceed a specific ratio of its adjusted income for tax purposes (OECD, 2017, pp. 198 ff.). Germany was the first[2] country to replace its existing thin capitalization rule with a standalone earnings stripping rule in 2008, followed by a variety of European countries that either instituted an earnings stripping rule standalone or on top of an existing thin capitalization rule (for an overview see: Gresik et al., 2017, p. 69).

Theoretical literature tends to favor earnings stripping rules over thin capitalization rules as welfare optimal, especially in developed high-tax countries (Gresik et al., 2017, p. 74; Kalamov, 2020, p. 61; Mardan, 2017, p. 10). Both Buslei and Simmler (2012, p. 20) and Alberternst and Sureth-Sloane (2016, p. 27) find a significant reduction in the debt-to-asset ratios of affected firms upon the implementation of the German earnings stripping rule in 2008 (see also: Dreßler & Scheuering, 2012, pp. 20 ff.). Buslei and Simmler (2012, p. 25) found no adverse effects on investment in the context of the German earnings stripping rule reform in 2008. Carrizosa et al. (2023) analyze the tightening of the existing regulation in America in 2017 and find that affected firms decreased their leverage significantly. Nonetheless, concerns about tax competition due to heterogeneity in interest limitation rules remained, and a growing voice called for a coordinated supranational approach to interest limitation rules (European Commission, 2007, p. 9; G20, 2012, p. 8; OECD, 1998, p. 9).

The first multilateral initiative to standardize interest limitation rules was launched by the G20/OECD (2013, p. 17) in 2013 as part of their action plan on Base Erosion and Profit Shifting (BEPS). In 2016, the European Commission issued the EU Anti-Tax Avoidance Directive 2016/1164 (ATAD I), which adopts the OECD's suggestion for an earnings stripping rule for its Member States as a mandatory adaptation until 2019. By analyzing the effect of the earnings stripping rule reform in the European Union in 2019 on the consolidated capital structure of the largest listed European corporations, I want to add to the existing literature in two main aspects. Firstly, no prior study estimates the effects of earnings stripping rules on the consolidated capital structure of large multinational firms. Prior studies on earnings stripping rules relied on individual financial statements (Alberternst and Sureth-Sloane, 2016, p. 14; Buslei and Simmler, 2012, p. 14; Dreßler and Scheuering, 2012, p. 14). I argue that the effectiveness of earnings stripping rules can only be thoroughly evaluated at the group level since adjustment effects due to tax restructuring may not be apparent at an individual firm level. De Mooij and Hebous (2018, p. 374) found a significant decrease in the external debt ratio of consolidated firms upon a restriction of total interest expenses, but do not specifically focus on multinational firms or earnings stripping rules. Secondly, as of November 2023, more than 140 countries committed to the implementation of the BEPS framework (OECD, 2023a). Empirical findings on the first multilateral implementation of the OECD's earnings stripping rule recommendation are relevant for policymakers to address existing deficiencies, understand intra-EU tax competition, and support policymakers in countries that plan to adopt the BEPS best practice approach into national law.

My analysis is based on the consolidated financial statements of the STOXX Europe 600 constituents from 2014 to 2022 and focuses on EU Member States. The firm-level financial data is augmented with data I collected on the European Union's pre- and post-reform country-level characteristics of interest limitation rules. The research design is based on a fixed-effects difference-in-difference (DID) approach. The control group consists of firms in those Member States with an equally effective interest limitation rule prior to the reform, as determined by the European Commission (2018, p. 1). I find a 2.1 percentage point reduction in the external debt-to-asset ratio of firms that were not limited in total interest expenses before the reform, which indicates an 8.3 percent decrease based on the pre-reform average external debt ratio of treated firms. The effect is lower at 1.3 percentage points but equally significant if I exclude French firms from the control group. I find a particularly consistent and strong reaction from firms that were only restricted in their internal interest expenses before the reform, which may indicate the "reversal" of the substitution effect between external and internal debt found in prior studies (Buettner et al., 2012, p. 936; Wamser, 2014, p. 22). I expect tax havens to act as net creditors in the context of debt shifting. I postulate that sink tax havens depend on external debt to fund the transfer of intangible assets. Consistent with my hypothesis, sink tax havens experienced a notable decrease of 4.3 percentage points in their external debt ratio following the reform. Lastly, the results indicate that firms operating in industries with inherently higher net-interest-to-EBITDA ratios are affected more strongly on average by the reform. How-

[1] Finland, Norway and Slovakia temporarily had earnings stripping rules in place that only limited internal interest expenses (De Mooij & Hebous, 2018, p. 370).

[2] The U. S. combined a thin capitalization rule with a variant of an earnings stripping rule back in 1989 (Umbrecht & Llewellyn, 1994, pp. 650 f.). The combined rule only restricted internal interest expenses of foreign affiliates and only if both the debt-to-equity and net-interest-to-adjusted-income ratios were exceeded. It can therefore not be directly compared to a modern standalone earnings stripping rule.

ever, outliers suggest potential variations in the availability of escape clauses across different industries. My results indicate that the earnings stripping rule reform in the European Union significantly affected the external debt bias of large multinational firms.

Section 2 of this paper starts with the institutional background of interest limitation rules. Section 3 provides a literature review on both thin capitalization and earnings stripping rules. Section 4 explains the best practice approach as recommended by the OECD. Section 5 gives an overview of the national implementation of the earnings stripping rule reform in the European Member States and their pre-reform interest limitation rules. Section 6 presents my four main hypotheses. Sections 7 and 8 outline the research design and descriptive statistics, respectively. Section 9 presents my empirical results, followed by robustness tests and extensions in Section 10. Lastly, Section 11 concludes and outlines potential avenues for future research.

2 Institutional Background of Interest Limitation Rules

Prior research largely categorizes earnings stripping rules as a subcategory of thin capitalization rules. Some authors refer to debt-to-equity ratio-based rules as "safe harbour" (Gresik et al., 2017; Ruf and Schindler, 2015) or "safe haven" rules (Mardan, 2017) and to earnings-related restrictions as "earnings stripping" rules. The majority of the literature on the German earnings stripping rule in 2008 uses the literal German translation "interest barrier" as a subcategory of thin capitalization rules (Alberternst and Sureth-Sloane, 2016; Buslei and Simmler, 2012; Dreßler and Scheuering, 2012). Ruf and Schindler (2015, p. 19) argue that the "differentiation [between thin capitalization rules and earnings stripping rules] is too semantic". This argument may have been sufficient when earnings stripping rules were still of secondary importance, but it is no longer appropriate for the current state of development. I would like to address three points that illustrate the deficiencies of the previous nomenclature.

Firstly, both terms "safe haven rule" and "safe harbor rule" for debt-to-equity rules are misleading. It is not apparent why a net-interest-to-earnings ratio would be less of a safe harbor than a debt-to-equity ratio, except that one might argue that a debt-to-equity ratio is easier to control. However, earnings stripping rules usually include additional net-interest safe harbors, which makes the terminology inconsistent. The standalone term "safe haven/harbor rule" is also very impractical for potential future discussions in the context of other tax legislation, for example, the safe harbor rules of Pillar II (OECD, 2022a). Secondly, both rules are inherently different in their technical design and target different categories of financial ratios. Thin capitalization rules are based on balance sheet information, and firms need to consider their balance sheet figures to avoid interest limitations. Earnings stripping rules rely on information from the profit-and-loss statement, and companies must consider their operating figures to avoid interest limitations. Thirdly, Ruf and Schindler (2015, p. 19) state that "both set of [...] rules have analogous effects". Earlier empirical studies (Alberternst and Sureth-Sloane, 2016, p. 27; Buslei and Simmler, 2012, p. 29; Carrizosa et al., 2023, pp. 13 ff. Dreßler and Scheuering, 2012, pp. 20 ff.) and theoretical literature (Gresik et al., 2017, p. 75; Mardan, 2017, p. 13) suggest that earnings stripping rules entail different adjustment effects than thin capitalization rules. Furthermore, thin capitalization rules tend to restrict internal interest expenses only, whereas earnings stripping rules tend to limit total interest expenses. Unfortunately, this is not a reliable distinction as some countries deviate from this classification (see Section 2.2). I specifically differentiate between thin capitalization and earnings stripping rules as two separate categories of interest limitation rules. This is consistent with the OECD's (2017, p. 202) updated final report on debt shifting.

2.1 Debt and Taxes

In a purely domestic scenario, the (corrected) seminal theoretical work by Modigliani and Miller (1958, p. 293) and Modigliani and Miller (1963, p. 434) states that tax advantages of debt are the only permanent advantages resulting from variation in corporate finance structure. Further developed extensions of the traditional static trade-off theory suggest that a firm finds its optimal capital structure when the marginal benefit (e. g. tax savings)[3] equals the associated marginal cost (e. g. bankruptcy and/or agency costs) of an increase in the share of debt (Bradley et al., 1984, p. 858; DeAngelo and Masulis, 1980, p. 21; Kraus and Litzenberger, 1973, p. 915; S. C. Myers, 1977, p. 172; Stiglitz, 1969, p. 788).[4] While the theories vary in assumptions and specifications, most stress that taxes are likely to be one of the driving determinants of optimal financing structure (Barclay and Smith, 2005, p. 9; Marsh, 1982, p. 122).

It has been particularly difficult to empirically prove the corporate debt sensitivity to tax incentives in a domestic setup and earlier studies found no significant (Mayer, 1990, p. 320) or ambiguous (Rajan & Zingales, 1995, p. 1444) results. However, studies employing a more sophisticated identification approach did find significant and economically strong effects of tax rates on corporate debt (Graham, 1996, p. 71; MacKie-Mason, 1990, p. 1483). Studies that employ a refined research method with the correct control variables, excluding multinational subsidiaries, find a substantial tax effect with a 2.3 percentage point increase in debt-to-assets ratios on average if the marginal tax rate increases by 10 percentage points (Feld et al., 2013, p. 2857). Heider and Ljungqvist (2015, p. 708) show that taxes have a first-order effect on capital structure. The inherent tax advantage of debt over equity is commonly referred to as debt bias

[3] Notably, more recent literature suggests that there are other benefits to debt, e. g. it can discipline overspending managers (Huizinga et al., 2008, p. 81).

[4] There are other theories on corporate finance which do not identify tax rates as first order decision relevant, see e. g. Pecking Order theory (Donaldson, 1961; S. C. Myers and Majluf, 1984).

(De Mooij and Hebous, 2018, p. 369; Heckemeyer and de Mooij, 2017, p. 53).

Apart from Miller (1977, p. 262), who hints in a footnote that multinational companies seem to face different debt-related tax incentives than domestic companies, early theoretical literature has largely excluded the special position of multinationals. To understand the aggravated tax incentives for multinationals, one needs to differentiate between the two general types of debt financing, external and internal debt. In this context, external debt is defined as loans borrowed from the external capital market. External debt constitutes a liability generating interest expenses at a consolidated corporate group level. Internal debt is defined as loans between related affiliates, thus a mere redistribution of cash within the group (Desai et al., 2004; Huizinga et al., 2008). Generally, intercompany loans on a national level do not generate tax advantages since the interest expense deductible at one affiliate is taxed at the same tax rate as interest income at the other.[5] This is different for cross-border intercompany financing since the tax rates between the different affiliates might differ significantly due to different national corporate tax regimes. Therefore, internal debt can be used to exploit the same tax advantage as external debt, only without the disadvantage of interest payments to third parties. It was Hodder and Senbet (1990, p. 1513) who extended the Miller equilibrium to an international setting and state that in equilibrium, multinational companies exploit international tax differentials by taking on debt in the most tax-advantageous jurisdictions.

The empirical research provides strong evidence on international tax arbitrage. Desai et al. (2004, p. 2452) and Huizinga et al. (2008, p. 81) find that a 10 percentage points higher local tax rate indicates a higher debt ratio of that affiliate on average between 2.4 and 2.8 percentage points (see also: Buettner et al., 2009, p. 310; Mintz and Weichenrieder, 2005, p. 10). Feld et al. (2013, p. 2852) analyze 46 prior empirical studies and find that the average marginal tax effects on debt for multinational companies are almost twice as strong compared to purely national firms. The results by Desai et al. (2004, p. 2452) suggest that internal leverage is especially sensitive to tax differentials, indicating that the incentive to use internal loans is strongly tax-influenced. The exploitation of international tax differentials through a corporate group's capital structure is commonly referred to as debt shifting (De Mooij and Hebous, 2018, p. 369; Heckemeyer and de Mooij, 2017, p. 53).

2.2 Thin Capitalization Rules and Earnings Stripping Rules

A far-reaching reform of tax law and equalization of debt and equity has often been sought after but has not yet been successful, mainly due to the complexity of a reform and budgetary constraints (De Mooij, 2012, p. 489).[6] Instead, countries have opted for various interest limitation rules to adapt to their existing tax system. Limiting existing systematic tax-related advantages of debt can arguably be seen as the second-best solution to combat excessive debt financing. The most prominent interest limitation rules are thin capitalization rules and earnings stripping rules.

Most thin capitalization rules determine a debt-to-equity ratio up to which all interest is deductible without restrictions.[7] The threshold is usually defined by the ratio of total or internal debt to equity. Above that ratio, the deduction of interest expenses is denied. Generally, only interest payments to related parties are restricted, regardless of whether the ratio is determined by internal or total debt. However, some countries restrict total interest expenses based on thin capitalization rules which has not been considered in earlier literature except by De Mooij and Hebous (2018, p. 370). One must be especially cautious when interpreting earlier studies on this matter, as some differentiate between the chosen debt ratio (total or internal debt) and some between the form of interest affected by the restriction (total or internal interest). Certain countries reclassify excess interest as dividends, which can entail additional withholding taxes. Originally, most thin capitalization rules only limited internal interest expenses to foreign affiliates and did not apply to internal finance by domestic investors (Blouin et al., 2014, p. 23; Buettner et al., 2012, p. 932).[8] This is intuitive, given that thin capitalization rules were specifically designed to curb international tax avoidance and should not restrict national firms.

However, the exclusive restriction of internal interest expenses to foreign subsidiaries has led to a groundbreaking legal dispute in Europe. In December 2002, the European Court of Justice (ECJ) ruled in the "Lankhorst-Hohorst case" (2002, C-324/00) that the German thin capitalization rule (prior to 2004) infringed upon the principle of "freedom of establishment" since it discriminated against investors from other EU countries. Hence, most EU members had to adjust their existing thin capitalization rules to comply with the internal market principles (for details, see: Ruf & Schindler, 2015, p. 20). Thin capitalization rules still play an important role in many jurisdictions. The last official report from the OECD (2020, p. 50) on this matter states that out of 134 members of the OECD's Inclusive Framework, 67 juris-

[5] In certain scenarios purely national intercompany loans can lower a group's effective tax rate if there are differences in the affiliates effective tax rates, e. g. differences in the legal forms of the two companies, the existence of loss carryforwards or differences in provincial/territorial income taxes.

[6] The most recent approaches to (partially) harmonize the treatment of equity and debt are the DEBRA (European Commission, 2022) project in Europe, which is currently suspended, and the blueprint "A Better Way for Tax Reform" (Ways and Means Committee, 2016) in the United States. A working group around Michael P. Devereux (Devereux et al., 2021) currently elaborates the details of a Destination-Based Cash Flow Taxation (DBCFT) system which would also ensure neutral treatment of debt and equity.

[7] New Zealand as an exception still employs a thin capitalization rule that relies on a debt-to-asset ratio (PwC, 2024a).

[8] De la Feria and Dourado (2008) provide an in-depth overview on thin capitalization rules in the European Union around 2007.

dictions employ an interest limitation rule and around two-thirds still solely rely on a thin capitalization rule.

> "Given the large potential macroeconomic damage from excess leverage [. . .], it is hard to see why – as now is often the case – debt finance should be systematically tax favored." – (IMF, 2009, p. 12)

While earlier tax policy mainly focused on debt-shifting via internal debt, the global financial crisis around 2008 increased the awareness of general systematic risks of external debt (Bachmann et al., 2015, p. 2; Heckemeyer and de Mooij, 2017, p. 500; IMF, 2009, p. 12). A high level of debt within a firm increases the likelihood of bankruptcy due to the consistent cash outflow of interest payments (Auerbach, 2002, p. 1267; Stiglitz, 1972, pp. 459 f.). General overborrowing in an economy significantly increases the probability and severity of a financial crisis (Bianchi, 2011, p. 3400). While tax-related debt bias is not the only reason for overborrowing, it certainly contributes to it (De Mooij, 2012, p. 498). Hence, jurisdictions aimed to address both debt bias and debt-shifting in future policies, contributing to the popularity of the so-called earnings stripping rules. Earnings stripping rules target interest expenses directly and deny interest expense deductibility above a certain net-interest-to-profitability ratio, usually a firm's earnings before interest, taxes, depreciation, and amortization (EBITDA). Even though the initial goal of Earnings stripping rules was to target total interest expenses, some countries apply a restriction on internal interest expenses only (see Table 15 in the Appendix). Currently, all European Member States (see Table 6) and around 60 percent of the G20 states apply an earnings stripping rule (see Table 15 in the Appendix).

2.3 Interest Limitation Rules – an Overview

So far, I have emphasized thin capitalization and earnings stripping rules as the main tools to limit excessive interest expenses. Even though those are the most prominent, they are not the only interest limitation rules countries apply to restrict debt shifting. Most countries combine multiple limitations to close loopholes or to address certain country-specific tax risks (OECD, 2017, pp. 23-26). I want to briefly highlight the most common regulations and why they are usually insufficient as standalone measures to prevent debt shifting. Table 1 gives an overview of the outlined regulations and their effect on debt bias and debt shifting.

Most jurisdictions (OECD, 2023b) apply the OECD transfer pricing guidelines, which state that any intra-group transaction must be priced at the arm's length principle, hence priced as it would have been between unrelated third parties. The often subjective and difficult estimation of a correct transfer price leaves leeway to set the most tax-advantageous transfer price in a range of acceptable prices. Scandals like the Luxembourg Leaks and their multinational companies' very accommodating Advanced Pricing Agreements (APA), as well as academic studies (Heckemeyer & Overesch, 2017; Sikka & Willmott, 2010; Swenson, 2001), have shown that manipulating transfer prices is one of the most prominent channels for international tax avoidance. Furthermore, even if a correct transfer price on interest is set, a tax advantage from debt shifting is only restricted but not fully prevented. Transfer pricing regulations do not limit the general debt bias regarding third-party debt.

Table 1: Categories of Interest Limitation Rules

Notes: Table 1 gives an overview of the most prominent anti-tax avoidance regulations countries employ to safeguard their tax base concerning debt bias and debt shifting. Source: own presentation, see text.

	Aimed at	
	Debt bias	Debt shifting
1. Arm's length principle	x	Yes
2. Withholding taxes	x	Yes
3. CFC-Rules	x	Yes
4. Thin Capitalization Rules	usually not	Yes
5. Earnings stripping rules	usually yes	Yes

Withholding taxes are, in theory, very well suited to curb tax incentives for cross-border intragroup financing, as they reallocate taxing rights to the source country. Therefore, interest income from internal loans is taxed at the same tax rate as interest expenses, preventing tax arbitrage. In reality, most bilateral tax treaties assign tax rights to the country of residency of the foreign bondholder and reduce or abolish the withholding taxes, which leaves them insufficient to prevent debt shifting fully (Buettner et al., 2009, p. 931). Multinational firms have been shown to specifically set up affiliates in countries with the most favorable tax treaty networks to avoid withholding taxes, so-called "treaty shopping" (Weyzig, 2013, p. 911). Like transfer pricing regulations, withholding taxes do not affect the general debt bias.

Lastly, many countries employ some form of Controlled Foreign Company (CFC) regulations (OECD, 2024a). Firms may establish affiliates in low-tax countries to reallocate mobile or passive activities. The OECD specifically lists financial and banking services as activities with a high risk for base erosion and profit shifting (OECD, 2015b, p. 43). CFC rules apply in the context of controlled foreign entities and usually only if the host country is below a certain effective or nominal tax rate. Consequently, the taxing rights on profits generated by the foreign-controlled company are (partly) reallocated to the parent company's country. Again, CFC rules do not address the general debt bias and may be circumvented through the interposition of conduit tax havens (Lejour, 2021, p. 2).

3 Literature Review

The heterogeneity of interest limitation rules makes consistent and correct categorization extremely challenging. Earlier literature partly relied on insufficient distinctions between types of interest limitation rules or even incorrect assumptions. Mardan (2017, p. 1) states that interest limitation rules generally restrict only internal interest expenses

and only vary in their threshold based on debt-to-equity or net-interest-to-EBITDA ratios. The author lists Germany and Spain as examples of internal interest restrictions through an earnings stripping rule, even though both countries restricted total interest expenses at the time of his study. Gresik et al. (2017, p. 69) state that earnings stripping rules restrict interest payments on internal debt, which is again incorrect or at least insufficient in detail, as most rules they list restrict internal and external interest expenses. Only Kalamov (2020, p. 39) acknowledges that earnings stripping rules vary in their scope of interest restriction. All of these studies assume that thin capitalization rules (or safe harbor rules, as some authors call them) only limit internal interest expenses, which is again not universally true (De Mooij & Hebous, 2018, p. 370). Therefore, all further outlined theoretical studies only consider a limitation of internal interest expenses in their model and only compare the effect of different threshold ratios.

Haufler and Runkel (2012, p. 1094) show that thin capitalization rules are inefficiently lax under tax competition since they can be used to attract foreign investment at a lower cost than changes in the statutory tax rate. Furthermore, multinational companies are tax-favored over national firms in a non-cooperative tax equilibrium, as national firms can not exploit the inefficiently lax allowance on internal debt. Thin capitalization rules are vulnerable to transfer price manipulation on interest rates as they do not directly consider interest payments under their threshold. Under an earnings stripping rule, a manipulation of transfer prices that leads to excessive interest payments is potentially restricted by the net-interest-to-EBITDA ratio.[9] The theoretical literature concludes that earnings stripping rules are more effective at curbing transfer price manipulations while thin capitalization rules incentivize transfer pricing manipulation (Gresik et al., 2017, p. 75; Kalamov, 2020, p. 55; Mardan, 2017, p. 9). Gresik et al. (2017, p. 74) argue that earnings stripping rules are superior to thin capitalization rules for any regime with a positive tax rate because the adverse effect on investment is only a second-order effect. Countries must be financially developed enough that a higher effective tax rate does not significantly lower investment (Mardan, 2017, p. 10). A multilateral coordination of interest limitation rules could reduce the adverse effects of tax competition and could reduce incentives for debt shifting more effectively (Haufler & Runkel, 2012, p. 1099). The theoretical literature deems earnings stripping rules to be welfare-optimal over thin capitalization rules as a tool to restrict internal interest expenses, especially in developed high-tax countries (Gresik et al., 2017, p. 75; Kalamov, 2020, p. 61; Mardan, 2017, p. 13).

[9] A recently published draft law by the German government contained an attempt to restrict transfer pricing regulations with regard to internal interest payments explicitly by law (German Parliament, 2023b, p. 133). This indicates that transfer price manipulation is still of relevance in an earnings stripping rule regime.

3.1 Thin Capitalization Rules

Table 2 provides an overview of key empirical results from prior studies on thin capitalization rules and their sample and time horizon. The table focuses on the effect on internal leverage ratios, which is the main emphasis of most studies, and therefore only shows a selection of the results. Most empirical studies were conducted before European Countries adapted their legislation due to the ECJ ruling. All studies only consider restrictions on internal interest expenses. The empirical research on thin capitalization rules has been heavily focused on Germany due to its comprehensive data collection on foreign direct investment (FDI) by the German Bundesbank and its role as a pioneer regarding interest limitation rules.

Both Overesch and Wamser (2010) and Weichenrieder and Windischbauer (2008, p. 14) examine German inbound FDI and find a significant reduction in internal leverage ratios upon the initial implementation of the German thin capitalization rule as well as its reforms in 2001 and 2004, which only restricted interest on internal debt. Overesch and Wamser (2010, p. 570) find that holding corporations react particularly quickly and strongly, with a decrease in internal debt of 10 percentage points, which is approximately a one-third reduction of the pre-reform mean of inter-company debt to capital. This indicates that holdings might have been more frequently used as a tax-planning entity. Buslei and Simmler (2012, p. 2) note that differentiation between external and internal debt may not be as easy in reality, especially in complex corporate groups that can potentially manipulate their shareholding structures. Wamser (2014, p. 22) analyses German inbound FDI around the thin capitalization rule reform in 2001 and finds a 5 percentage point decrease in a firm's internal debt ratio. Furthermore, the results indicate that around 55 percent of the reduction in internal debt is substituted by an increase in external debt (Wamser, 2014, p. 22).

The substitutability between internal and external debt is confirmed by Buettner et al. (2012, p. 935), who analyze the effect of thin capitalization rules on the capital structure of subsidiaries of German multinationals in 36 countries. Buettner et al. (2012, p. 936) consider the heterogeneity of thin capitalization rules between different jurisdictions and find that their effectiveness varies strongly with their specifications. The authors conclude that thin capitalization rules, which rely on a total-debt-to-equity ratio seem to reduce tax incentives for internal debt more successfully (Buettner et al., 2012, p. 937). I argue that the causality of this observation is not necessarily that total debt ratios are more efficient than internal debt ratios but rather that internal debt ratios are too lax in relative comparison, which becomes apparent based on their overview of thin capitalization rules (Buettner et al., 2012, p. 932). Naturally, a 3:1 debt-to-equity ratio is stricter if defined by total debt instead of only internal debt, as the total debt ratio is inherently at least as high as the internal debt ratio. The measure for the strictness of a rule employed by the authors does not consider the difference in baseline levels of total and internal debt (Buettner et al., 2012, p. 932).

Table 2: Overview of Empirical Studies on Thin Capitalization Rules

Notes: Table 2 gives an overview of the empirical research on thin capitalization rules, focusing on their effect on the internal leverage ratio and investment. Information on the sample, observation period, and other notable results is also included. Source: own presentation.

Study	internal leverage ratio		Investment	Sample	Timeframe	Other results
Buettner et al. (2012)	D/C	decreases	-	German outbound FDI in 36 countries	1996–2004	substitution effects
Blouin et al. (2014)	D/E	-6.3 percentage points	-	US outbound FDI in 54 countries	1982–2004	decrease in consolidated firm value
Overesch and Wamser (2010)	D/C	decreases	-	German inbound FDI	1996–2004	-
Buettner et al. (2018)	-	-	decreases	German outbound FDI in 39 countries	1996–2007	decrease in employement
Weichenrieder and Windischbauer (2008)	D/E	decreases	no effect	German inbound FDI	1989–2003	increase in equity
Wamser (2014)	D/C	-5 percentage points	-	German inbound FDI	around 2001	substitution effects
De Mooij and Liu (2021)	-	-	decreases	MNE affiliates in 34 countries	2006–2014	-
Buettner et al. (2016)	D/C	decreases	-	German outbound FDI in 36 countries	1996–2004	capital structure becomes insensitive to tax rate change above threshold
Goyvaerts and Roggeman (2020)	D	decreases	-	non-listed Belgium firms	2008–2014	increase in equity

Buettner et al. (2012, p. 936) only observe significant substitution effects of external for internal debt under total debt thresholds. This is counterintuitive, as an increase in external debt under a total debt threshold restricts internal interest expenses indirectly. Again, I argue that this effect is driven by the fact that total debt thresholds are stricter than internal debt thresholds relative to the underlying financial ratio, not because they focus on total debt.

Buettner et al. (2016, 2018) use the data sample on German outbound FDI in two additional studies focusing on the specific effects of thin capitalization rules. In their first follow-up study, the authors analyze the tax sensitivity of capital structures under a thin capitalization regime (Buettner et al., 2016, p. 787). They find that the capital structure of firms restricted in their interest expenses is insensitive to changes in the statutory tax rate (Buettner et al., 2016, p. 793). This aligns with the theoretical implication that thin capitalization rules remove the tax incentive to use internal debt for firms above the threshold. The second follow-up study by Buettner et al. (2018, pp. 565, 568) indicates that thin capitalization rules adversely affect investment and slightly decrease employment, especially in high-tax countries. Furthermore, the authors find that FDI is about twice as sensitive to tax rate changes under a strict thin capitalization regime than an unrestricted regime (Buettner et al., 2018, p. 566). This is consistent with the theoretical reasoning that companies cannot or only partly use increased debt financing to mitigate the impact of a tax rate increase if they are restricted in their interest expenses. De Mooij and Liu (2021, p. 4) find that multinational affiliates reduce their investment by 20.8 percent based on the mean asset level upon the implementation of thin capitalization rules in countries with a 27 percent statutory corporate tax rate. Weichenrieder and Windischbauer (2008, p. 29) find no adverse effects on investment upon the tightening of the German thin capitalization rule in 2001 on German inbound FDI, which may indicate that the reaction of investment differs between highly developed and other countries, in line with the theoretical by Mardan (2017, p. 10).

The study by Blouin et al. (2014, p. 29) suggests that both total-debt and internal-debt-ratio-based thin capitalization rules effectively decreased individual US affiliates' targeted leverage ratios in 54 countries. They do not find a significant reaction of the firm-wide debt-to-asset ratios upon implementation of thin capitalization rules. This is unsurprising, given that the authors focus on thin capitalization rules that restrict internal interest expenses only and, therefore, do not directly affect the external debt ratio of the consolidated group. Even though the consolidated leverage ratios were not affected, the consolidated interest expenses decreased. This could indicate that the adjustment of the firm's financial structure led to less creditor risk, for example through higher equity financing of affiliates, thus lowering interest rates (Blouin et al., 2014, p. 19). Furthermore, Blouin et al. (2014, p. 19) find a significant negative effect on consolidated firm value based on Tobin's q. This aligns with the theoretical implication that a reduced tax advantage of debt finance adversely affects investment and, therefore, firm valuation (Summers et al., 1981, p. 69).

3.2 Earnings Stripping Rules

Table 3 provides an overview of key empirical results from prior studies on earnings stripping rules, as well as their sample and time horizon. Empirical studies are few and suffer from little diversity in research design. There are five empirical studies on the effect of earnings stripping rules, four of which analyze the implementation of the German "interest barrier" in 2008 (Alberternst and Sureth-Sloane, 2016;

Table 3: Overview of Empirical Studies on Earnings Stripping Rules

Notes: Table 3 gives an overview of the empirical research on earnings stripping rules, focusing on their effect on the total leverage ratio and investment. Information on the sample, observation period, and other notable results is also included. Source: own presentation.

Study	Debt-to-asset ratio	Investment	Sample	Timeframe	Other results
Buslei and Simmler (2012)	-5.3 percentage points	no effect	German firms	2004–2008	firms split up their assets
Carrizosa et al. (2023)	-2.9%	-	U. S. firms	2014–2018	firms predominantly decrease long-term debt
Dreßler and Scheuering (2012)	No robust effect	-	German limited liability corporations	2005–2010	Only above-average leveraged firms react
Alberternst and Sureth-Sloane (2016)	-4.7 percentage points	-	German Firms	2006, 2008	Companies without liquidity constraints react more strongly
Knauer and Sommer (2012)	-	-	26 leveraged buyouts in Germany	1997–2008	Significant reduction of tax shield in leveraged buyouts
De Mooij and Hebous (2018)	-5 percentage points	-	consolidated firms in 60 countries	2005–2016	restrictions on total interest expenses are more effective

Buslei and Simmler, 2012; Dreßler and Scheuering, 2012; Knauer and Sommer, 2012) and one which examines the tightening of the existing earnings stripping rule through the Tax Cuts and Jobs Act (TCJA) in the U. S. in 2017 (Carrizosa et al., 2023). All five studies are based on individual firm financial statements. Empirical results on the German earnings stripping rule in 2008 should be treated with caution since those coincided with both a major corporate tax reform and the aftermath of the financial crisis, potentially diluting causal effects (Dreßler & Scheuering, 2012, p. 1).

Nevertheless, results on the German earnings stripping rule are of particular interest since the OECD recommended a best-practice approach based on the German role model (OECD, 2017, p. 29). Germany already had a thin capitalization rule prior to the reform. Therefore, the studies estimate the effect of a switch from a thin capitalization rule to an earnings stripping rule but do not cover the full impact of an earnings stripping rule from an "unregulated" regime.

Buslei and Simmler (2012, p. 24) find that in the unrestricted sample, German firms reduce their debt-to-asset ratio by 2 percentage points on average upon introduction of the earnings stripping rule in 2008. Another specification based on restricted control and treatment groups around the exemption limit indicates a significantly stronger reduction in the debt ratio of around 6 percentage points. This is in line with the expectation that the reform is only relevant for firms threatened by an application of the threshold. Furthermore, Buslei and Simmler (2012, p. 24) find suggestive evidence that firms try to bypass the interest restriction by splitting up their assets to make repeated use of the exemption limit. The study does not show any significant effects on investment levels in the short run. The authors conclude that firms either have low substitution cost of equity for debt, that firms find ways to avoid the interest limitation in the long term or that most affected firms do not shift profits.

Dreßler and Scheuering (2012) only find a significant effect on the debt ratio of German firms upon the reform in 2008 if the treatment group showed a higher average debt ratio pre-reform than the control group. The authors conclude that highly indebted firms generally try to lower their debt ratio or that the cut in the corporate income tax causes the adjustment in capital structure (Dreßler & Scheuering, 2012, p. 28). Only Dreßler and Scheuering (2012, p. 12) differentiate between the reaction of external and internal debt and find that the reaction is driven by a decrease in external debt. This is conclusive, given that the pre-reform thin capitalization rule had already restricted internal interest expenses. Alberternst and Sureth-Sloan find that treated German firms reduced their debt ratio by 4.7 percentage points upon the earnings stripping rule reform in 2008. Alberternst and Sureth-Sloane (2016, p. 11) define their treatment group as firms that cannot use escape clauses and are above the net-interest-to-EBITDA threshold based on their pre-reform characteristics in 2006. Unfortunately, the authors do not report whether their treatment group had a significantly higher absolute debt ratio before the reform, which could result in a similar bias as observed by Dreßler and Scheuering (2012).

Knauer and Sommer (2012, p. 209) analyze the effects of the German earnings stripping rule reform on the tax shield in the context of leveraged buyouts. The tax shield consists of the present value of all tax-induced savings from debt financing. The authors find that around 78 percent of buyouts should be affected by the new earnings stripping rule in the first year after the reform, significantly weakening the incentive for excessive debt-financed buyouts (Knauer & Sommer, 2012, p. 223). Concerning a change in tax revenues upon the earnings stripping rule reform in Germany, Broer (2009, p. 411) estimates an increase between EUR 809m and EUR 2,537m and Bach and Buslei (2009) an increase of around EUR 800m based on microsimulation models. Broer (2009, p. 411) concludes that after the reform, there is no tax incentive for parent companies in other EU Member States to favor debt over equity financing for affiliates in Germany, except for Estonia and Cyprus. Carrizosa et al. (2023) are the only authors to study the effects of earnings stripping rules outside of Germany. They find that affected U.S. firms decreased their leverage by 2.9 percent of assets upon tightening of the existing earnings stripping rule, which was implemented as part

of the Tax Cuts & Jobs Act in 2017.

Lastly, I want to highlight the study by De Mooij and Hebous (2018), which is the only study to include both earnings stripping and thin capitalization rules in its analysis based on consolidated firms in 60 countries between 2005 and 2016. Contrary to prior literature, they do not differentiate between earnings stripping and thin capitalization rules but rather between a restriction of internal or total interest expenses. Therefore, the observed firm-wide effects cannot be attributed to the type of interest limitation rule but only to the type of interest that is limited. De Mooij and Hebous (2018, p. 374) find a 5 percentage point decrease in the firm-wide debt ratio on average upon a restriction of total interest expenses. The results suggest that a limitation on total interest expenses is highly effective at reducing the debt bias of firms, even at the consolidated group level.

All prior empirical studies on earnings stripping rules coincide with major tax law reforms and, in the case of Germany, the global financial crisis around 2008. The only study that considers firm-wide effects of a restriction of total interest expenses does not focus on earnings stripping rules in particular and does not differentiate between national and multinational firms (De Mooij & Hebous, 2018, p. 11). We know surprisingly little about the impact of earnings stripping rules on the capital structure of large multinational groups, which are at least the focus of earnings stripping rules in aspects of debt shifting. As such, this paper aims to extend our knowledge of the firm-wide effects of earnings stripping rules based on the earnings stripping rule reform in the European Union in 2019, which builds on the OECD's best practice approach recommendations.

4 The OECD's Best Practice Approach to Interest Limitation

For almost twenty years, the OECD (1987, p. 33) specifically discouraged Member States from implementing "hard and fast debt-equity" (OECD, 1979, pp. 88 f.) rules and recommended a flexible approach to judge the adequacy of interest expenses on an individual case basis. It was not until 1998 that the OECD (1998, p. 9) realized that a purely individual case basis is widely insufficient to prevent debt shifting and identified debt shifting as a harmful tax practice of significant importance. The first multilateral initiative to standardize interest limitation rules was launched by the G20/OECD (2013, p. 17) in 2013 as part of the action plan on Base Erosion and Profit Shifting (BEPS), which emphasizes 15 action points to tackle international tax avoidance. The OECD released its recommendation on limiting BEPS based on interest deductions and other financial payments in 2015 (OECD, 2015c, p. 25), followed by a more comprehensive updated version in 2017 (OECD, 2017, p. 29). The updated version is mostly identical with the addition of details on the proposed group ratio rule as well as recommendations on how to deal with banks and insurance groups (OECD, 2017, p. 117 and 169). The earnings stripping rule proposed by the OECD (2023a) is of particular interest as 145 states are members of the inclusive framework on BEPS as of November 2023.

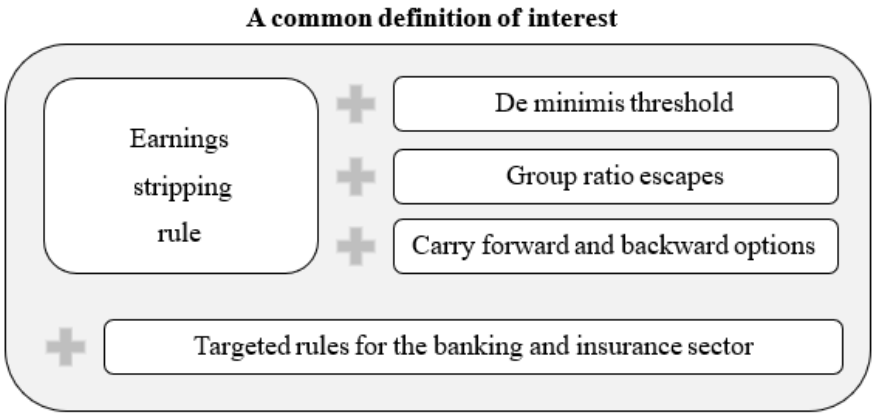
In line with prior research results and national trends, an earnings stripping rule with a fixed ratio of net-interest-to-EBITDA, very similar to the German "interest barrier", was proposed as a best practice approach by the OECD in 2015 (OECD, 2015a, p. 25). Proposals for a de minimis threshold, a group ratio escape, and certain carry forward and carry backward options amend the core fixed ratio limitation. Additionally, countries may consider a common definition of interest and implement targeted rules to address country-specific risks, as well as specific rules for the banking and insurance sectors. Table 4 gives an overview of the proposed earnings stripping rule with optional amendments. Furthermore, I aim to put previous literature into context with the recommendations of the OECD to assess whether the best practice approach aligns with theoretical and empirical results.

The OECD (2017, p. 29) encourages Member States to implement an earnings stripping rule that limits an entity's net interest deductions above a fixed percentage of its profits, measured by EBITDA based on tax numbers. Net interest expenses are interest expenses less interest income. The restriction of both external and internal interest expenses is recommended to prevent the observed substitution effects between internal and external debt in prior empirical studies that were noted earlier (Blouin et al., 2014, p. 16; Buettner et al., 2012, p. 935) and to reduce both tax avoidance and general risks of insolvency based on the German role model (German Parliament, 2007, p. 31). As the rule only restricts net interest expenses, entities can always deduct their interest expenses at least in the amount of their interest income. The advantage over the classic thin capitalization rule is that the limit is directly linked to economic activity and is not as easy to manipulate. The OECD's recommendation is in line with the general theoretical research on earnings stripping rules (Gresik et al., 2017, p. 75; Mardan, 2017, p. 12) as well as the empirical results (Alberternst and Sureth-Sloane, 2016, p. 27; Buslei and Simmler, 2012, p. 29; Dreßler and Scheuering, 2012, p. 28) and microeconomic simulations on the introduction of the German earnings stripping rule in 2008 (Bach and Buslei, 2009, p. 32; Broer, 2009, p. 411). Even though EBITDA is determined as the first measure of choice, it can also be reasonable to use Earnings Before Interest and Taxes (EBIT) or, in rare cases, based on asset values instead of earnings.[10] If a country relies on net-interest-to-EBITDA as a limitation, the OECD (2017, p. 53) recommends applying a ratio between 10% and 30%.

The OECD provides three leading suggestions meant to reduce companies' economic burden. Firstly, the OECD (2017, p. 39) recommends a de minimis threshold to relieve low-risk entities and the general administrative burden. The guidelines do not specify a specific amount for the

[10] Denmark had an earnings stripping rule relying on a debt-to-asset ratio in place until 2019, when they switched to an net-interest-to-EBITDA ratio (OECD, 2017, p. 51).

Table 4: The OECD's Best Practice Approach to Interest Limitation Rules



Notes: Table 4 gives an overview of the most critical aspects of the OECD's proposed earnings stripping rule. Source: own presentation based on OECD (2015a, p. 25).

exemption limit. The threshold should instead be based on an absolute monetary value of net interest expenses that reflects aspects of the local economy, such as interest rates or the tax regime. Any entity below the threshold can deduct its interest expenses without restrictions, whereas entities above the threshold must apply the earnings stripping rule to its full extent. Notably, the OECD suggests that the threshold should consider the total net interest expenses of all group entities in one country. The recommendation explicitly advises including an anti-fragmentation rule to prevent groups from establishing multiple entities and repeatedly exploiting the threshold. This recommendation aligns with empirical research by Buslei and Simmler (2012, p. 29) and a recent draft law by the German Parliament (2023b, p. 130), which confirms that entities otherwise split up their assets to exploit the threshold multiple times. Prior research also states that the effectiveness of an earnings stripping rule is highly dependent on the specific threshold amount, especially in the absence of an anti-fragmentation rule (Buslei & Simmler, 2012, p. 29). As we will see later from the actual national implementation, the EU Directive does not include such an anti-fragmentation rule, thus leaving room for tax restructuring.

Secondly, the group ratio escape is an effort to consider the circumstances of individual entities and groups and soften the effect of the blunt fixed ratio approach (OECD, 2017, p. 61). Different sectors and industries may naturally vary in leverage due to industry-specific factors unrelated to tax avoidance efforts. A group ratio option would allow entities to deduct net interest expenses up to the higher of their own or the group's overall fixed ratio, which does not necessarily have to be a net-interest-to-EBITDA ratio. For example, a new affiliate that is barely profitable with high interest expenses due to heavy initial investments could be additionally burdened by the interest limitation rule due to its low EBITDA. The regulation does not intend to restrict economic financing decisions excessively. Prior empirical research has no reliable estimates on the impact of group ratio escapes due to a lack of relevant variables (Alberternst and Sureth-Sloane, 2016, p. 17; Buslei and Simmler, 2012, p. 10; Dreßler and Scheuering, 2012, p. 10).

Thirdly, options for carrying forward and carrying back disallowed and unused interest capacity are meant to reduce the adverse effects of fluctuations in EBITDA and net interest expenses (OECD, 2017, p. 72). An unexpected slump in earnings with constant net interest expenses could additionally strain entities that are suddenly affected by an interest expense restriction. In other cases, an entity might have initial high net interest expenses due to significant investments that will be offset by additional earnings in the future. In both cases, the entity may be subject to an interest limitation in the first year simply because there is a timing mismatch between net interest expenses and EBITDA. Recurring volatility issues could lead to unwarranted double taxation since interest income is taxed at the lender while the borrower accumulates non-deductible interest expenses. Carry forward and carry backward options are an attempt to reduce distortions over the total period. While those options restrict double taxation, they also introduce the potential for debt shifting as they weaken the overall interest limitation effect. The OECD issues concerns especially over unlimited EBITDA carry forwards, which may add up to a significant tax asset and incentivize entities to increase their future net interest expenses (OECD, 2017, p. 73).

Banks and insurance groups vary significantly from other sectors in the role interest plays in their main business activities (OECD, 2017, p. 79). Due to its restriction on net interest expenses, a general earnings stripping rule is unsuitable to prevent debt shifting of business models with naturally positive net interest income, which will regularly be the case for banks and insurance groups. The OECD (2017, p. 185) recommends excluding financial institutions when

applying the fixed ratio to an overall local group, as their excessive interest income might be used to shelter the interest expenses of other group members. Financial entities are usually already restricted in their capital structure due to non-tax-related regulatory requirements.[11] Nonetheless, prior literature suggests that banks engage in profit shifting (Langenmayr & Reiter, 2022, p. 824), that their capital structure is indeed sensitive to changes in local tax rates (Heckemeyer and de Mooij, 2017, p. 73; De Mooij and Keen, 2016, p. 27) and that both banks and insurance groups could use a variety of arrangements to engage in tax avoidance (OECD, 2017, p. 79). Empirical research on interest limitation rules has so far excluded financial service providers due to their particular capital structure or because they are exempt from interest limitation rules (Alberternst and Sureth-Sloane, 2016, p. 14; Buettner et al., 2012, p. 934; Buslei and Simmler, 2012, p. 15).

The OECD (2017, p. 80) explicitly concludes that banks and insurance groups should be targeted by interest limitation rules, but does not provide a universal approach to do so. Instead, countries should implement unilateral rules harmonizing with existing national regulatory requirements for financial institutions. This inconsistent approach may leave room for tax competition and tax avoidance arrangements in the financial sector. The recommendation is especially concerning as prior studies found the cost of debt bias to be especially high in the financial industry (De Mooij, 2012, p. 500; Pazarbasioglu et al., 2010, p. 24). In 2023, 35 of the 100 largest groups were banks or insurance companies (Forbes, 2023). Given that banks and insurance companies are fundamentally important to the global economy, future research and policymakers should analyze closely whether openings for debt-shifting in the financial sector remain.

Lastly, I want to highlight Chapter 2 in the OECD's (2017, p. 33) best practice guide, which deals with the definition of interest and payments that are economically equivalent to interest. The OECD (2017, p. 34) provides some examples of essential aspects of a common interest definition but does not provide a concrete recommendation. I believe a common definition of interest expenses is one of the linchpins of any successful interest limitation rule on an international level. In jurisdictions with lax debt and interest expense definitions, companies might substitute interest expenses for economically equivalent expenses not affected by limitations on their deductibility for tax purposes, thus optimizing their overall tax burden. Different national definitions of interest expenses might be exploited to achieve interest income in one jurisdiction, which lowers net interest expenses, and non-interest expenses in another jurisdiction, which are not restricted by the earnings stripping rule. This simplified consideration suggests that, in theory, opportunities for tax arbitrage remain under an inconsistent definition of interest expenses between jurisdictions.

[11] See e. g. Basel IV for equity requirements of banks and the Solvency II Directive for requirements on risk management and equity quotas of insurance groups in the European Union.

5 The Earnings Stripping Rule Reform in the European Union

In the OECD's last official progress report on Action 4, out of 134 members of the Inclusive Framework, only 33 jurisdictions employed an earnings stripping rule in general, and not all of them are necessarily in line with the recommended best practice approach (OECD, 2020, p. 50). As of today, almost 10 years later, out of the 19 sovereign countries of the G20 Members, only 11 have implemented the best practice approach suggested by the OECD, or at least a similar version (see Table 15 in the Appendix). The low participation is particularly surprising as the project was carried out in cooperation with the G20 countries (OECD, 2013, p. 25).

Tørsløv et al. (2023, p. 1522) estimate that the European Union is the region most affected by profit shifting, with profit losses up to 1.5% of its GDP in 2015. The study suggests that debt shifting accounted for 15% of overall profits shifted to tax havens, not accounting for debt shifting between non-haven jurisdictions. This is significant given that most European Member States had already implemented some form of unilateral interest limitation rules at that point (De Mooij and Hebous, 2018, p. 370; OECD, 2017, p. 197). It is hardly surprising that the European Union plays a pioneering role in the implementation of the OECD's earnings stripping rule reform, in line with its announcement to take coordinated action against debt shifting already back in 2007 (European Commission, 2007, p. 9).

In 2016, the European Commission issued the EU Anti-Tax Avoidance Directive 2016/1164 (ATAD) which specifically aims to implement the action plan against BEPS by the OECD. Article 4 of the Directive contains the provisions on the earnings stripping rule that are to be transposed into national law. The Directive provides a framework for a minimum standard and Member States are free to implement stricter variants of the proposed rule wherever necessary (European Commission, 2016, p. 2). It is somewhat surprising that the EU decided to rely on the German "interest barrier rule" as template for the Directive, which is considered to be potentially in violation of Germany's constitution by the German Federal Fiscal Court (BFH) (2015, I R 20/15) and has been pending before the German Constitutional Court (BVerfGE) since 2016 (2 BvL 1/16).

I collected information on the interest limitation rules in European Member States before the reform as well as the details on the national implementation of the reform. The data stems from the International Bureau of Fiscal Documentation (IBFD), the PWC Worldwide Tax Summarize (PwC, 2024b) and the EY Worldwide Corporate Tax Guide (EY, 2023). For historical information on interest limitation rules I relied on the OECD database (OECD, 2019), KPMG EU Tax Center, and the overview provided by De Mooij and Hebous (2018, p. 370). The next section outlines the interest limitation rules in the European Union prior to the reform, followed by information on each Member State's national implementation of the Directive.

Table 5: Interest Limitation Rules Prior to the Reform in the European Union

Notes: Table 5 gives an overview of the interest limitation rule in place before the individual reform year of each European Member State and the United Kingdom. Source: own presentation, see text.

Country	Date of Impl.	Prior ILR		keep prior ILR
		TCR	ESR	
Austria	2021	x	x	x
Belgium	2019	total debt	x	Yes
Bulgaria	2019	total debt	x	Yes
Croatia	2019	total debt	x	Yes
Cyprus	2019	x	x	x
Czech Republic	2019	related party debt	x	Yes
Denmark	2019	total debt	total interest	Yes
Estonia	2019	x	x	x
Finland	2019	x	related interest	x
France	2019	related party debt	related interest	Yes
Germany	2024	x	total interest	x
Greece	2019	x	total interest	x
Hungary	2019	total debt	x	x
Ireland	2022	x	x	x
Italy	2019	x	total interest	x
Latvia	2018	total debt	x	Yes
Lithuania	2019	total debt	x	Yes
Luxembourg	2019	x	x	x
Malta	2019	x	x	x
Netherlands	2019	x	x	x
Poland	2018	related party debt	x	x
Portugal	2019	x	total interest	x
Romania	2018	total debt	x	x
Slovak Republic	2024	x	related interest	x
Slovenia	2024	related party debt	x	Yes
Spain	2024	x	total interest	x
Sweden	2019	related party debt	x	x
United Kingdom	2017	x	x	x

5.1 Interest Limitation Rules Before the Reform

Table 5 provides an overview of the interest limitation rules in the European Union prior to the reform. The information is based on the year before a country's individual implementation date of the reform. In line with prior research, I differentiate between thin capitalization rules with a total debt threshold and an internal debt threshold, as all of them restrict only interest expenses to related parties (Buettner et al., 2012, p. 932).[12] Earnings stripping rules are categorized by the form of interest restricted, as all of them consider total interest expenses in regard to their threshold. In this overview, I only consider interest limitation rules codified in the tax law.[13]

The interest limitation regimes prior to the reform varied greatly between Member States. Seven Member States, excluding the United Kingdom, had no interest limitation rule in place. Thin capitalization rules were the most popular form of interest limitation rules, with five Member States employing a related party debt threshold and eight Member States a total debt threshold.

Notably, most Member States maintained their existing thin capitalization rule after the reform. Three countries, namely Finland, France, and the Slovak Republic relied on an earnings stripping rule prior to the reform that only restricted internal interest expenses. France and Denmark were the only two Member States to combine a thin capitalization rule with an earnings stripping rule prior to the reform. Based on classic tax haven lists such as the one by Tørsløv et al. (2023, p. 1528) it becomes apparent that none of the EU tax havens (Belgium, Ireland, Luxembourg, Malta, Netherlands) employed any form of interest limitation rule prior to the reform except for Belgium. This is in line with the expectation that tax havens generally aim to allow an unrestricted flow of capital and rely on low effective tax rates.

5.2 Earnings Stripping Rules After the Reform

The next section outlines the specifics of the rule's national adaptations at the time of its implementation and the enactment schedule as outlined by the ATAD Directive. According to Article 11, paragraph 1, all Member States must apply national law, regulations, and administrative provisions in line with Article 4 of the Directive from January 1,

[12] Bulgaria's pre-reform thin capitalization rule theoretically also limits external interest expenses. However, interest expenses on loans and leases are exempt if not paid to related parties (KPMG, 2018, p. 4). Romania restricted interest only on loans from non-financial institutions above a certain threshold, hence mainly targeting internal interest expenses (KPMG, 2017, p. 3).

[13] Austria had customary debt-to-equity ratios based on administrative practice and court rulings, which I do not consider to be equally effective as a codified rule (KPMG, 2020).

Table 6: Earnings Stripping Rule Reform in the European Union

Notes: Table 6 illustrates the specifics of the national earnings stripping rule at the time of its implementation in European Member States and the UK. Currency conversions of the Member States are based on December 31, 2018. Notably, some Member States changed their earnings stripping rules in the years after the implementation. Source: own presentation, see text.

			Escape Rules				Carry Forward	
					Financial Group Esc.			
Country	Date of Impl.	30% of EBITDA Threshold	De Minimis	Stand-alone	Equity	EBITDA	Interest	EBITDA
Austria	2021	Yes	3m	Yes	Yes	x	Indefinetly	5 years
Belgium	2019	Yes	3m	Yes	x	x	Indefinetly	x
Bulgaria	2019	Yes	3m	x	x	x	Indefinetly	x
Croatia	2019	Yes	3m	Yes	x	x	3 years	x
Cyprus	2019	Yes	3m	Yes	Yes	Yes	5 years	5 years
Czech Republic	2019	Yes	3.2m	Yes	x	x	Indefinetly	x
Denmark	2019	Yes	3m	x	x	Yes	Indefinetly	5 years
Estonia	2019	Yes	3m	Yes	Yes	Yes	x	x
Finland	2019	25%/EBITD	3m	Yes	Yes	x	Indefinetly	x
France	2019	Yes	3m	Yes	Yes	Yes	Indefinetly	5 years
Germany	2024	Yes	3m	Yes	Yes	x	Indefinetly	5 years
Greece	2019	Yes	3m	x	x	x	Indefinetly	x
Hungary	2019	Yes	2.5m	x	Yes	Yes	Indefinetly	Indefinetly
Ireland	2022	Yes	3m	Yes	Yes	Yes	Indefinetly	5 years
Italy	2019	Yes	x	x	x	x	Indefinetly	5 years
Latvia	2018	Yes	3m	x	x	x	x	x
Lithuania	2019	Yes	3m	Yes	Yes	Yes	Indefinetly	x
Luxembourg	2019	Yes	3m	Yes	Yes	Yes	Indefinetly	5 years
Malta	2019	Yes	3m	Yes	Yes	Yes	Indefinetly	5 years
Netherlands	2019	Yes	1m	x	x	x	Indefinetly	x
Poland	2018	Yes	0.7m	x	x	x	5 years	x
Portugal	2019	Yes	1m	x	x	Yes	5 years	5 years
Romania	2018	Yes	1m	Yes	x	x	Indefinetly	x
Slovak Republic	2024	Yes	3m	x	x	x	5 years	x
Slovenia	2024	Yes	1m	Yes	x	x	x	x
Spain	2024	Yes	1m	x	x	x	Indefinetly	5 years
Sweden	2019	Yes	0.45m	x	x	x	6 years	x
United Kingdom	2017	Yes	2.2m	x	x	Yes	Indefinetly	5 years

2019, onwards. Table 6 gives an overview of the year of implementation, the majority being 2019. However, Article 11, paragraph 6, opens the possibility of delaying national implementation until January 1, 2024, at the latest if the Member States can prove that existing regulations are at least as effective as the provisions in Article 4. A regulation was deemed as equally effective only if it had both "(1) the legal similarity and (2) the economic equivalence" (European Commission, 2018, p. 1) as the new measures proposed by the Directive. This differentiation is the basis for my categorization into treatment and control groups in my empirical analysis.

The European Commission (2018, p. 1) deemed existing measures by Greece, France, Slovakia, Slovenia, and Spain as equally effective as Article 4 of the ATAD. Germany was allowed to postpone the implementation until 2024 since their prior earnings stripping rule served as prototype for Article 4 (European Parliament, 2022, p. 21). Both France and Greece voluntarily transposed the new measure into national law already in 2019.

Austria and Ireland applied to postpone the implementation of Article 4 but the European Commission (2019, p. 9) deemed their current interest limitation rules to not comply with the requirements of Article 4. Austria and Ireland applied the new interest limitation rules in accordance with Article 4 from 2021 and 2022 onwards, respectively. Latvia, Poland, and Romania already implemented an adjusted interest limitation rule in accordance with Article 4 from 2018 onwards. The United Kingdom was the first country to adapt its national legislation in 2017. Germany, the Slovak Republic, and Spain delayed the implementation until 2024. Slovenia is currently in the legislation process and expected to apply the new earnings stripping rule retroactively starting January 1, 2024.

The core earnings stripping rule is in line with the OECD's recommendation of a best practice approach. Paragraph 1 of Article 4 outlines a fixed ratio rule in the form of an earnings stripping rule that limits net-interest expenses that exceed 30 percent of an entity's EBITDA. An entity can be an individual taxpayer, a consolidated group for tax purposes, or an entity of a group that does not consolidate its results for tax purposes. All EU Member States implemented a 30 percent EBITDA threshold except for Finland, which chose a 25 percent Earnings Before Interest, Taxes, and Depreciation (EBITD) threshold. EBITD is generally higher[14] than EBITDA, which could be offset by the 5-percentage point reduction in the fixed ratio, depending on the individual entity. Arguably, Finland's fixed ratio might be less prone to manipulation since intangible assets are excluded.

Paragraphs 3 and 5 of Article 4 include four possible escape clauses that Member States are free to adapt into na-

[14] Unless the appreciation of intangible assets outweighs the depreciation.

tional law. The Directive includes options for a "de Minimis escape" and two escape clauses based on ratios of the consolidated financial group, which can also be found in the OECD guidance. All Member States allow firms to apply the net-interest-to-EBITDA ratio at the level of the consolidated group for tax reasons. Additionally, a "standalone escape" is offered, which would allow Member States to exclude entities without any affiliated entities from the application of the earnings stripping rule.

Every Member State except Italy applies some form of "de Minimis" threshold as defined in paragraph 3(a). Eighteen Member States incorporated the recommended threshold of EUR 3m for overall net interest expenses into national law. Finland is the only country that differentiates between external and internal debt and introduced a 3-million-euro threshold for external debt, but only EUR 0.5 million for internal debt. The Netherlands, Portugal, Romania, Slovenia, and Spain set the threshold significantly lower at only EUR 1m. Sweden and Poland have the lowest threshold with only around EUR 0.45m and 0.7m, respectively. Around half of the Member States exclude entities from the application of the rule if they are not part of a consolidated group for financial reasons and have no associated enterprise or permanent establishment as defined in paragraph 3. There is some diversity regarding the financial group escape options and Member States tend to implement either both escape clauses or none. The equity escape allows an entity to avoid the application of the rule if it can demonstrate that its equity over asset ratio is not more than two percentage points lower than that of the consolidated group for financial purposes. Under the EBITDA group escape, entities can choose to test for the net-interest-to-EBITDA threshold at the consolidated financial group level instead of each entity individually. Almost all Member States allow for a carry-forward of denied interest expense deductions, most of them indefinitely. Only Estonia, Latvia, and Slovenia do not allow any form of interest expense carry-forward, possibly to avoid administrative costs for the relatively small countries. Additionally, around half of the Member States allow firms to carry forward unused EBITDA capacity, usually for five years. The only exception is Hungary, which allows for an indefinite interest as well as indefinite EBITDA capacity carry-forward.

There is certain heterogeneity between the national implementation of the reform, which already indicates that tax competition within the EU by means of interest limitation rules is not fully ruled out by the Directive. The classical tax havens Luxembourg, Ireland, and Malta implemented the most lenient versions of the rule. On the other hand, Belgium and the Netherlands, generally considered tax havens, implemented almost none of the escape rules (Tørsløv et al., 2023, p. 1506). France and Germany also offer most escape clauses, potentially to avoid the adverse effects of tax competition. The specifics of the national implementation may be dependent on a variety of factors, such as the prior interest limitation regime, other additional existing interest limitation rules, and economic conditions. Notably, many Member States adjusted the original earnings stripping rule in the years after the reform. Greece implemented an equity and EBITDA escape on the financial group level (EY, 2023, p. 678). Netherlands lowered their fixed ratio to 20 percent net-interest-expenses-to-EBITDA (EY, 2023, p. 1266). Finland restricted the equity escape clause in 2022 and again in 2023 in certain situations (PwC, 2024c).

6 Development of Hypotheses

In the following, I empirically investigate whether the mandatory adaptation of the earnings stripping rule in 2019 in the European Union significantly affected the capital structure of the 600 largest listed European firms at the consolidated group level. For several reasons, the earnings stripping rule reform in the European Union is a promising quasi-experimental setup for an empirical study based on a Difference-in-Differences (DID) design. Firstly, the European Commission (2018, p. 1) assigned Member States to one of two categories, dependent on their interest limitation rule prior to the reform. Only Member States that did not apply an "equally effective" interest limitation rule before the reform are expected to show a significant reaction to the Directive. Based on this classification, I consider firms with their domicile in countries that did not have an equally effective interest limitation rule prior to the reform as the treatment group and those that already had equally effective measures in their country of domicile as the control group. Secondly, even though tax law is, in principle, a state matter, the European Union has established harmonized standards between Member States in the field of corporate taxation, especially in matters of international taxation (European Union, 2024). The European single market creates a certain comparability of large companies within the EU, ensuring arguably sufficient homogeneity between the observations of treatment and control groups. Furthermore, the reform does not contain other provisions that directly affect the usage of external debt which could distort a potential treatment effect.

I assign the treatment classification based on a firm's country of domicile. The 600 largest European listed firms are all multinationals which are not exclusively affected by their domestic tax system but also by all foreign tax systems with at least one affiliate present. Therefore, identifying a multinational firm's treatment status purely based on its country of domicile is an approximation, as the initial effect of the earnings stripping rule is determined at the level of the individual affiliates. The model by Huizinga et al. (2008, p. 93) states that the capital structure of multinational firms is dependent on the relative effective tax rates of all affiliates, weighted by their share of assets to firm-wide total assets. The study by Oh (2009, p. 339) on the regional activity of the biggest multinational European firms states that around 50 percent of their assets are located in their country of domicile. Therefore, it seems reasonable to assume that changes in the tax law in a European multinational's country of domicile have the single strongest effect on average on its capital structure, as opposed to changes in the tax

law of other countries with smaller affiliates present. This approach is in line with the research approach by De Mooij and Hebous (2018, p. 371), who also assign the treatment classification of consolidated firms based on their country of domicile.

Spillover effects are not fully ruled out in this research design and must be considered in the interpretation of the results. A corporate group based in a control group Member State may have affiliates in a Member State that is treated. Upon the reform, the affiliates in treated Member States are subject to stricter interest limitation rules and may decrease their debt ratio accordingly. The corporate group can not shift excess debt to other affiliates in the European Union, as those are already subject to equally effective interest limitation rules, which act as "backstop".[15] Hence, upon the reform, the consolidated debt ratio of firms in the control group should also decrease on average. This may introduce a downward bias to the treatment coefficient. I do not see potential spillover effects as a critical limitation of my research design. My study can not provide direct information on potential changes in firms' internal debt structures based on consolidated financial statements. Due to restrictions in the data sample, I was not able to analyze the effect of differences in the national implementation of the earnings stripping rule, for example, regarding the escape clauses.

6.1 Hypothesis 1 – General Effect

The Member States in the treatment group vary in their interest limitation rules prior to the reform. One can rightly assume that the effect of the reform may vary between countries based on their pre-reform interest limitation rule. In my first hypothesis, I want to abstract from the details of the pre-reform interest limitation rule characteristics and purely focus on the European Commission's classification as not equally effective.

The European Commission specifically states that existing measures are only comparably effective if they lead to a "similar or higher tax liability for a majority of large undertakings" (European Commission, 2018, p. 2). Conversely, this means that large undertakings in countries without equally effective regulations prior to the reform face a higher tax liability on average upon implementation of the reform, irrespective of their prior interest limitation regime. Therefore, in every treated Member State, the new earnings stripping rule can be seen as a stricter interest limitation rule, be it through an extension of the scope of application from internal to total interest, a stricter interest deduction limit, or by closing loopholes. The tax advantage of debt is driven by the possibility of deducting interest expenses from corporate income. As the interest deduction becomes limited above a certain threshold, debt loses its marginal tax-related advantage. The associated marginal costs of debt remain unaltered. Based on earlier theoretical and empirical research, the average multinational firm is expected to decrease its debt ratio so that the marginal benefit and marginal cost of debt are in equilibrium again (Buettner et al., 2012, p. 932; Buslei and Simmler, 2012, p. 3; DeAngelo and Masulis, 1980, p. 21; S. Myers, 1984, p. 4). This assumes that companies are not cash-constrained or at least able to substitute equity for debt if necessary.

Multinational firms may be able to substitute debt in other high-tax countries without an interest limitation rule in place. If the cross-border financing decisions of multinational firms were perfectly elastic, then there is no effect at the consolidated group level to be expected upon the reform as long as at least one affiliate is present in a country without an interest limitation rule and the same effective tax rate. However, tax restructuring, to a certain extent, may be unattractive due to high deadweight costs or for non-tax reasons (Huizinga & Laeven, 2008, p. 1166). Furthermore, as the new earnings stripping rule covers all European Member States and since many developed non-European countries have at least implemented some variant of interest limitation rule (see Table 15 in the Appendix), it seems rather unlikely that the average European multinational firm can fully circumvent an interest restriction purely through reallocation of debt. Nonetheless, tax restructuring is not fully ruled out and may decrease the observed effect of the reform on a consolidated firm level. This leads to my first hypothesis:

> H 1: Firms that were not subject to an equally effective interest limitation in their country of domicile before the reform decreased their external debt ratio to a greater extent than firms that were already subject to an equally effective interest limitation rule.

6.2 Hypothesis 2 – Pre-Reform Characteristics

My second hypothesis considers the characteristics of existing interest limitation rules prior to the reform. Unfortunately, the sample limits an in-depth analysis of the complex pre-reform heterogeneity in interest limitation rules. Instead, I focus on the type of interest expenses that were restricted, in line with earlier research by De Mooij and Hebous (2018, p. 369). Prior to the reform, a firm could be subject to a limitation of internal interest expenses, total interest expenses, or no limitation at all.

Firms based in Member States that already restricted total interest expenses through an earnings stripping rule prior to the reform should show the least change in their debt ratio. Firms in Member States that did not limit total interest expenses were not explicitly restricted in using external debt before the reform and are, therefore, expected to show a significant decrease in their debt ratio. Those firms could have been subject to no form of interest limitation at all or only a limitation of internal interest expenses.

Firms based in Member States that only restricted internal interest expenses are of particular interest. Prior liter-

[15] This assumes that all other affiliates are optimally financed at their debt ratio limit and that the heterogeneity between the new earnings stripping rules is not sufficient to incentivize strong debt shifting between Member States.

ature found that firms substitute external for internal debt upon the implementation of a thin capitalization rule for both related party and total debt-to-equity thresholds at the individual firm level (Buettner et al., 2012, p. 936; Wamser, 2014, p. 22). Therefore, firms under a thin capitalization rule should show a higher external debt ratio than firms under no interest limitation rule at all.[16] This assumption should also apply under an earnings stripping rule limiting only internal interest expenses. Indeed, the average pre-reform external debt ratio of firms in my sample subject to only an internal interest restriction before the reform is almost 5 percent (1.2 percentage points) higher than that of firms with no prior interest limitation. Therefore, firms that were subject to only a limitation of internal interest expenses are expected to show a particularly strong decline in their external debt ratio. Additionally, firms may substitute internal for external debt after the reform if they deem internal debt to be more advantageous and only deviated from their internal debt level preference due to the restriction pre-reform. This "reversal" of the original substitution effect would also reduce the debt ratio of the overall firm as internal loans are generally not observable at the consolidated group level but only applies if the Member State did not keep its specific restriction on internal interest expenses upon the reform. My second hypothesis is as follows:

> H 2: Firms in Member States without an explicit limitation of total interest expenses before the reform significantly decreased their external debt ratio upon the reform.

6.3 Hypothesis 3 – Sink and Conduit Tax Havens

Prior research found that tax havens play a key role in cross-border financing and earnings-stripping practices of multinational enterprises (Coppola et al., 2021, p. 1544; Fuest et al., 2022, p. 30; Gravelle, 2009, p. 732). In the first step of the classic earnings stripping scenario, a firm in a tax haven grants a loan to an affiliate located in a foreign high-tax country, which is refinanced through the external capital market or cash holdings at the level of the tax haven. In line with the OECD's (2022, p. 67) transfer pricing guidelines, firms in tax havens should show a slightly positive net interest income on risk-free intra-group financing activities, as the interest received from the foreign affiliate has to equal the interest payments of the refinanced loan plus a risk-free rate of return. If firms in tax havens are generally net creditors, is there any relevance to interest limitation rules in tax havens? I want to highlight why the reaction of tax havens to the earnings stripping rule reform might be more complex than apparent at an initial glance, based on the categorization of tax havens into sink and conduit tax havens by Garcia-Bernardo et al. (2017, p. 6).

Garcia-Bernardo et al. (2017, p. 2) define sink tax havens as countries that attract and, most importantly, retain foreign capital. Traditional sink tax havens are an attractive final target location for capital and future cash flows, not least due to their generally low effective tax rates (Hines & Rice, 1994, p. 150). I propose that the relevance of interest limitation rules in tax havens is driven not by debt shifting but rather by other measures of profit shifting, specifically the allocation of intangibles. Prior research suggests that the relocation of intangible assets, in combination with transfer pricing manipulation, is one of the main drivers behind profit-shifting (Dischinger and Riedel, 2011, p. 693; Grubert, 2003, p. 239; Langenmayr and Reiter, 2022, p. 828; Tørsløv et al., 2023, p. 1518). An interest limitation rule in a tax haven restricts the possibility, or at least makes it more costly, to finance the relocation of intangibles, which is subject to transfer pricing regulations (OECD, 2022b, p. 245). This assumption is in line with the observation that none of the European sink tax havens had a codified interest limitation rule in place prior to the reform as well as their tendency to implement rather lenient versions of the earnings stripping rule. The earnings stripping rule reform may directly target one of the main advantages of sink tax havens: the ability to relocate future profits at a low cost. As traditional European sink tax havens introduced a less strict earnings stripping version on average, the question is whether the least strict version of the new earnings stripping rule can still significantly affect firms in sink tax havens. Given that the goal of the European Union was to set a minimum standard that effectively decreases the tax-related advantages of debt, I expect firms based in European sink tax havens to lower their external debt ratio. Following Garcia-Bernardo et al. (2017, p. 6), I define Luxembourg and Malta as sink tax havens.

Conduit tax havens are defined as tax havens that reroute profits from high-tax countries to their final offshore financial location (OFC) (Garcia-Bernardo et al., 2017, p. 2; Tørsløv et al., 2023, p. 1527). Conduit tax havens are attractive to foreign investors due to their comprehensive treaty networks, low withholding taxes, tax rulings, tax reliefs, and typically low or zero taxes on capital transfers (Garcia-Bernardo et al., 2017, p. 2; Lejour, 2021, p. 8). Firms can avoid the application of tax avoidance rules by routing profits through a conduit tax haven first, instead of directly to the sink tax haven, if they show a low level of real activity in the conduit tax haven (Tørsløv et al., 2023, p. 1527). Therefore, I expect firms in conduit tax havens to be net creditors on average and not as strongly affected as firms in sink tax havens. Following Garcia-Bernardo et al. (2017, p. 6), I define the Netherlands, the United Kingdom, Ireland, and Belgium[17] as conduit tax havens. For my third hypothesis, I formulate:

> H 3: Firms based in sink tax havens decreased their external debt ratio significantly upon the reform. Firms based in conduit tax havens showed a smaller decrease in their external debt ratio than firms in sink tax havens.

[16] Nonetheless, their total debt ratio should be lower, as the substitution between external and internal debt is imperfect (Desai et al., 2004, p. 20).

[17] Belgium is usually categorized as small conduit tax haven (Garcia-Bernardo et al., 2017, p. 6; Tørsløv et al., 2023, p. 1506).

6.4 Hypothesis 4 – Industry-Specific Effects

The new earnings stripping rule generally restricts a firm's interest expenses as they exceed the 30 percent net-interest-to-EBITDA threshold. Therefore, one would expect the relevance of the reform to vary based on a firm's pre-reform characteristics. A firm without any interest expenses is not affected by the reform, whereas firms above or close to the threshold need to consider the implications of the reform on their capital structure. I focus on industry characteristics instead of firm-level characteristics to avoid biased results. Naturally, leverage between different industries varies.[18] Industries that have naturally higher net-interest-to-EBITDA ratios should be affected more strongly by the new regulation and are expected to decrease their debt ratio above average. Notably, the threshold is applied at the individual firm level in the first step. Therefore, the net-interest-to-EBITDA ratio of the consolidated group is only an approximation of the actual effect. I do not see this as a critical limitation, as a high average group-wide net-interest-to-EBITDA ratio indicates that the average subsidiary has a high net-interest-to-EBITDA ratio. Lastly, I propose that:

> H 4: Firms in industries with an above-average net-interest-to-EBITDA ratio before the reform decreased their external debt ratio more strongly than those with a below-average net-interest-to-EBITDA ratio.

7 Research Design

As explained in the introduction to my hypotheses, I divide the Member States into a treatment group and a control group based on the country-level classification provided by the European Commission (2018, p. 1). Only firms in Member States that did not have an equally effective interest limitation rule prior to the reform are expected to adjust their capital structure. Consistent with the discussion in the previous section, I postulate that the average firm's debt ratio would have changed at the same rate across the control and treatment groups in the absence of the reform. I provide visual confirmation for the parallel trend assumption in Section 8. Consistent with earlier research on the interaction between taxes and the capital structure on a consolidated firm level, I rely on the external-debt-to-total-asset (D/A) ratio as a measure for financial leverage as my dependent variable (Blouin et al., 2014, p. 35; De Mooij and Hebous, 2018, p. 374). The baseline equation is the following:

$$\begin{aligned}\frac{D}{A}_{it} &= \beta_0 + \beta_1 Treat_i + \beta_2 Reform_t \\ &\quad + \beta_3 Treat_i \cdot Reform_t + \beta_4 STRx_{ct} \\ &\quad + \lambda_t + \mu_i + \varepsilon_{it}\end{aligned} \tag{1}$$

$Treat_i$ is a dummy variable and indicates whether a firm is treated based on its country of domicile and turns one (zero) if the firm is in the treatment (control) group. $Reform_{it}$ is a dummy variable indicating the timing of the reform and equals zero pre-reform (2014-2018) and one for every observation starting in the year 2019 until 2022. The $Treat_i \cdot Reform_{it}$ interaction term identifies each firm observation that has been treated by the reform. The coefficient of the interaction term is the variable of interest, which indicates if and how the average treated firm adjusts its debt ratio upon the reform compared to non-treated firms. A significant and negative coefficient would be in line with my general hypothesis. $STRx_{ct}$ is the combined statutory tax rate in country c at time t.[19] As I analyze consolidated financial statements of multinational firms, the combined statutory tax rate in a firm's country of domicile is admittedly an imperfect tax measure (De Mooij & Hebous, 2018, p. 371). I do not resort to average effective tax rates as they may also reflect the effect of the reform.

Due to the long observation period, I include both firm (μ_i) and time (λ_t) level fixed effects to avoid biased results based on unobserved factors. The time-fixed effects cover general trends that are constant over all observations and vary over time. The firm fixed effects are particularly important and reflect time-constant effects that influence a firm's baseline debt ratio, which varies strongly. The firm fixed effects capture the $Treat_i$ dummy and the time fixed effects the $Reform_t$ dummy due to collinearity. Therefore, all time and firm-constant variables are not identified and thus not reported. The final regression equation with additional controls is:

$$\begin{aligned}\frac{D}{A}_{it} &= \beta_0 + \beta_1 Treat_i + \beta_2 Reform_t \\ &\quad + \beta_3 Treat_i \cdot Reform_t + \beta_4 STRx_{ct} \\ &\quad + \beta_5 Tangibility_{it} + \beta_6 ROA_{it} \\ &\quad + \beta_7 lnAssets_{it} + \beta_8 Loss_{it} \\ &\quad + \beta_{10} Inflation_{ct} + \lambda_t + \mu_i + \varepsilon_{it}\end{aligned} \tag{2}$$

$Tangibility_{it}$ depicts a firm's property, plant & equipment over its total assets and is a measure for collateral that may affect lending conditions. I expect a positive coefficient for $Tangibility_{it}$ as it improves lending conditions and incentivizes debt (Rajan & Zingales, 1995, p. 1451). Tangible assets could also provide a non-debt tax shield through higher depreciation, which may reduce taxable income and decrease the advantage of debt in certain scenarios (DeAngelo & Masulis, 1980). However, I expect the general advantage in improved lending conditions to outweigh the adverse effect on leverage incentives due to non-tax shields, which is situational. The ROA_{it} ratio measures a firm's profitability and consists of EBITDA over total assets. On the one hand,

[18] For details on industry specific differences in leverage, see Bowen et al. (1982).

[19] Some countries such as Germany and Luxembourg have additional subcentral government corporate income taxes which are included in this specification.

profitable firms may have easier access to credit, hence increasing their leverage (Blouin et al., 2014, p. 11). On the other hand, firms with higher profitability can pay down their debt faster and are less reliant on external financing to realize investments (see also: pecking order theory (Donaldson, 1961; S. C. Myers and Majluf, 1984)). The direction of the coefficient is not ex-ante clear. I use $lnAssets_{it}$ as the natural logarithm of total assets, which would otherwise be heavily left-skewed as an indicator for firm size (Buslei and Simmler, 2012, p. 13; Carrizosa et al., 2023, p. 25).[20] Larger firms may be less prone to bankruptcy and are more likely to have access to different capital markets. Hence, $lnAssets_{it}$ is expected to show a positive coefficient.

I incorporate $Loss_{it}$ as a dummy variable that turns one (zero) if a firm has less than zero (at least zero) EBITDA as an approximation. Loss-making firms are likely to react differently to an interest limitation as tax deductions are not immediately relevant in a loss situation (Alberternst & Sureth-Sloane, 2016, p. 26). However, due to the possibility of both loss-carry forwards and interest expense carry forwards they may become relevant in the future. Furthermore, the variable is only an approximation as it indicates a loss at the consolidated group level for financial reasons which does not out rule that individual affiliates might be profitable and subject to taxes. Due to the complexity and potential varying reasons for a loss situation, the coefficient is not ex-ante clear. The existence of a loss carryforward would also be of interest. Unfortunately, I cannot access information on the loss carry forward positions of firms in my sample. Lastly, I include $Inflation_{ct}$ as a country-specific variable that approximates the change in interest rates (Blouin et al., 2014, p. 11; De Mooij and Hebous, 2018, p. 373). The coefficient is expected to be negative as higher interest payments increase the marginal cost of debt, especially under an earnings stripping rule.

8 Data and Descriptive Statistics

For the empirical analysis, I rely on the Thomson Reuters database for information on the consolidated financial statements of the STOXX Europe 600 constituents in the period of 2014 to 2022. It is theoretically possible that the reform in 2019 affected the composition of the index, in addition to the general fluctuation. Therefore, the sample is based on the end-of-2018 index constituents to avoid a biased sample and to ensure a common trend analysis based on the same set of firms over the observation period. The empirical results are robust to a sample based on end-of-2017 constituents as well as the combination of all unique constituents from 2017 to 2020. Self-selection into the treatment and control group was ruled out by comparing whether firms changed their country of domicile between 2018 and 2019, which could have affected their treatment classification. All financial statements in the sample follow the IFRS accounting standards and are generally comparable. The information from the Thomson Reuters database is augmented with country-level data on the combined statutory tax rates from the OECD tax database (OECD, 2024b) as well as details on country-specific inflation rates obtained by the World Bank (World Bank, 2024). Finally, these are combined with the information I collected about the interest limitation rules in the EU before and after the reform.

The analysis is based on the period from 2014 to 2022 for three main reasons. Firstly, this enables enough pre-reform observations to establish a robust common trend assumption. Secondly, no distortions should be expected by the direct aftermath of the European debt crisis due to the restriction to years from 2014 onwards (European Parliament, 2019, p. 1). Thirdly, starting the observation period in 2014 excludes distortions caused by prior interest rule changes in the treatment and control groups.[21] The observation period is not extended further into the present as the data on financial statements and additional control variables are incomplete from 2023 onwards at the time of this study.

The original sample with 600 firms is adjusted to ensure consistency in the research design. The STOXX Europe 600 index lists corporations from the European Economic Area as well as Switzerland. Since the research questions specifically focus on the earnings stripping rule reform based on the Directive in the European Union, only observations on firms based in a European Member State are kept. In the second step, all countries that fit neither the definition of the treatment nor the control group are dropped. Those are Latvia, Poland, and Romania, which voluntarily implemented the reform in 2018, and Austria and Ireland, which delayed their implementation until 2021 and 2022, respectively. Additionally, all observations based in the United Kingdom are dropped due to its implementation of the reform already in 2017 and the distortions that may arise due to its exit from the European Union (DESTATIS, 2023). In line with prior research, firms in the financial industry, based on their Industry Classification Benchmark (ICB), are excluded due to their particular capital structure (Buettner et al., 2012, p. 934; Buslei and Simmler, 2012, p. 15) and because most Member States explicitly exclude the financial sector from the application of the earnings stripping rule. Additionally, observations without information on key variables such as the year of the financial statement, EBITDA, total debt, total assets, and property, plant, and equipment (PPE) are dropped. The final sample stands at 298 individual firms in 2018 and 2,631 observations in the observation period from 2014 until 2022. All variables are winsorized at the 0.25 percent level.

Table 7 illustrates the number of firms in the treatment and control group in the pre- and post-reform periods, their key financial variables, and the assigned country-level data.

[20] Another popular control for firm size is the logarithm of sales (Blouin et al., 2014, p. 935; Buettner et al., 2018, p. 563). However, around 10 percent of the observations in my sample do not report net sales. The results are robust if I use ln(Sales) instead of ln(Assets), see Section 10.

[21] I have cross checked prior literature on interest limitation rules (De Mooij and Hebous, 2018, p. 370; Gresik et al., 2017, p. 70) with country specific news on interest limitation rule changes from 2013 onwards.

Table 7: Summary Statistics

Notes: Table 7 illustrates the summary statistics for pre- and post-reform observations separately reported for the treatment and control groups. The table reports statistics on the mean, standard deviation, and 1 and 99 percentiles of the variable of interest and all control variables. Variables are defined in Table 14 in the Appendix. Source: own calculations.

	Pre-reform (2014-2018): 1.477 observations							
	Treat=0 (Observations 802)				Treat=1 (Observations 675)			
Variable	Mean	Std. dev.	p1	p99	Mean	Std. dev.	p1	p99
DtA	0.260	0.158	0.000	0.657	0.252	0.170	0.000	0.872
STR	33.138	5.100	25.000	44.433	24.898	4.091	20.000	33.990
Tangibility	0.208	0.184	0.001	0.848	0.224	0.191	0.000	0.839
ROA	0.111	0.061	0.002	0.302	0.136	0.091	-0.066	0.492
ln(Assets)	23.257	1.388	19.935	26.354	23.347	1.535	19.256	26.374
Loss	0.010	0.099	-	-	0.015	0.121	-	-
Inflation	0.843	0.676	-0.500	1.956	0.831	0.682	-0.208	2.126
	Post-reform (2019-2022): 1.154 observations							
	Treat=0 (Observations 623)				Treat=1 (Observations 531)			
	Mean	Std. dev.	p1	p99	Mean	SD	p1	p99
DtA	0.304	0.150	0.010	0.694	0.276	0.166	0.005	0.950
STR	29.527	2.739	25.000	34.433	23.660	3.134	20.000	31.500
Tangibility	0.223	0.171	0.001	0.765	0.238	0.189	0.001	0.894
ROA	0.100	0.061	-0.077	0.252	0.134	0.084	-0.028	0.450
ln(Assets)	23.644	1.263	20.879	26.446	23.747	1.423	20.283	26.473
Loss	0.029	0.168	-	-	0.021	0.143	-	-
Inflation	2.507	2.381	-0.323	8.391	3.136	3.144	-0.138	10.001

The treatment and control group are remarkably similar in all financial statement categories, which provides suggestive evidence that the groups are generally comparable. The biggest deviation can be observed in ROA, where firms in the treatment group are, on average, 23 percent more profitable than those in the control group, both pre- and post-reform. In both groups, the debt-to-asset ratio is higher in the post-reform period. From 2019 onwards, the new IFRS 16 standard introduced stricter rules for recognizing leases and related liabilities in the balance sheet which was expected to increase the debt ratio for financial statements applying IFRS significantly (IFRS, 2016, p. 3). The treatment group debt ratio increases less (+ 9.5 percent) than the control group (+ 16.9 percent), which is in line with the general hypothesis that the earnings stripping rule reform has an adverse effect on the relative leverage of the treatment group. Furthermore, the average statutory tax rate in the control group is significantly higher than in the treatment group. This is in line with the observation that countries with high corporate tax rates, on average, already introduced strict interest limitation rules to safeguard their tax base before the reform.

Figure 1 depicts the mean debt-to-asset ratio development for both the treatment and control groups over the observation period for a visual confirmation of the common trend assumption. Besides a deviation from 2015 to 2016, the control and treatment groups showed a strong parallel development before the reform in 2019. Estimation results are robust if I exclude the years 2014 and 2015, see Section 10. However, it also becomes apparent that the two groups are different regarding their baseline debt-to-asset ratio. If both groups were perfectly identical, one would expect the treated group to have a higher baseline debt ratio, as it is less restricted in its capital structure pre-reform. I do not see this as a major limitation for my research design, as my fixed effects strategy should cover the baseline difference. An additional two-sample t-test strongly indicates that the development of the mean debt ratio of both groups is not significantly different pre-reform.

Lastly, Figure 2 gives an overview of the composition of treatment and control groups based on the observations per country over the nine-year observation period. The final sample contains observations in 12 out of 27 Member States, as many countries do not have a firm present in the STOXX Europe 600 index. Therefore, one needs to be cautious when drawing a generalized conclusion, as firms in countries not represented in the sample could, in theory, react differently. I am nonetheless confident that the sample is diverse enough to obtain a substantiated estimation of the effect of the reform.

9 Empirical Results

This section presents the empirical results based on my hypotheses. I control for firm and time-fixed effects in all specifications of the following regressions. Standard errors are clustered at the country level (Abadie et al., 2022). The analysis and interpretation for Hypotheses H2 to H4 are based on the coefficients of subsamples of treated firms.

9.1 Hypothesis 1 – Regression Results

Table 8 presents my results on the estimated treatment effect of the reform on the external debt-to-asset ratio of

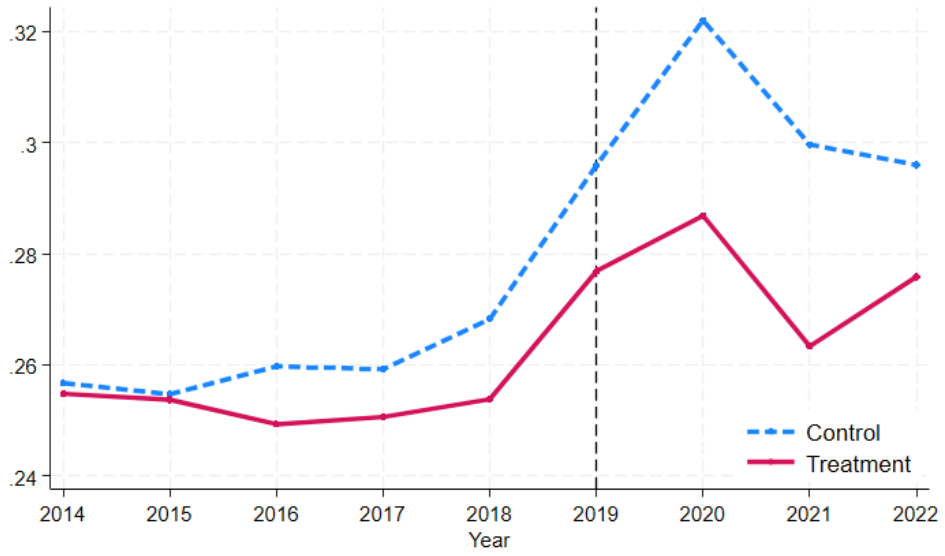


Figure 1: Average Debt to Asset Ratio per Year by Treatment Classification

Notes: Figure 1 illustrates the development of the mean debt-to-asset ratio over the observation period separately for the treatment and control group. The dashed black line indicates 2019 as the year of the reform. Source: own calculations.

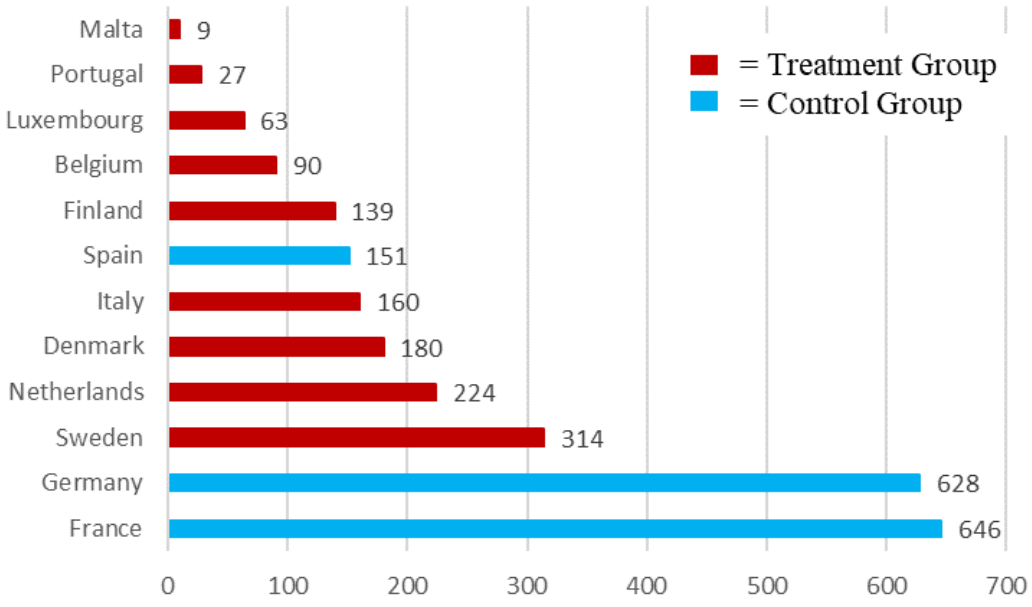


Figure 2: Country-Level Composition of Treatment and Control Group

Notes: Figure 2 gives an overview of the composition of the treatment and control groups based on observations per Member State. Source: own calculations.

firms in the treatment group compared to those in the control group. The first column shows the results of my basic regression without additional controls. The coefficient of the interaction term between treatment and reform is -0.019 and significant at the 5 percent level. Hence, the external debt ratio of treated firms decreased by 1.9 percentage points on average compared to firms in the control group. The combined statutory corporate tax rate coefficient is -0.001 and significant at the 5 percent level. If large consolidated multinational firms react to tax rate implications in their home country in line with the reactions of individual firms found in prior studies, the coefficient should be positive and sig-

nificant (Buettner et al., 2012, p. 935; Desai et al., 2004, p. 2452; Huizinga et al., 2008, p. 81). However, as already addressed in my research design, the domestic statutory tax rate is an imperfect measure of the actual tax effect on a consolidated multinational firm. I refrain from further interpretation as the tax coefficient varies slightly around zero in all further regressions and specifications. The adjusted R^2 varies between 0.8 and 0.9 in most regression specifications, which is higher than in related studies (Blouin et al., 2014, p. 35; De Mooij and Hebous, 2018, p. 374). This is not exceptionally surprising, as I run a firm-level fixed-effects strategy, which contributes strongly to the model's explanatory effect.

The second specification includes additional control variables and slightly alters the coefficient of my variable of interest. The coefficient is lower at -0.014 and only significant at the 10 percent level. Nonetheless, the direction of the effect is in line with my prediction that treated firms decrease their debt ratio upon the reform relative to non-treated firms. A 1.4 percentage points reduction in debt ratio would indicate a 5.5 percent decrease based on treated firms' pre-reform average external debt ratio. The result is lower than estimates reported by previous studies investigating the effects of interest limitation rules, which range between minus 6.3 and minus 21 percent (Blouin et al., 2014, p. 15; Buettner et al., 2012, p. 936; Buslei and Simmler, 2012, p. 20; De Mooij and Hebous, 2018, p. 374). This is unsurprising given that in my setup some Member States already had a limitation on total interest expenses in place prior to the reform, which is expected to decrease the average treatment effect.

Table 8: Results of the Regression for Hypothesis H1

Notes: Table 8 outlines the estimated effect of the earnings stripping rule reform on the debt ratio of treated firms compared to those in the control group. The dependent variable it the total debt-to-asset ratio. Specification (1) presents the basic estimation results without additional controls. Specification (2) includes the additional controls STR, Tangibility, ROA, ln(Assets), Loss and Inflation. Variables are defined in Table 14 in the Appendix. * Significant at 10-percent level. | ** Significant at 5-percent level. | *** Significant at 1-percent level. Source: own calculation.

	(1)	(2)
Dependent variable	DtA	DtA
Treat x Reform	−0.019**	−0.014*
	(0.008)	(0.008)
STR	−0.001**	−0.001**
	(0.000)	(0.001)
Tangibility		0.475***
		(0.058)
ROA		−0.282***
		(0.045)
ln(Assets)		0.050***
		(0.013)
Loss		0.035*
		(0.019)
Inflation		0.000
		(0.001)
Observations	2631	2631
Treated Firms	136	136
Adjusted R^2	0.83	0.85

Tangibility, as the ratio of PP&E to total assets, is significant at the 1 percent level and positive in this and all further regressions. The positive coefficient is in line with the expectation that firms with a higher share of potential collaterals have better access to credit and are, therefore, more heavily indebted. ROA as a measure of profitability has a strongly significant coefficient of -0.282 and indicates that more profitable firms indeed tend to pay off debts instead of applying for more credit, thus providing suggestive evidence for the pecking order theory. The coefficients are directionally identical, albeit smaller in magnitude, to those found by Blouin et al. (2014, p. 35) and De Mooij and Hebous (2018, p. 374). Firm size seems to have an influence on borrowing conditions or the preference for leverage as ln Assets has a positive coefficient of 0.05, significant at the 1 percent level. Studies that rely on other indicators for firm size, such as sales, also find a significant positive correlation between firm size and debt ratio (Blouin et al., 2014, p. 35; Buettner et al., 2012, p. 936). The coefficient for the dummy variable Loss is positive and significant at the 10 percent level. This aligns with the ROA coefficient and indicates that firms pay off their debt more quickly when they are profitable. The Inflation coefficient is zero and insignificant in this specification. The time-fixed effects strategy might cover most of the effect since it should reflect changes in the European Central Bank's official interest rates, which strongly affect country-specific inflation and interest rates.

9.2 Hypothesis 2 – Regression Results

The regression results in Table 9 are based on a sample split where treated firms are allocated to one of the subsamples based on the type of interest that was restricted before the reform in their country of domicile. Additionally, I include information on the treated firms and their country of domicile in each subsample. In line with my prediction, column (1) indicates no significant effect on the external debt ratio of firms that were already subject to a limitation of total interest before the reform. The coefficient of zero is surprisingly low. Portuguese firms are the only firms to show a positive treatment coefficient which is an anomaly I can not explain based on other changes in their tax law. Both Denmark and Italy show a slightly negative coefficient on an individual country basis, in line with the expectation that the reform implemented a stricter limitation of total interest.

The estimated coefficient of the treatment effect in column (2) suggests that firms that were not restricted in total interest expenses before the reform reacted particularly strongly, with a 2.1 percentage point reduction of their external debt ratio, significant at the 5 percent level. A 2.1 percentage point reduction in the external debt ratio indicates an 8.3 percent decrease based on the pre-reform average external debt ratio of treated firms of 25.2 percent. The coefficient is slightly lower at 1.3 percentage points if I exclude French firms from the control group, but equally significant, see Section 10. The observed effect aligns with Hypothesis H2 and indicates that the earnings stripping rule reform significantly

Table 9: Results of the Regression for Hypothesis H2

Notes: Table 9 outlines the estimated effect of the earnings stripping rule reform on treated firms based on the type of interest that was limited in their country of domicile pre-reform. Specification (1) presents the results based on the subsample which was subject to a limitation of total interest before the reform. Specification (2) presents the results based on the subsample, which was not subject to a limitation of total interest before the reform. Specification (3) contains only those firms from column (2) that were subject to no interest limitation rule at all before the reform. Specification (4) shows the effect on firms that were subject to a restriction of internal interest before the reform.
Variables are defined in Table 14 in the Appendix.
* Significant at 10-percent level. | ** Significant at 5-percent level. | *** Significant at 1-percent level.
Source: own calculations.

	(1)	(2)	(3)	(4)
Dependent variable	DtA	DtA	DtA	DtA
Treated Subsample	prior limitation on total interest	no prior limitation on total interest	no prior limitation	prior limitation on internal interest only
Treat x Reform	0.000	−0.021**	-0.016	−0.021**
	(0.008)	(0.007)	(0.013)	(0.005)
STR	−0.002**	−0.001**	−0.002**	−0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Tangibility	0.445***	0.510***	0.454***	0.552***
	(0.068)	(0.050)	(0.062)	(0.037)
ROA	−0.327***	−0.260***	−0.324***	−0.269**
	(0.037)	(0.059)	(0.032)	(0.073)
ln(Assets)	0.056***	0.048**	0.043**	0.058***
	(0.012)	(0.014)	(0.015)	(0.014)
Loss	0.047***	0.038	0.053***	0.041*
	(0.008)	(0.022)	(0.003)	(0.018)
Inflation	0.000	0.001	0.001	0.001
	(0.002)	(0.001)	(0.001)	(0.002)
Observations	1792	2264	1721	1968
Treated Firms	41	95	34	61
Countries	DK, IT, PT	BE, LU, MT, NL, SE	LU, MT, NL	BE, FI, SE
Adjusted R^2	0.85	0.86	0.85	0.86

affected the debt bias of large multinational firms in the European Union. My coefficient is lower than the average 5 percentage points decrease in the external debt ratio found by De Mooij and Hebous (2018, p. 374). However, the authors include smaller and purely national groups, which will most likely not have the same flexibility in reallocating debt as the multinational firms in my sample. Firms in my sample may have partly substituted debt in countries outside of the European Union that do not have comparable interest limitation rules, such as Switzerland. The significant treatment effect in my analysis indicates that a complete substitution of external debt in other countries might be too costly or impossible due to non-tax reasons.

I split the subsample observed in column (2) into two more subsamples. Column (3) outlines the treatment effect on firms not subject to any interest limitation rule before the reform. The coefficient of -0.016 is not significant at conventional levels. Interestingly, the treatment group only contains firms in classical tax havens, such as Luxembourg, Malta, and the Netherlands. The observed heterogeneity in the reaction between those three countries is outlined under the regression results for Hypothesis H3. Due to the special composition of the treatment group, no general conclusions can be drawn on how the average firm, which was not subject to any interest limitation rule before the reform, reacts.

The estimated coefficient of the treatment effect in column (4) suggests that firms that were only restricted in internal interest expenses before the reform were the drivers of the effect observed in column (2). The sub-sample contains treated firms from Belgium, Finland, and Sweden. The results are robust on the individual country level and independent of the form of internal interest expense restriction before the reform. The estimation results could provide suggestive evidence for the substitution effects observed by prior studies, but the comparison to column (3) might be flawed due to its particular country composition (Buettner et al., 2012, p. 936; Wamser, 2014, p. 22). The coefficient is lower than that observed in earlier studies on the switch from a limitation of internal interest only to an earnings stripping rule, which range between 2 and 5 percentage points and consider both changes in internal and external debt ratios at the individual firm level (Alberternst and Sureth-Sloane, 2016, p. 22; Buslei and Simmler, 2012, p. 20; Dreßler and Scheuering, 2012, p. 21). Internal debt has been shown to be more responsive to changes in tax incentives (Desai et al., 2004, p. 23; Dreßler and Scheuering, 2012, p. 21). It is unsurprising that I find a smaller effect, given that I can only observe changes in the external debt ratio in my empirical setup. Furthermore, as noted earlier, the relocation of debt and tax restructuring might reduce the consolidated effect. Importantly, my results do not mean that the overall capital structure of firms in column (4) changed the most. Based

Table 10: Results of the Regression for Hypothesis H3

Notes: Table 10 outlines the estimated effect of the earnings stripping rule reform on treated firms based on their classification as a sink or conduit tax haven in columns (1) and (2). The dependent variable is the total debt-to-asset ratio. Specification (3) to (6) outline the reaction of each tax haven individually. Variables are defined in Table 14 in the Appendix.
* Significant at 10-percent level. | ** Significant at 5-percent level. | *** Significant at 1-percent level. Source: own calculations.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	DtA	DtA	DtA	DtA	DtA	DtA
Treatment Subsample	sink tax haven	conduit tax haven	Luxembourg	Malta	Belgium	Netherlands
Treat x Reform	-0.043***	-0.009	-0.044***	-0.038**	-0.021**	-0.004
	(0.006)	(0.007)	(0.006)	(0.007)	(0.004)	(0.005)
STR	-0.002***	-0.002**	-0.002**	-0.002***	-0.002***	-0.002**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tangibility	0.498***	0.464***	0.493***	0.515***	0.510***	0.464***
	(0.040)	(0.068)	(0.041)	(0.045)	(0.044)	(0.070)
ROA	-0.336***	-0.350***	-0.330***	-0.357***	-0.366***	-0.345***
	(0.031)	(0.021)	(0.033)	(0.021)	(0.023)	(0.021)
ln(Assets)	0.053**	0.043*	0.052**	0.055**	0.053**	0.044*
	(0.012)	(0.016)	(0.013)	(0.014)	(0.013)	(0.018)
Loss	0.054***	0.047***	0.054***	0.051***	0.047***	0.051***
	(0.004)	(0.006)	(0.004)	(0.003)	(0.007)	(0.003)
Inflation	0.000	0.000	0.000	0.001	0.000	0.000
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.000)
Observations	1497	1739	1488	1434	1515	1649
Treated Firms	8	36	7	1	10	26
Adjusted R^2	0.85	0.87	0.85	0.85	0.87	0.86

on the imperfect substitution between external and internal debt, I would expect to find the strongest reaction of firms in column (3) if I could observe changes in both internal and external debt. The control variables are consistent over the different subsamples and those found in the total sample.

9.3 Hypothesis 3 – Regression Results

Table 10 presents the results from estimating Hypothesis H3 based on a firm's country of domicile and its classification as sink (column 1) or conduit tax haven (column 2). In line with my Hypothesis H3, sink tax havens show a highly significant reaction at the 1 percent level and a strong decrease of 4.3 percentage points in their external debt ratio. This could indicate that firms in sink tax havens rely on external debt to finance the relocation of intangibles from other group members. A restriction on total interest expenses in sink tax havens seems to reduce debt bias and potentially, indirectly, the attractiveness of profit shifting through intangibles. The strong reaction from firms in sink tax havens could indicate that the reform successfully decreased incentives for tax avoidance. However, only eight firms in my sample are based in a sink tax haven, seven in Luxembourg, and one in Malta. It would be interesting to see whether the results stay consistent with more treated observations and a broader sample of sink tax havens.

Column (2) presents the results for firms in conduit tax havens which show no significant reaction to the reform. This aligns with my argument that firms in conduit tax havens tend to show a positive net interest income due to their position as middlemen to other offshore financial locations. However, Belgium, as a conduit tax haven, also shows a significant decline in the external debt ratio of firms, albeit not as strong as those in sink tax havens. Belgium already employed a limitation of internal interest expenses prior to the reform so that this effect may be driven by the observations made in Hypothesis H2. Goyvaerts and Roggeman (2020, p. 6) define Belgium only as a borderline conduit tax haven, which could also explain the stronger reaction compared to the Netherlands. However, further interpretation based on only two different conduit tax havens would be highly speculative. The weak reaction of conduit tax havens is strongly influenced by firms in the Netherlands, which show no significant reaction to the reform.

9.4 Hypothesis 4 – Regression Results

Table 11 outlines my results on Hypothesis H4, which predicts different effects on industries based on their pre-reform net-interest-to-EBITDA ratios. The sample contains firms in 10 industries, based on the ICB industry classification, six of which I consider in this analysis. Columns (1) and (2) contain the estimation results of the three industries with the highest (lowest) net-interest-to-EBITDA ratio based on the average 2016 to 2018. I do not include the four industries in the midfield, as their ratios are very similar and their relative position dependent on their net-interest-to-EBITDA ratio changes frequently.

In column (1), the estimated coefficient of the treatment effect is -0.014 and significant at the 5 percent level. Surprisingly, the coefficient in column (2) is higher in magnitude, at -0.022, albeit with a notably higher standard error and insignificant at conventional levels. The results generally support Hypothesis H4 and suggest that firms closer to the threshold of the new earnings stripping rule show a stronger reaction to the reform on average. Columns (3) to (8) show

Table 11: Results of the Regression for Hypothesis H4

Notes: Table 11 outlines the estimated effect of the earnings stripping rule reform on treated firms based on their industry. Columns (1) and (2) show the effect of the reform on the three industries with the highest (lowest) net-interest-to-EBITDA average in the years 2016-2018. The dependent variable is the external debt-to-asset ratio. Specifications (3) to (5) outline the reaction of the 3 industries with the highest net-interest-to-EBITDA ratio. Specifications (6) to (8) outline the reaction of the 3 industries with the lowest net-interest-to-EBITDA ratio. Sequence in descending order. Variables are defined in Table 14 in the Appendix.
* Significant at 10-percent level. | ** Significant at 5-percent level. | *** Significant at 1-percent level. Source: own calculations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable	DtA	DtA	DtA	DtA	DtA	DtA	DtA	DtA
Subsample: industry	top 3	bot 3	Real Estate	Utilities	Telecomu.	Health Care	Consumer	Technology
Treat x Reform	-0.014**	-0.022	-0.014	-0.044**	-0.024	-0.002	-0.058*	-0.009
	(0.006)	(0.015)	(0.019)	(0.015)	(0.044)	(0.015)	(0.026)	(0.033)
STR	-0.002*	-0.004***	-0.004	0.002*	-0.003	-0.002	-0.002	-0.007***
	(0.001)	(0.001)	(0.003)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)
Tangibility	0.597***	0.543***	0.014	0.731***	0.347	-0.203	0.624***	0.358*
	(0.084)	(0.108)	(0.160)	(0.083)	(0.203)	(0.292)	(0.076)	(0.147)
ROA	0.003	-0.420***	-0.638*	-0.018	-0.073	-0.427**	-0.339	-0.274
	(0.199)	(0.112)	(0.301)	(0.295)	(0.439)	(0.177)	(0.201)	(0.184)
ln(Assets)	0.026**	0.072**	0.044	0.080*	0.002	0.051	0.122**	0.088**
	(0.011)	(0.030)	(0.047)	(0.035)	(0.031)	(0.040)	(0.040)	(0.033)
Loss	-0.016	0.058**		-0.001		-0.036	0.110*	
	(0.074)	(0.024)		(0.041)		(0.030)	(0.050)	
Inflation	0.002	-0.002	-0.001	0.009	-0.007	-0.009	0.001	0.000
	(0.004)	(0.003)	(0.005)	(0.008)	(0.009)	(0.007)	(0.006)	(0.005)
Observations	487	991	143	173	171	341	462	188
Treated Firms	23	42	5	8	10	21	15	6
Adjusted R^2	0.87	0.81	0.62	0.90	0.88	0.80	0.85	0.69

the estimation results for each industry separately from the left in descending order of their net-interest-to-EBITDA ratio. A trend is visible, according to which the reform's effect on the external debt ratio decreases as the average net-interest-to-EBITDA ratio per industry decreases, in line with my prediction. There are two outliers, the real estate and the consumer discretionary industry. Corporate groups in the real estate industry usually set up individual entities for each big project as to limit the risk of insolvency on the whole group. I see the insignificant coefficient as an indicator that the real estate industry is inherently able to exploit the de minimis threshold repeatedly, which is in line with prior observations from research and policymakers (Buslei and Simmler, 2012, p. 29; German Parliament, 2023b, p. 130). The strong reaction of firms in the consumer discretionary industry could indicate that other factors, such as an industry's traditional corporate group structure, may affect the usage of escape clauses. The results support concerns by De Mooij and Hebous (2018, p. 377) about the fairness of interest limitation rules that do not differentiate between industries.

10 Robustness Tests and Extensions

France voluntarily implemented the earnings stripping reform in 2019, which could potentially affect French firms in the control group if it influenced incentives for external debt. As France's pre-reform interest limitation regime was deemed at least equally effective, the reform could either make no difference or represent a relaxation of the existing interest limitation regulations if these were stricter than the minimum standard required by the Directive. France has kept its existing related party thin capitalization rule upon the reform but adjusted other aspects of its interest limitation regime. Prior to 2019, only 75% of interest expenses above 3 million net interest expenses were deductible for tax reasons under the "French rabot" restriction (Deloitte, 2018, p. 7). From 2019 onwards, the rule was transformed and now only applies to firms that use the equity or EBITDA financial group ratio escape (KPMG, 2019, p. 6). Firms that apply for the equity or EBITDA escape face the same interest restriction after the reform as they did before the reform. Firms that cannot use an escape clause and are above the new earnings stripping rule net-interest-to-EBITDA ratio are more strongly restricted in their interest expenses after the reform than before the reform, as interest deductions above 3 million are fully denied. Firms that are below the net-interest-to-EBITDA ratio of the new earnings stripping rule but above the 3 million net interest expense threshold face a less strict interest limitation regime after the reform. If the reform had relaxed the existing interest limitation regime for the average large multinational firm in France, French firms would have increased their debt ratio in 2019, leading to an overestimation of my coefficient of interest.

Table 12 outlines the main estimation results for hypotheses H1 to H4. Indeed, the general effect observed on the total sample was driven by the reaction of French firms, and it is not significant with only Germany and Spain in the control group. The estimation results for hypotheses H2 and H3 stay consistent, albeit smaller in magnitude. The standard error of the estimation results for Hypothesis H4 increases, but the

Table 12: Robustness Test of All Four Hypotheses Without France in the Control Group

Notes: Table 12 outlines the estimated effect of the earnings stripping rule, excluding firms based in France from the control group as a robustness test. Columns (1) to (9) repeat my main regression results based on the restricted control. Variables are defined in Table 14 in the Appendix.
* Significant at 10-percent level. | ** Significant at 5-percent level. | *** Significant at 1-percent level
Source: own calculations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable	DtA	DtA	DtA	DtA	DtA	DtA	DtA	DtA	DtA
Hypothesis	1	2				3		4	
Subsample	Full sample	Total ILR	No Total ILR	No prior ILR	Int. IRL	Sink haven	Conduit haven	Top 3 indust.	Bot. 3 indust.
Treat x Reform	-0.005	0.007	-0.013**	-0.008	-0.016**	-0.024**	0.000	-0.012	-0.004
	(0.006)	(0.008)	(0.005)	(0.007)	(0.004)	(0.004)	(0.002)	(0.007)	(0.009)
STR	0.001	0.001	0.001	0.004	-0.001	0.003	0.000	-0.003	0.002
	(0.001)	(0.002)	(0.002)	(0.003)	(0.001)	(0.003)	(0.001)	(0.002)	(0.004)
Tangibility	0.456***	0.388**	0.498***	0.401***	0.552***	0.463***	0.404**	0.592**	0.443***
	(0.072)	(0.085)	(0.064)	(0.076)	(0.048)	(0.031)	(0.093)	(0.095)	(0.123)
ROA	-0.281***	-0.346***	-0.249**	-0.336***	-0.254*	-0.346***	-0.367***	0.182	-0.459***
	(0.053)	(0.041)	(0.073)	(0.041)	(0.095)	(0.043)	(0.029)	(0.140)	(0.128)
ln(Assets)	0.046**	0.050**	0.044**	0.036	0.054**	0.048*	0.035	0.024**	0.083**
	(0.015)	(0.016)	(0.017)	(0.019)	(0.017)	(0.017)	(0.020)	(0.008)	(0.035)
Loss	0.026	0.041**	0.029	0.051***	0.035	0.053***	0.042**	-0.013	0.063**
	(0.029)	(0.013)	(0.038)	(0.005)	(0.031)	(0.006)	(0.013)	(0.106)	(0.025)
Inflation	-0.001	0.000	0.002	0.003	0.002	0.008	0.000	0.000	-0.002
	(0.003)	(0.004)	(0.002)	(0.005)	(0.003)	(0.004)	(0.002)	(0.005)	(0.006)
Observations	1985	1146	1618	1075	1322	851	1093	381	703
Adjusted R^2	0.85	0.84	0.85	0.85	0.85	0.84	0.88	0.88	0.80

relative effect of the reform between industries with different net-interest-to-EBITDA ratios is more in line with my prediction.

As stated in Section 6, the treatment coefficient is expected to suffer from a downward bias due to spillover effects in my research design. I kept Portuguese firms in the sample, the only ones showing a positive treatment coefficient, which may introduce downward bias if it is indeed an anomaly. I adjusted neither the countries in the control group nor the treatment group but stayed consistent with the classification provided by the European Commission. Accordingly, I decided to keep France in the control group as it does not change the relative effects found between the subsamples or the interpretation of my hypotheses.

Five Member States implemented the earnings stripping rule reform before or after 2019, three of which contain firms in the STOXX Europe 600. Table 13 shows the reaction of the external debt ratio of firms in Ireland (2022), the United Kingdom (2017), and Poland (2018) upon their respective reform years. Irish firms showed an average decrease of 2 percentage points in their external debt ratio, significant at the 1 percent level. Firms in the United Kingdom showed no significant reaction, which indicates that my classification as a conduit tax haven, which includes both Ireland and the United Kingdom, is not a consistent predictor. Additionally, the United Kingdom left the European Union on January 31, 2020, which may introduce distortions in the observation of British firms (DESTATIS, 2023).The coefficient of the treatment effect on Polish firms of -0.035 is high in magnitude but insignificant at conventional levels. The high coefficient is consistent with Hypothesis H2, which predicts that firms subject to only a limitation of internal interest expenses prior to the reform particularly strongly decreased their external debt ratio. The high standard deviation is unsurprising due to the small sample of only five Polish firms.

Table 16 in the Appendix contains additional regression specifications for each control variable. The results are not driven by one single control variable. Furthermore, I included ln(Sales) instead of ln(Assets) as a measure of firm size which does not significantly change the results but slightly lowers the number of observations. Table 17 in the Appendix contains three additional regression specifications that test for robustness in different observation periods. In the first specification, I only include observations from 2016 onwards to exclude the observed divergence of treatment and control groups from 2015 to 2016. In column (2) I only include observations in 2017 and 2019 to control for adjustment effects in the pre-reform year. The third specification includes observations only until 2019 to exclude any heterogeneity that might be induced by the COVID-19 pandemic or subsequent changes in the earnings stripping rule. The observed effect on the full sample is overall robust to changes in the observation period, albeit slightly lower in magnitude and significance.

11 Conclusion

This paper estimates the impact of the earnings stripping rule reform through the ATAD Directive in 2019 on the firm-wide external debt ratio of large multinational firms in the European Union. The analysis is based on the consolidated financial statements of listed firms in 12 different Member States between 2014 and 2022.

Table 13: Extended Regression Results for Ireland, United Kingdom, and Poland

Notes: Table 13 outlines additional regression results for the reform effect in Ireland, United Kingdom and Poland based on each country's individual implementation date. Variables are defined in Table 14 in the Appendix.
* Significant at 10-percent level. | ** Significant at 5-percent level. | *** Significant at 1-percent level
Source: own calculations.

	(1)	(2)	(3)
Dependent variable	DtA	DtA	DtA
Treated Subsample	Ireland	United Kingdom	Poland
Treat x Reform	-0.020***	-0.010	-0.035
	(0.002)	(0.007)	(0.021)
STR	-0.002***	-0.001*	-0.002***
	(0.000)	(0.001)	(0.001)
Tangibility	0.506***	0.529***	0.513***
	(0.033)	(0.037)	(0.041)
ROA	-0.277*	-0.406***	-0.282***
	(0.091)	(0.049)	(0.055)
ln(Assets)	0.052**	0.046***	0.057***
	(0.012)	(0.007)	(0.007)
Loss	0.057***	0.080***	0.059***
	(0.008)	(0.019)	(0.019)
Inflation	0.001	0.000	0.002
	(0.002)	(0.004)	(0.003)
Observations	1514	1968	1792
Treated Firms	11	117	5
Adjusted R^2	0.84	0.77	0.85

The estimation results indicate that firms not subject to a limitation of total interest expenses before the reform decreased their external debt ratio by 2.1 percentage points, which indicates an 8.3 percent decrease based on their pre-reform average external debt ratio. The effect is lower at 1.3 percentage points but equally significant if I exclude French firms from the control group. I find a particularly consistent and strong reaction from firms that were only restricted in their internal interest expenses before the reform, which may indicate the "reversal" of the substitution effect between external and internal debt found in prior studies (Buettner et al., 2012, p. 936; Wamser, 2014, p. 22). I assume tax havens to be net creditors in aspects of debt shifting. I postulate that sink tax havens rely on external debt to finance the relocation of intangible assets, which has been shown to be the main channel for profit shifting. In line with my prediction, sink tax havens show a significant decline of 4.3 percentage points in their external debt ratio upon the reform. Lastly, my results suggest that firms in industries with naturally higher net-interest-to-EBITDA ratios significantly reduce their external debt ratio. I observe outliers, which may indicate that the accessibility of escape clauses varies between industries. The overall estimation results of this study suggest that the earnings stripping rule reform in the European Union significantly affected the debt bias of large multinational firms.

Since I can only observe changes in external debt based on consolidated financial statements, I can not directly conclude that the reform was effective against debt shifting. The reduction of external debt observed in firms located in sink tax havens may have indirectly decreased the profitability of relocating intangibles and, therefore, incentives for profit shifting. A more detailed understanding of the reform could be achieved by analyzing unconsolidated financial statements, which allow the observation of changes in internal and external debt and, thus, a more direct conclusion on its effect on tax avoidance. Unfortunately, due to limitations in the data set, I was not able to analyze the effect of heterogeneity in the national implementation of the reform between Member States. It would be interesting to determine to what extent escape clauses, especially based on the financial group level, incentivize tax restructuring, whether they significantly change the reform's effectiveness in general, and their effect on tax competition within the EU. An analysis of affiliates of EU Member States in Switzerland could give an insight into the potential redistribution effects of debt, as Switzerland has no explicit interest limitation rule. Future research may also examine potential adverse effects on investment due to an increase in the cost of capital (Fazzari et al., 1988, p. 186), which have been observed in the context of thin capitalization rules (Buettner et al., 2018, p. 564; De Mooij and Liu, 2021, p. 5).

Ironically, the European Commission (2016, p. 2) lists the legal certainty of the reform as one of its advantages, while its role model has been pending before the constitutional court in Germany since 2016 (2 BvL 1/16). Finally, I would like to express my doubts about whether a regulation whose effect varies between industries, potentially triggers double taxation, and is so deeply involved in firms' fundamental financing decisions can fulfill its self-proclaimed goal: fair and effective taxation (European Commission, 2016, p. 2).

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2022). When Should You Adjust Standard Errors for Clustering? *The Quarterly Journal of Economics*, *138*(1), 1–35. https://doi.org/10.1093/qje/qjac038

Alberternst, S., & Sureth-Sloane, C. (2016). Interest Barrier and Capital Structure Response. *WU International Taxation Research Paper Series No. 2015-09*. https://doi.org/http://dx.doi.org/10.2139/ssrn.2563572

Auerbach, A. J. (2002). Taxation and Corporate Financial Policy. In *Handbook of Public Economics* (pp. 1251–1292, Vol. 3). Elsevier. https://doi.org/10.1016/S1573-4420(02)80023-4

Bach, S., & Buslei, H. (2009). Empirische Analysen zur Zinsschranke auf Grundlage von Handelsbilanzdaten. *DIW Berlin Research Notes 30*.

Bachmann, C., Lahmann, A., & Schuler, C. (2015). The Impact of Thin-Capitalization and Earnings Stripping Rules in the EU-15 on the Tax Shield. *University of Leipzig*.

Barclay, M. J., & Smith, C. W. (2005). The Capital Structure Puzzle: The Evidence Revisited. *Journal of Applied Corporate Finance*, *17*(1), 8–17. https://doi.org/10.1111/j.1745-6622.2005.012_2.x

Bianchi, J. (2011). Overborrowing and Systemic Externalities in the Business Cycle. *American Economic Review*, *101*(7), 3400–3426. https://doi.org/10.1257/aer.101.7.3400

Blouin, J., Nicodème, G., Huizinga, H., & Laeven, L. (2014). Thin Capitalization Rules and Multinational Firm Capital Structure. *IMF Working Paper WP/14/12*. https://doi.org/10.2778/32784

Bowen, R. M., Daley, L. A., & Huber, C. C. (1982). Evidence on the Existence and Determinants of Inter-Industry Differences in Leverage. *Financial Management*, *11*(4), 10. https://doi.org/10.2307/3665227

Bradley, M., Jarrell, G. A., & Kim, E. H. (1984). On the Existence of an Optimal Capital Structure: Theory and Evidence. *The Journal of Finance*, *39*(3), 857–878. https://doi.org/10.1111/j.1540-6261.1984.tb03680.x

Broer, M. (2009). Ziele, Wirkungsweise und Steueraufkommen der neuen Zinsschranke. *Schmollers Jahrbuch*, *129*(3), 391–413. https://doi.org/10.3790/schm.129.3.391

Buettner, T., Overesch, M., Schreiber, U., & Wamser, G. (2009). Taxation and Capital Structure Choice—Evidence from a Panel of German Multinationals. *Economics Letters*, *105*(3), 309–311. https://doi.org/10.1016/j.econlet.2009.09.001

Buettner, T., Overesch, M., Schreiber, U., & Wamser, G. (2012). The Impact of Thin-Capitalization Rules on the Capital Structure of Multinational Firms. *Journal of Public Economics*, *96*(11-12), 930–938. https://doi.org/10.1016/j.jpubeco.2012.06.008

Buettner, T., Overesch, M., & Wamser, G. (2016). Restricted Interest Deductibility and Multinationals' Use of Internal Debt Finance. *International Tax and Public Finance*, *23*(5), 785–797. https://doi.org/10.1007/s10797-015-9386-8

Buettner, T., Overesch, M., & Wamser, G. (2018). Anti Profit-Shifting Rules and Foreign Direct Investment. *International Tax and Public Finance*, *25*(3), 553–580. https://doi.org/10.1007/s10797-017-9457-0

Buslei, H., & Simmler, M. (2012). The Impact of Introducing an Interest Barrier: Evidence from the German Corporation Tax Reform 2008. *DIW Berlin Discussion Paper No. 1215*. https://doi.org/10.2139/ssrn.2111316

Carrizosa, R., Gaertner, F. B., & Lynch, D. P. (2023). Debt and Taxes? The Effect of TCJA Interest Limitations on Capital Structure. *The Journal of the American Taxation Association*, *45*(2), 35–55. https://doi.org/10.2308/JATA-2021-010

Coppola, A., Maggiori, M., Neiman, B., & Schreger, J. (2021). Redrawing the Map of Global Capital Flows: The Role of Cross-Border Financing and Tax Havens. *The Quarterly Journal of Economics*, *136*(3), 1499–1556. https://doi.org/10.1093/qje/qjab014

De la Feria, R., & Dourado, A. P. (2008). Thin Capitalization Rules in the Context of the CCCTB. In *Common Consolidated Corporate Tax Base* (pp. 1–40). Linde Verlag. https://ssrn.com/abstract=2293447

De Mooij, R. A. (2012). Tax Biases to Debt Finance: Assessing the Problem, Finding Solutions. *Fiscal Studies*, *33*(4), 489–512. http://www.jstor.org/stable/24440192

De Mooij, R. A., & Hebous, S. (2018). Curbing Corporate Debt Bias: Do Limitations to Interest Deductibility Work? *Journal of Banking & Finance*, *96*, 368–378. https://doi.org/10.1016/j.jbankfin.2018.07.013

De Mooij, R. A., & Keen, M. (2016). Debt, Taxes, and Banks. *Journal of Money, Credit and Banking*, *48*(1), 5–33. https://doi.org/https://doi.org/10.1111/jmcb.12289

De Mooij, R. A., & Liu, L. (2021). At a Cost: The Real Effects of Thin Capitalization Rules. *Economics Letters*, *200*, 1–6. https://doi.org/10.1016/j.econlet.2021.109745

DeAngelo, H., & Masulis, R. W. (1980). Optimal Capital Structure under Corporate and Personal Taxation. *Journal of Financial Economics*, *8*(1), 3–29. https://doi.org/10.1016/0304-405X(80)90019-7

Deloitte. (2018). France - Global Guide to M&A Tax: 2018 Edition. https://www.taxand.com/wp-content/uploads/2018/08/France_Tax and_MA_guide_2018_v2.pdf

Desai, M. A., Foley, C. F., & Hines, J. R. (2004). A Multinational Perspective on Capital Structure Choice and Internal Capital Markets. *The Journal of Finance*, *59*(6), 2451–2487. https://doi.org/10.1111/j.1540-6261.2004.00706.x

DESTATIS. (2023). Europe - Brexit Monitor. https://www.destatis.de/Europa/EN/Topic/Brexit/_node.html

Devereux, M. P., Auerbach, A. J., Keen, M., Oosterhuis, P., Schön, W., & Vella, J. (2021). Destination-Based Cash Flow Taxation. In *Taxing Profit in a Global Economy* (1st ed., pp. 267–334). Oxford University Press. https://doi.org/10.1093/oso/9780198808060.003.0007

Dischinger, M., & Riedel, N. (2011). Corporate Taxes and the Location of Intangible Assets within Multinational Firms. *Journal of Public Economics*, *95*(7-8), 691–707. https://doi.org/10.1016/j.jpubeco.2010.12.002

Donaldson, G. (1961). *Corporate Debt Capacity: A Study of Corporate Debt Policy and the Determination of Corporate Debt Capacity*. Harvard University.

Dreßler, D., & Scheuering, U. (2012). Empirical Evaluation of Interest Barrier Effects. *ZEW Discussion Paper No. 12-046*. https://doi.org/10.2139/ssrn.2112117

Entwurf eines Unternehmensteuerreformgesetzes 2008. (2007). https://dserver.bundestag.de/btd/16/048/1604841.pdf

European Commission. (2007). Communication from the Commission to the Council, the European Parliament and the European Economic and Social Committee - The Application of Anti-Abuse Measures in the Area of Direct Taxation – within the EU and in Relation to Third Countries. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A52007DC0785

European Commission. (2016). COUNCIL DIRECTIVE (EU) 2016/1164. https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A32016L1164

European Commission. (2018). Commission Notice - Measures Considered Equally Effective to Article 4 of the Anti-Tax Avoidance Directive. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018XC1207(01)&from=EN

European Commission. (2019). November Infringements Package: Key Decisions. https://ec.europa.eu/commission/presscorner/detail/en/inf_19_6304

European Commission. (2022). Proposal for a Councel Directive on Laying Down Rules on a Debt-Equity Bias Reduction Allowance and on Limiting the Deductibility of Interest for Corporate Income Tax Purposes. https://taxation-customs.ec.europa.eu/taxation-1/debt-equity-bias_en

European Parliament. (2019). A Decade on from the Crisis - Main Responses and Remaining Challenges. https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/642253/EPRS_BRI%282019%29642253_EN.pdf

European Parliament. (2022). Assessment of Recent Anti-Tax Avoidance and Evasion Measures (ATAD & DAC 6). https://www.europarl.europa.eu/RegData/etudes/STUD/2022/703353/IPOL_STU(2022)703353_EN.pdf:~:text=such%20a%20rule%20prior%20to,KEY%20FINDINGS

European Union. (2024). Taxation. https://eur-lex.europa.eu/summary/chapter/taxation.html?root_default=SUM_1_CODED%3D21&locale=en

EY. (2023). EY Worldwide Corporate Tax Guide. https://www.ey.com/en_gl/technical/tax-guides/worldwide-corporate-tax-guide

Fazzari, S. M., Hubbard, R. G., Petersen, B. C., Blinder, A. S., & Poterba, J. M. (1988). Financing Constraints and Corporate Investment. *Brookings Papers on Economic Activity*, *1988*(1), 141. https://doi.org/10.2307/2534426

Feld, L. P., Heckemeyer, J. H., & Overesch, M. (2013). Capital Structure Choice and Company Taxation: A Meta-Study. *Journal of Banking & Finance*, *37*(8), 2850–2866. https://doi.org/10.1016/j.jbankfin.2013.03.017

Forbes. (2023). The Global 2000. https://www.forbes.com/lists/global2000/

Fuest, C., Hugger, F., & Neumeier, F. (2022). Corporate Profit Shifting and the Role of Tax Havens: Evidence from German Country-by-Country Reporting Data. *Journal of Economic Behavior & Organization*, *194*, 454–477. https://doi.org/10.1016/j.jebo.2021.11.016

G20. (2012). G20 Leaders Declaration. https://www.oecd.org/g20/summits/los-cabos/2012-0619-loscabos.pdf

Garcia-Bernardo, J., Fichtner, J., Takes, F. W., & Heemskerk, E. M. (2017). Uncovering Offshore Financial Centers: Conduits and Sinks in the Global Corporate Ownership Network. *Scientific Reports*, *7*(1), 6246. https://doi.org/10.1038/s41598-017-06322-9

German Parliament. (2007). Entwurf eines Unternehmensteuerreformgesetzes 2008.

German Parliament. (2023a). Wachstumschancengesetz Gesetzesentwurf. https://dserver.bundestag.de/btd/20/086/2008628.pdf

German Parliament. (2023b). Wachstumschancengesetz Gesetzesentwurf.

Goyvaerts, D., & Roggeman, A. (2020). The Impact of Thin Capitalization Rules on Subsidiary Financing: Evidence from Belgium. *De Economist*, *168*(1), 23–51. https://doi.org/10.1007/s10645-019-09353-x

Graham, J. R. (1996). Debt and the Marginal Tax Rate. *Journal of Financial Economics*, *41*(1), 41–73. https://doi.org/10.1016/0304-405X(95)00857-B

Gravelle, J. G. (2009). Tax Havens: International Tax Avoidance and Evasion. *National Tax Journal*, *62*(4), 727–753. http://www.jstor.org/stable/41790645

Gresik, T. A., Schindler, D., & Schjelderup, G. (2017). Immobilizing Corporate Income Shifting: Should It Be Safe to Strip in the Harbor? *Journal of Public Economics*, *152*, 68–78. https://doi.org/10.1016/j.jpubeco.2017.06.001

Grubert, H. (2003). Intangible Income, Intercompany Transactions, Income Shifting, and the Choice of Location. *National Tax Journal*, *56*(1.2), 221–242. https://doi.org/10.17310/ntj.2003.1S.05

Haufler, A., & Runkel, M. (2012). Firms' Financial Choices and Thin Capitalization Rules under Corporate Tax Competition. *European Economic Review*, *56*(6), 1087–1103. https://doi.org/10.1016/j.euroecorev.2012.03.005

Heckemeyer, J. H., & de Mooij, R. A. (2017). Taxation and Corporate Debt: Are Banks Any Different? *National Tax Journal*, *70*(1), 53–76. http://www.jstor.org/stable/26349715

Heckemeyer, J. H., & Overesch, M. (2017). Multinationals' Profit Response to Tax Differentials. *The Canadian Journal of Economics / Revue canadienne d'Economique*, *50*(4), 965–994. https://www.jstor.org/stable/48581982

Heider, F., & Ljungqvist, A. (2015). As Certain as Debt and Taxes: Estimating the Tax Sensitivity of Leverage from State Tax Changes. *Journal of Financial Economics*, *118*(3), 684–712. https://doi.org/10.1016/j.jfineco.2015.01.004

Hines, J. R., & Rice, E. M. (1994). Fiscal Paradise: Foreign Tax Havens and American Business. *The Quarterly Journal of Economics*, *109*(1), 149–182. https://doi.org/10.2307/2118431

Hodder, J. E., & Senbet, L. W. (1990). International Capital Structure Equilibrium. *The Journal of Finance*, *45*(5), 1495–1516. https://doi.org/10.1111/j.1540-6261.1990.tb03725.x

Huizinga, H., & Laeven, L. (2008). International Profit Shifting within Multinationals: A Multi-Country Perspective. *Journal of Public Economics*, *92*(5-6), 1164–1182. https://doi.org/10.1016/j.jpubeco.2007.11.002

Huizinga, H., Laeven, L., & Nicodeme, G. (2008). Capital Structure and International Debt Shifting. *Journal of Financial Economics*, *88*(1), 80–118. https://doi.org/10.1016/j.jfineco.2007.05.006

IFRS. (2016). IFRS 16 Leases - Effects Analysis. https://www.ifrs.org/content/dam/ifrs/project/leases/ifrs/published-documents/ifrs16-effects-analysis.pdf

IMF. (2009). Debt Bias and Other Distortions: Crisis-Related Issues in Tax Policy, IMF Policy Paper.

Irish Tax Institute. (2021). ATAD Implementation Article 4 Interest Limitation - Response to Feedback Statement. https://taxinstitute.ie/wp-content/uploads/2021/03/2021-03-08-Irish-Tax-Institute-Response-to-Feedback-Statement-on-ATAD-implementation-of-Article-4-Interest-Limitation.pdf

Kalamov, Z. Y. (2020). Safe Haven or Earnings Stripping Rules: A Prisoner's Dilemma? *International Tax and Public Finance*, *27*(1), 38–76. https://doi.org/10.1007/s10797-019-09545-w

Knauer, T., & Sommer, F. (2012). Interest Barrier Rules as a Response to Highly Leveraged Transactions: Evidence from the 2008 German Business Tax Reform. *Review of Accounting and Finance*, *11*(2), 206–232. https://doi.org/10.1108/14757701211228228

KPMG. (2017). Romania Country Profile. https://assets.kpmg.com/content/dam/kpmg/ro/pdf/tax-card-2017-en-web.pdf

KPMG. (2018). Bulgaria Country Profile. https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2018/06/country-profile-bulgaria-2018.pdf

KPMG. (2019). France Country Profile. https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2020/10/france-country-profile-2020.pdf

KPMG. (2020). Austria Country Profile. https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2020/10/austria-country-profile-2020.pdf

Kraus, A., & Litzenberger, R. H. (1973). A State-Preference Model of Optimal Financial Leverage. *The Journal of Finance*, *28*(4), 911. https://doi.org/10.2307/2978343

Langenmayr, D., & Reiter, F. (2022). Trading Offshore: Evidence on Banks' Tax Avoidance*. *The Scandinavian Journal of Economics*, *124*(3), 797–837. https://doi.org/10.1111/sjoe.12507

Lejour, A. M. (2021). The Role of Conduit Countries and Tax Havens in Corporate Tax Avoidance. *CentER Discussion Paper Nr. 2021-014*. https://doi.org/10.2139/ssrn.3843734

MacKie-Mason, J. K. (1990). Do Taxes Affect Corporate Financing Decisions? *The Journal of Finance*, *45*(5), 1471. https://doi.org/10.2307/2328746

Mardan, M. (2017). Why Countries Differ in Thin Capitalization Rules: The Role of Financial Development. *European Economic Review*, *91*, 1–14. https://doi.org/10.1016/j.euroecorev.2016.09.003

Marsh, P. (1982). The Choice Between Equity and Debt: An Empirical Study. *The Journal of Finance*, *37*(1), 121–144. https://doi.org/10.1111/j.1540-6261.1982.tb01099.x

Mayer, C. (1990). Financial Systems, Corporate Finance, and Economic Development. In *Asymmetric Information, Corporate Finance, and Investment* (pp. 307–332). National Bureau of Economic Research, Inc. https://EconPapers.repec.org/RePEc:nbr:nberch:11477

Miller, M. H. (1977). Debt and Taxes. *The Journal of Finance*, *32*(2), 261. https://doi.org/10.2307/2326758

Mintz, J., & Weichenrieder, A. J. (2005). Taxation and the Financial Structure of German Outbound FDI. *CESifo Working Paper No. 1612*. https://doi.org/10.2139/ssrn.875306

Modigliani, F., & Miller, M. H. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review*, *48*(3), 261–297. http://www.jstor.org/stable/1809766

Modigliani, F., & Miller, M. H. (1963). Corporate Income Taxes and the Cost of Capital: A Correction. *The American Economic Review*, *53*(3), 433–443. http://www.jstor.org/stable/1809167

Myers, S. (1984). *Capital Structure Puzzle* (tech. rep. No. w1393). National Bureau of Economic Research. https://doi.org/10.3386/w1393

Myers, S. C. (1977). Determinants of Corporate Borrowing. *Journal of Financial Economics*, *5*(2), 147–175. https://doi.org/10.1016/0304-405X(77)90015-0

Myers, S. C., & Majluf, N. S. (1984). Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not

Have. *Journal of Financial Economics*, *13*(2), 187–221. https://doi.org/10.1016/0304-405X(84)90023-0

OECD. (1979). Transfer Pricing and Multinational Enterprises. https://www.oecd-ilibrary.org/content/publication/9789264167773-en

OECD. (1987). Thin Capitalisation - Taxation of Entertainers, Artistes and Sportsmen. https://www.oecd-ilibrary.org/content/publication/5b556cb3-en

OECD. (1998). Harmfull Tax Competition - An Emerging Global Issue. https://www.oecd.org/ctp/harmful/1904176.pdf

OECD. (2013). Action Plan on Base Erosion and Profit Shifting. https://www.oecd.org/ctp/BEPSActionPlan.pdf

OECD. (2015a). Limiting Base Erosion Involving Interest Deductions and Other Financial Payments, Action 4 - 2015 Final Report.

OECD. (2015b). Designing Effective Controlled Foreign Company Rules, Action 3 - 2015 Final Report. https://www.oecd-ilibrary.org/content/publication/9789264241152-en

OECD. (2015c). Limiting Base Erosion Involving Interest Deductions and Other Financial Payments, Action 4 - 2015 Final Report. https://www.oecd-ilibrary.org/content/publication/9789264241176-en

OECD. (2017). Limiting Base Erosion Involving Interest Deductions and Other Financial Payments, Action 4 - 2016 Update. https://www.oecd-ilibrary.org/content/publication/9789264268333-en

OECD. (2019). Interest Limitation Rules (IRL). https://qdd.oecd.org/subject.aspx?Subject=ILR

OECD. (2020). Corporate Tax Statistics, Second Edition. https://www.oecd-ilibrary.org/taxation/corporate-tax-statistics-second-edition_ff4d4ce8-en

OECD. (2022a). Safe Harbours and Penalty Relief: Global Anti-Base Erosion Rules (Pillar Two), OECD/G20 Inclusive Framework on BEPS. https://www.oecd.org/tax/beps/safe-harbours-and-penalty-relief-global-anti-base-erosion-rules-pillar-two.pdf

OECD. (2022b). OECD Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations 2022. https://www.oecd-ilibrary.org/content/publication/0e655865-en

OECD. (2023a). Members of the OECD/G20 Inclusive Framework on BEPS. https://www.oecd.org/tax/beps/inclusive-framework-on-beps-composition.pdf

OECD. (2023b). Transfer Pricing Country Profiles. https://www.oecd.org/tax/transfer-pricing/transfer-pricing-country-profiles.htmtpcountryprofiles

OECD. (2024a). Controlled Foreign Company (CFC) Rules. https://qdd.oecd.org/data/CFC/.ALL

OECD. (2024b). OECD Tax Database. https://stats.oecd.org/Index.aspx?QueryId=78166

Oh, C. H. (2009). The International Scale and Scope of European Multinationals. *European Management Journal*, *27*(5), 336–343. https://doi.org/10.1016/j.emj.2008.12.002

Overesch, M., & Wamser, G. (2010). Corporate Tax Planning and Thin-Capitalization Rules: Evidence from a Quasi-Experiment. *Applied Economics*, *42*(5), 563–573. https://doi.org/10.1080/00036840701704477

Pazarbasioglu, C., Keen, M., & Claessens, S. (2010). Financial Sector Taxation The IMF's Report to the G-20 and Background Material. https://www.imf.org/external/np/seminars/eng/2010/paris/pdf/090110.pdf

PwC. (2024a). New Zealand - Worldwide Tax Summaries. https://taxsummaries.pwc.com/new-zealand/corporate/group-taxation

PwC. (2024b). Worldwide Tax Summaries Online. https://taxsummaries.pwc.com/

PwC. (2024c). Finland - Worldwide Tax Summaries. https://taxsummaries.pwc.com/finland/corporate/group-taxation

Rajan, R. G., & Zingales, L. (1995). What Do We Know about Capital Structure? Some Evidence from International Data. *The Journal of Finance*, *50*(5), 1421–1460. https://doi.org/10.1111/j.1540-6261.1995.tb05184.x

Ruf, M., & Schindler, D. (2015). Debt Shifting and Thin-Capitalization Rules – German Experience and Alternative Approaches. *Nordic Tax Journal*, *2015*(1), 17–33. https://doi.org/10.1515/ntaxj-2015-0002

Sikka, P., & Willmott, H. (2010). The Dark Side of Transfer Pricing: Its Role in Tax Avoidance and Wealth Retentiveness. *Critical Perspectives on Accounting*, *21*(4), 342–356. https://doi.org/10.1016/j.cpa.2010.02.004

Stiglitz, J. E. (1969). A Re-Examination of the Modigliani-Miller Theorem. *The American Economic Review*, *59*(5), 784–793. http://www.jstor.org/stable/1810676

Stiglitz, J. E. (1972). Some Aspects of the Pure Theory of Corporate Finance: Bankruptcies and Take-overs. *The Bell Journal of Economics and Management Science*, *3*(2), 458. https://doi.org/10.2307/3003033

Summers, L. H., Bosworth, B. P., Tobin, J., & White, P. M. (1981). Taxation and Corporate Investment: A q-Theory Approach. *Brookings Papers on Economic Activity*, *1981*(1), 67. https://doi.org/10.2307/2534397

Swenson, D. L. (2001). Tax Reforms and Evidence of Transfer Pricing. *National Tax Journal*, *54*(1), 7–25. https://doi.org/10.17310/ntj.2001.1.01

Tørsløv, T., Wier, L., & Zucman, G. (2023). The Missing Profits of Nations. *The Review of Economic Studies*, *90*(3), 1499–1534. https://doi.org/10.1093/restud/rdac049

Umbrecht, R. L., & Llewellyn, D. W. (1994). Planning Pitfalls and Opportunities for Foreign Owned Corporations under the Earnings Stripping Rules. *The Tax Lawyer*, *47*(3), 641–712. http://www.jstor.org/stable/20771630

Verfassungsmäßigkeit der sog. Zinsschranke - Billigkeitsmaßnahme. (2015). https://www.bundesfinanzhof.de/de/entscheidung/entscheidungen-online/detail/STRE201610031/

Wamser, G. (2014). The Impact of Thin-Capitalization Rules on External Debt Usage – A Propensity Score Matching Approach. *Oxford Bulletin of Economics and Statistics*, *76*(5), 764–781. https://doi.org/10.1111/obes.12040

Ways and Means Committee. (2016). A Better Way for Tax Reform. https://www.novoco.com/public-media/documents/ryan_a_better_way_policy_paper_062416.pdf

Weichenrieder, A. J., & Windischbauer, H. (2008). Thin-Capitalization Rules and Company Responses - Experience from German Legislation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1299533

Weyzig, F. (2013). Tax Treaty Shopping: Structural Determinants of Foreign Direct Investment Routed through the Netherlands. *International Tax and Public Finance*, *20*(6), 910–937. https://doi.org/10.1007/s10797-012-9250-z

World Bank. (2024). World Development Indicators - DataBank. https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://databank.worldbank.org/source/world-development-indicators&ved=2ahUKEwie5Lbbk6uFAxXmBdsEHc2ZDVAQFnoECBgQAQ&usg=AOvVaw2izQ-SvJ7A84W2Mu5znrYz

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Der Einfluss unterschiedlicher Expertengruppen auf die ertragsteuerliche Gesetzgebung in Deutschland: Eine Analyse von schriftlichen Stellungnahmen

Thomas Zimmermann

Universität Paderborn

Abstract

Despite the pivotal role of tax legislation for the economy and society, empirical research on how interest groups influence German tax legislation remains scarce. This study addresses this gap by examining international corporate tax policy. Based on 94 statements from eleven legislative procedures (2013 - 2023), it analyzes the factors determining whether demands from various interest groups during expert hearings in the Finance Committee are adopted into final legislation. Methodologically, the paper combines qualitative content analysis with a quantitative probit regression. While descriptive data and initial analyses suggest varying success rates among expert groups, these differences disappear when controlling for the specific direction of the proposed changes. No significant group influence remains. The findings suggest that in the later stages of the legislative process, lawmakers select external expertise primarily based on substantive, technical criteria rather than systematically favoring specific actors. Thus, the study contributes to the political economy of taxation by highlighting the importance of technical content over pure actor power.

Zusammenfassung

Trotz der zentralen Bedeutung des Steuerrechts für Wirtschaft und Gesellschaft fehlen in der Literatur bislang empirische Untersuchungen dazu, inwiefern Interessengruppen die deutsche Steuergesetzgebung tatsächlich beeinflussen. Die Arbeit versucht diese Forschungslücke am Beispiel der internationalen Ertragsteuerpolitik ein Stück weit zu schließen. Auf Basis von 94 Stellungnahmen aus elf Gesetzgebungsverfahren zwischen 2013 und 2023 wird analysiert, welche Faktoren darüber entscheiden, ob Forderungen verschiedener Gruppen nach Expertenanhörungen im Finanzausschuss in die finale Gesetzgebung übernommen werden. Methodisch kombiniert die Arbeit eine qualitative Inhaltsanalyse mit einer quantitativen Probit-Regression. Deskriptive Daten und erste Regressionsanalysen lassen zunächst Unterschiede im Erfolg der verschiedenen Expertengruppen vermuten. Sobald jedoch für die spezifische Änderungsrichtung der betrachteten Forderungen kontrolliert wird, kann kein signifikanter Einfluss der Gruppen festgestellt werden. Die Ergebnisse legen nahe, dass der Gesetzgeber im späteren Verlauf des Gesetzgebungsprozesses externe Expertise primär nach inhaltlichen Kriterien auswählt und keine Interessengruppe systematisch bevorzugt. Damit liefert die Untersuchung neue Erkenntnisse zur politischen Ökonomie der Besteuerung, indem sie die Bedeutung technischer Inhaltsmerkmale gegenüber reiner Akteursmacht hervorhebt.

Keywords: lobbyismus; steuergesetzgebung; interessengruppen; stellungnahmen; einfluss

1 Einleitung

1.1 Motivation

Die externe Einflussausübung auf politische Entscheidungsfindung hat eine große gesamtgesellschaftliche Relevanz und ist regelmäßig Teil des politischen Diskurses. Insbesondere in Fragen der Besteuerung besteht ein großes Interesse von allen Seiten, da jegliche Änderungen in dieser Hinsicht mittelbare Auswirkungen auf den Staatshaushalt, die Liquidität von Unternehmen und auch auf den persönlichen Haushalt jedes Einzelnen[1] entfalten. In diesem Zuge wird bei steuerlichen Aspekten gar über den “heiligen Gral” des Lobbyismus gesprochen (vgl. Barrick und Alexander, 2014, S. 43). Fragen der Steuerbelastung oder Steuergerechtigkeit tangieren alle Interessengruppen. Umso relevanter ist es zu erfahren, ob diese Gruppen die steuerliche Gesetzgebung im Land durch ihre Expertise beeinflussen können und ob zwischen diesen Gruppen Unterschiede in ihrer Einflussintensität bestehen. Die bisherige Lobbyliteratur

Ich danke Yuri Piper und Prof. Dr. Dr. h.c. Dr. h.c. Caren Sureth-Sloane für die hilfreiche und angenehme Betreuung der Masterarbeit, sowie den Teilnehmerinnen und Teilnehmern meines Kolloquiums für die konstruktiven Beiträge zum damaligen Arbeitsstand.

[1] In dieser Arbeit wird ausschließlich zum Zweck der besseren Lesbarkeit auf die geschlechtsspezifische Schreibweise verzichtet. Alle personenbezogenen Bezeichnungen sind geschlechts-neutral zu verstehen.

DOI: https://doi.org/10.5282/jums/v11i3pp671-691

© The Author(s) 2026. Published by Junior Management Science.
This is an Open Access article distributed under the terms of the CC-BY-4.0 (Attribution 4.0 International). Open Access funding provided by ZBW.

setzt vor allem bei Regierungsbeeinflussung an, da der größte Einfluss in frühen Stadien des Legislativprozesses zu erwarten ist (vgl. Sutton, 1984, S. 88; Yackee und Yackee, 2006, S. 137). Demzufolge werden bis zur Veröffentlichung eines Regierungsentwurfs die Hauptlinien des Gesetzes determiniert. Mögliche Detailänderungen erfolgen jedoch eher im fortgeschrittenen Stadium im Parlament, sodass dieses Stadium ebenfalls Relevanz genießt.

"*Kein Gesetz verlässt den Bundestag so, wie es hereingekommen ist*" lautet das im politischen Berlin bekannte und nach dem ehemaligen Verteidigungsminister Peter Struck benannte "strucksche Gesetz" (Kurbjuweit und Medick, 2022, S. 29), welches auch in der Steuergesetzgebung seine Gültigkeit behält. Angesichts divergierender Motive unterschiedlicher Gruppen könnten Gesetzesvorhaben dort weiterhin in verschiedene Richtungen gelenkt werden. Die bisherige Literatur zu politischer Einflussnahme signalisiert häufig Unterschiede hinsichtlich der Einflussintensität unterschiedlicher Interessengruppen in verschiedenen Szenarien, für die es differenzierte Gründe gibt. Allerdings existieren bisher keine Erkenntnisse darüber, welche Gruppen ihre Interessen am ehesten in den steuerlichen Gesetzgebungsprozess in Deutschland einfließen lassen können. Die an diese Lücke anknüpfende Forschungsfrage lautet daher: "Welchen Expertengruppen folgen die Parlamentarier des Deutschen Bundestags bei der Implementierung von neuer Steuergesetzgebung?". Unter Zuhilfenahme von schriftlichen Stellungnahmen, die im Rahmen von Expertenanhörungen im Finanzausschuss des Bundestags eingereicht worden sind, soll ihr Einfluss auf mögliche Änderungen im Gesetzgebungsverfahren mithilfe einer Regressionsanalyse gemessen werden, um die Forschungsfrage zu beantworten. Dabei wird der Fokus auf Vorhaben gesetzt, die die internationale Besteuerung betreffen, da dieses Themenfeld insbesondere in den vergangenen Jahren enorm an fiskalpolitischer Bedeutung gewonnen hat, wie das Vorgehen der OECD zur Bekämpfung von Gewinnverkürzung und -verlagerung (BEPS) unter Beweis stellt (vgl. OECD, 2015, S. 4; Mulligan und Oats, 2016, S. 63).

Die Beantwortung der Forschungsfrage ist für alle im demokratischen Prozess beteiligten Akteure relevant: Je nach Ergebnis können für die lobbyierenden Gruppen etwa ihre Erfolgschancen gegenüber anderen Gruppen evaluiert und ggf. für spätere Handlungsänderungen in ihrer Strategie verwendet werden. Für die politischen Entscheidungsträger erwachsen Erkenntnisse über ihr Berücksichtigungsverhalten unterschiedlicher Gruppen und bei einer möglichen Bevorzugung oder Benachteiligung einen Anstoß zur Einschätzung, aus welchen Gründen diese erfolgt. Auch gesamtgesellschaftlich sind die Ergebnisse relevant; geben sie doch Aufschluss über den Lobbyerfolg einzelner Gruppen und damit Transparenz, welche Kollektive gegenüber anderen mehr Einfluss ausüben. Bei der Feststellung von starken Ungleichgewichten könnten sich gar Fragen der demokratischen Legitimation ergeben (van de Vijer, 2022, S. 58).

1.2 Gang der Arbeit

Um die Beantwortung der Forschungsfrage geordnet voranzutreiben, verfolgt die Arbeit den folgenden Gang: Zu Beginn werden in Kapitel 2 mit einem dreiteiligen Literaturüberblick die bisherigen Erkenntnisse zur untersuchten Thematik aufgegriffen. Nachdem zunächst auf den allgemeinen Einfluss von Lobbyismus eingegangen wird, folgt eine Untersuchung der Ergebnisse bei Gesetzgebungsprozessen auch unter Differenzierung von gruppenspezifischen Ergebnissen, bevor danach auf Erkenntnisse der Literatur zum Lobbyismus bei öffentlichen Anhörungen bzw. schriftlichen Stellungnahmen geblickt wird. Kapitel 3 bildet den Übergang von Theorie hin zur eigentlichen Untersuchung, indem durch die Beleuchtung des institutionellen Hintergrundes der Ansatz der Analyse im deutschen Gesetzgebungsprozess nachvollziehbar wird. In Kapitel 4 folgt die Vorstellung des angewendeten Forschungsdesigns, wobei neben der Entwicklung der zugrundeliegenden Hypothese auch die durchzuführenden Kodierungen und der Schätzungsansatz im analytischen Modell berücksichtigt wird. Kapitel 5 beinhaltet die Untersuchungsergebnisse, welche deskriptive Statistiken, qualitative Analysen, Ergebnisse der Regressionsanalysen sowie Robustheitstests und die Benennung relevanter Limitationen umfassen. Das Fazit der Arbeit in Kapitel 6 bildet anschließend den Schluss.

2 Literaturüberblick

Um eine fundierte Auseinandersetzung über die Einflussnahme von Expertengruppen auf die Steuergesetzgebung zu ermöglichen, soll im Folgenden zunächst mithilfe der bisherigen Literatur eine theoretische und evidenzbasierte Grundlage zur Thematik etabliert werden.

Das Konzept des Einflusses wird in verschiedenen Wissenschaften in unterschiedlichem Maße verwendet. Im Zuge dieser Arbeit bietet es sich deswegen an, mit einem allgemeinen Verständnis über "Einfluss" zu arbeiten, der gleichzeitig in politik-, sozial- und wirtschaftswissenschaftlichen Sphären anwendbar ist. Ein solches interdisziplinäres Konzept ist wünschenswert, da die Verbindung aus unterschiedlichen Wissenschaftssträngen zur Verbesserung eines ausführlichen theoretischen Hintergrunds zum Themenkomplex der politischen Einflussnahme beiträgt (vgl. Barrick und Brown, 2019, S. 75; Mellado und Parte, 2021, S. 4). Ein geeignetes und vielzitiertes Konzept hierfür liefert Parsons (1963). Dieser definiert Einfluss als "*einen Weg durch (nicht notwendigerweise rationale) beabsichtigte Handlungen eine Auswirkung auf Ansichten und Meinungen anderer auszuüben – die Auswirkung könnte ein Wechsel der Meinung oder die Verhinderung einer möglichen Änderung sein*" (Parsons, 1963, S. 38).

Inwiefern Expertengruppen solche Auswirkungen im Rahmen ihrer schriftlichen Stellungnahmen bei Anhörungsverfahren zu deutscher Steuergesetzgebung ausüben können, wird im weiteren Verlauf Gegenstand der Untersuchung sein.

2.1 Einfluss von Lobbyismus

Bei der Betrachtung von Einflussmöglichkeiten im politischen und wirtschaftlichen Umfeld erfolgt korrespondierend eine Berücksichtigung des Lobbyismus. Das Konzept des Lobbyismus ist Gegenstand von einer Vielzahl theoretischer und empirischer Untersuchungen. War Lobbyismus ursprünglich nur eine von mehreren möglichen Strategien zur politischen Einflussnahme (vgl. Keim und Zeithaml, 1986, S. 829), fallen vor allem im europäischen Raum sämtliche politischen Beeinflussungsstrategien unter dieses Konzept (vgl. Hillman und Hitt, 1999, S. 834). Im Folgenden soll jedoch auf eine wissenschaftlich gängige und im angelsächsischen Raum verbreitete Definition von Lobbyismus abgestellt werden, da diese den hiesigen Untersuchungsgegenstand präziser beschreiben kann.

In diesem Zuge bedeutet Lobbyismus eine Bereitstellung von Informationen an politische Entscheidungsträger mit der Intention einer Bekanntmachung von eigenen Positionen und der Beeinflussung von politischen Ergebnissen (vgl. Keim und Zeithaml, 1986, S. 830; Hillman und Hitt, 1999, S. 834; Barrick und Alexander, 2014, S. 24; de Figueiredo und Richter, 2014, S. 164). Somit werden Personen, die sich bei Lobbyaktivitäten beteiligen, oftmals als Lobbyisten bezeichnet. Generell fallen unter solche Lobbyaktivitäten Bemühungen, Kommunikationskanäle mit Regulatoren und Gesetzgebungsinstitutionen sowie deren Mitarbeitern zu etablieren und zu verwenden (vgl. Keim und Zeithaml, 1986, S. 830). Über solche Kanäle können auf vielfältige Arten Informationen fließen, z.B. in Form von Teilnahmen an offiziellen Anhörungen, der Bereitstellung von Positionspapieren, Berichten, Statistiken, Argumenten, Prognosen oder eine Kombination dergleichen (vgl. Hillman und Hitt, 1999, S. 834; de Figueiredo und Richter, 2014, S. 164; Barrick und Brown, 2019, S. 73). In repräsentativen Demokratien ist ein solches Ersuchen von gewählten Vertretern zur möglichen Beeinflussung grundsätzlich ein zentrales Recht, welches individuell oder kollektiv ausgeübt werden darf (vgl. de Figueiredo und Richter, 2014, S. 164). In Deutschland dienen die Meinungs-, Presse- und Koalitionsfreiheit des Grundgesetzes als grundrechtliche Legitimation für Interessenvertretung jeglicher Art (vgl. Gerig, 2014, S. 247).

Es ist zwar anzunehmen, dass durch Lobbyisten beschaffte und für Politiker vorher unbekannte Informationen für die Entscheidungsträger potenziell dabei nützlich sein können, fundiertere politische Entscheidungen zu treffen. Dabei versuchen die Interessengruppen die Entscheidungsträger davon zu überzeugen, dass jene vertretenen Positionen auch ihren Interessen und die der Gesellschaft entsprechen (vgl. Grossman und Helpman, 2001, S. 96).

Doch Grossman und Helpman (2001) weisen auch darauf hin, dass sofern die Rezipienten keine übereinstimmenden Ziele mit den Lobbyisten aufweisen, die erhaltenen Informationen aufgrund möglicher Verzerrungen durch Missinterpretationen oder Übertreibungen (= Bias) mit Vorsicht betrachtet werden müssen (vgl. S. 97). Politiker sind sich solchen Verzerrungen bewusst, weswegen Lobbyaktivitäten nur eine überzeugende Wirkung entfalten können, wenn der berichtende Lobbyist auch einen Anreiz für die Tätigung wahrheitsgemäßer Angaben hat. Ein solcher Anreiz und damit der größte Nutzen für die Empfänger resultiert aus einer Situation, in der verschiedene Interessengruppen die Entscheidungsträger in jeweils unterschiedliche Richtungen bewegen möchten; hierdurch fungiert jede Gruppe als Kontrollinstanz für die Kredibilität der Angaben konkurrierender Lobbyisten (vgl. Grossman und Helpman, 2001, S. 101, 129). Eine solche Situation stellt die Expertenanhörung in Ausschüssen des Deutschen Bundestags inklusive der Möglichkeit zur Einreichung von Stellungnahmen zweifelsfrei dar.

Interessanterweise hatten frühe Grundprämissen über Theorien zu Lobbyismus die Einigkeit, dass die angewandten Methoden das ultimative Ziel einer Präferenzänderung bei den adressierten Entscheidungsträgern verfolgen (vgl. Grossman und Helpman, 2001, S. 96; Hall und Deardorff, 2006, S. 69). Doch mittlerweile findet eine andere Perspektive in der Literatur eine ähnliche Beachtung: Mit der Formulierung der Lobbytheorie einer “legislativen Subventionierung” vollziehen Hall und Deardorff (2006) einerseits einen Wechsel von einer Orientierung nach Präferenzen hin zu einer Fokussierung auf (Zeit-)Budgets (vgl. ebd., S. 69) und andererseits eine Änderung der Adressaten von zu überzeugenden Entscheidungsträgern hin zu Politikern mit bereits übereinstimmenden Ansichten (vgl. ebd. S. 72, ähnlich Smith, 1995, S. 97). So können sich die zeitlich beschränkten Abgeordneten mit der Hilfe von Lobbyisten, die diesen mit spezialisierten Informationen in Form von individuellen Entwürfen, Argumenten, Strategieanweisungen oder Folgeabschätzungen versorgen und damit signifikant zuarbeiten (vgl. Hall und Deardorff, 2006, S. 74). Hauptprämisse der Strategie ist somit die Unterstützung verbündeter Entscheidungsträger zur Erreichung gemeinsamer Ziele. Schnakenberg (2017) entwickelt diese Annahme in seinem Modell weiter: Bei einer gleichzeitigen Betrachtung verschiedener Entscheidungsträger schlussfolgert er, dass Interessengruppen ihre Lobbyaktivitäten auf verbündete Abgeordnete richten und diese als Intermediäre agieren, damit sie die erhaltenen Informationen für die Überzeugung von Abgeordneten mit anderweitigen Präferenzen nutzen können (vgl. Schnakenberg, 2017, S. 130, 137). Ein deutlicher Bruch zur Prämisse der allgemeinen Überzeugungsarbeit besteht somit nicht. Auch die eigenen Erfahrungen von Lobbyisten aus der Praxis zeigen, dass Maßnahmen zur Präferenzänderung eine wichtige Komponente für ihren Lobbyerfolg darstellen (vgl. Schnakenberg, 2017, S. 131). Die Theorien sind somit nicht klar voneinander abgrenzbar; dienen aber als mögliche Erklärungsansätze für die Wirkungsweise angewandter Lobbystrategien und könnten damit bei der Untersuchung zum Einfluss der Stellungnahmen auf Steuergesetze relevant sein.

Relevanz in der Lobbyliteratur genießt vor allem Sutton (1984), der in seiner Arbeit für die Motivation hinter solchen Praktiken ein ökonomisches Kalkül betrachtet: Nur wenn die aus den Lobbystrategien zu erwartenden Vorteile die hierdurch entstehenden Kosten überwiegen, wird sich

eine Interessengruppe an Lobbyaktivitäten beteiligen (vgl. Sutton, 1984, S. 82f). Eine ähnliche Argumentation verfolgt Olson (1965) bei seinen Überlegungen, in welchen Fällen sich gleichgesinnte Akteure zu einer Gruppe organisieren, um ihre gemeinsamen Ziele zu erreichen (vgl. Olson, 1965, S. 31). Eine solche numerische Betrachtung könnte insbesondere die Motivation für Unternehmen erklären, sich in politischer Einflussnahme und dabei auch in Organisationen zu engagieren. Allerdings ist diese Erklärung für die übrigen Gruppen unzureichend, da diese häufig auch non-monetäre oder sogar idealistische Motive verfolgen. Für alle Interessengruppen ist jedoch gemein, dass sie in spezialisierten Gruppen auftreten sollten, da diese gegenüber Einzelakteuren besser organisiert sind, sowie mit anderweitigen Anreizen eine attraktive Anlaufstelle für ähnliche Akteure bleiben, um am effektivsten die gemeinsamen Ziele zu erreichen (vgl. Olson, 1965, S. 133, 145; Mellado und Parte, 2021, S. 10). In Deutschland erfolgt eine solche Organisation meistens über Verbände, Kammern oder Gewerkschaften, die u.a. auch die Interessen ihrer Mitglieder gegenüber politischen Entscheidungsträgern vertreten und regelmäßig an Expertenanhörungen teilnehmen.

2.2 Einfluss auf Gesetzgebungsprozesse

Nachdem grundlegende Aspekte zum Lobbyismus betrachtet wurden, bietet sich nun die Untersuchung von Erkenntnissen an, die konkret Einflüsse auf bestimmte Gesetzgebungs- oder Standardsetzungsprozesse betreffen.

2.2.1 Allgemein

In der Literatur wird angemahnt, dass Lobbyaktivitäten relativ früh im Legislativprozess angestoßen werden sollten. Da sich schon vor Ausformulierung eines Gesetzentwurfs die Ansichten der Entscheidungsträger kontinuierlich ausprägen, sind Lobbyansätze vor einer Entwurfsveröffentlichung effektiver; auch wenn danach noch erfolgreiches Lobbyieren möglich ist (vgl. Sutton, 1984, S. 88). Dazu passt, dass Regierungen bzw. die federführenden Ministerien den ersten und wichtigsten Anlaufpunkt für Interessenvertreter darstellen, obwohl auch Parlamente weiterhin eine wichtige Rolle spielen (vgl. Eising und Spohr, 2017, S. 314). Dennoch werden z.B. von Unternehmen die späteren Etappen des Prozesses zur Festlegung neuer Rechnungslegungsstandards stärker für Lobbyaktivitäten verwendet; insbesondere nach der Veröffentlichung entsprechender Entwürfe (vgl. Georgiou, 2004, S. 225f). Obwohl manche der Unternehmen die obige Logik nachvollziehen können, erachtet ihre Mehrheit ihr Vorgehen paradoxerweise dennoch als effektiv (vgl. ebd., S. 228). Auch in der Forschung werden Lobbyisten signifikante Einflussmöglichkeiten auf den Legislativprozess attestiert, wobei sich das informative Lobbyieren früh als die meistgenutzte Strategie etablierte (vgl. Aplin und Hegarty, 1980, S. 446; Smith, 1995, S. 97).

Darüber hinaus wurde in Einklang mit Olson (1965) empirisch bestätigt, dass spezialisierte Gruppen gegenüber allgemeinorientierten Organisationen in Gesetzgebungsprozessen wesentlich mehr Einfluss ausüben können (vgl. Schneider und Baltz, 2003, S. 273). Eine wahrscheinliche Erklärung dafür ist, dass sie durch ihre speziellen Kenntnisse reichhaltigere und detailliertere Argumente liefern können als ihre allgemeingehaltenen Kollegen. Für Entscheider sind in ihrer täglichen Arbeit erstere weitaus nützlicher, um die behandelte Thematik tiefgehender durchdringen zu können (vgl. Mellado und Parte, 2021, S. 16).

Eine objektive Beurteilung der Effektivität gestaltet sich jedoch schwierig. Mögliche Ergebnisse von Lobbyismus sind nicht immer kodifiziert, wodurch sie in der Gesetzgebung nicht wahrnehmbar sein könnten (vgl. Richter et al., 2009, S. 895). Des Weiteren fußen die meisten Resultate über politische Lobbyergebnisse nur auf erzählbasierten Fallstudien oder Umfragen (vgl. Smith, 1995, S. 117), sodass solche Auswirkungen lediglich fallbasiert evaluiert werden können. Dennoch sehen viele Organisationen eine Teilnahme am politischen Prozess als sehr wichtig für ihre (subjektiv erfolgreiche) Gesamtlobbystrategie an (vgl. Furlong und Kerwin, 2005, S. 356). McLeay et al. (2000) argumentieren, dass die bei Steuergesetzen üblichen präzisen und hochspezifischen Angaben der Lobbyisten jedoch eine objektivere Identifizierung ihrer Präferenzen sowie ein verlässlicheres Maß an Einfluss ermöglichen (vgl. ebd., S. 80f). Darüber hinaus identifiziert Mahoney (2007) einige Faktoren, die den Erfolg von Lobbyaktivitäten beeinflussen: So nimmt die Wahrscheinlichkeit für eine erfolgreiche Einflussnahme ceteris paribus ab, je breiter die Thematik in der Öffentlichkeit diskutiert wurde oder je umfangreicher dieses in seiner Gänze ist (vgl. Mahoney, 2007, S. 47–50). Eher erfolgreich sind Lobbyversuche, die den Status quo möglichst erhalten wollen (vgl. Mahoney, 2007, S. 52). Neben auf Expertise gestützter Strategien scheinen auch persönliche Verbindungen zu den Entscheidungsträgern relevante Einflussfaktoren zu sein (vgl. Aplin und Hegarty, 1980, S. 447; Bertrand et al., 2014, S. 3917; de Figueiredo und Richter, 2014, S. 167). Zusätzlich verbessert bereits bestehende Lobbyerfahrung sowie die Verwendung rhetorischer Stilmittel die Fähigkeit für politische Einflussnahme (vgl. Georgiou, 2004, S. 233; Hoffmann und Zülch, 2014, S. 721).

2.2.2 Differenzierung nach Expertengruppen

Für die zu untersuchenden Expertengruppen selbst bestehen in der Literatur im Umfang sowie auch inhaltlich unterschiedliche Erkenntnisse zur politischen Einflussnahme.

Die mit Abstand meisten Untersuchungen gibt es zu **Unternehmen** bzw. Akteuren mit wirtschaftlichen Interessen. Diese sind die Interessengruppe mit den höchsten Ausgaben für Lobbyismus (vgl. de Figueiredo und Richter, 2014, S. 165). Es ist davon auszugehen, dass sie für (gegen) Änderungen lobbyieren, die neue Vorteile (Kosten) für ihre Unternehmen bedeuten (vgl. Watts und Zimmerman, 1978, S. 113). Dabei lobbyieren Unternehmen ohne gewisse Größe und/oder Expertise eher kollektiv in Verbänden, um bei der gemeinsamen Verfolgung derselben Ziele in den Genuss von Skalierungseffekten zu kommen (vgl. Hillman und Hitt, 1999, S. 831). Durch die Bildung von Koalitionen kann

die Lobbywirkung sogar verstärkt werden, da durch große übereinstimmende Gruppenpositionen eine breit angelegte Unterstützung der vertretenen Meinung signalisiert wird (vgl. Keim und Zeithaml, 1986, S. 839). Hierbei sind die Unternehmen aufgrund ihrer homogeneren Interessen gegenüber anderen Interessengruppen eher im Vorteil (vgl. Sutton, 1984, S. 86; Eising und Spohr, 2017, S. 316). Ein weiterer Vorteil ist ihre größere Kapazität zur Erfassung von komplexen zugrundeliegenden Studien und Daten, zur Gewinnung eigener Daten sowie ihrem Zugang zu erfahrenen Anwälten, Lobbyisten oder Experten (vgl. Yackee und Yackee, 2006, S. 131). Dieser Umstand könnte erklären, warum diese Interessengruppe im Vergleich zu anderen tendenziell eine höhere Qualität bei der Bereitstellung ihrer Informationen und eine stärkere Wahrnehmung von Expertise genießt (vgl. ebd., S. 130). Solche Charakteristika könnten Erklärungsansätze dafür sein, dass Unternehmen in empirischen Studien durchweg als Interessengruppe mit den meisten Einflussmöglichkeiten hervorgehoben werden: Im Vergleich zwischen Produzenten und Konsumenten stellen Schneider und Baltz (2003) fest, dass die Interessen der politisch wesentlich aktiveren Produzenten gegenüber Konsumenten im europäischen Kontext vorteilig behandelt werden (vgl. ebd., S. 15, 19). Yackee und Yackee (2006) stellen in einem amerikanischen Umfeld fest, dass Unternehmen bei regelungsbasierten Vorhaben einen direkten Einfluss ausüben können, der von anderen Gruppen nicht im selben Maße erreicht wird (vgl. ebd., S. 133, 135). In Untersuchungen im deutschen Kontext, die dieses Ungleichgewicht bestätigen, soll dabei die Signalisierung einer möglichst starken Opposition gegenüber einem Gesetzesvorschlag die erfolgreichste Strategie für Unternehmen sein, um möglichst viele Änderungen zu ihren Gunsten erreichen zu können (vgl. McLeay et al., 2000, S. 92; Eising und Spohr, 2017, S. 328).

Dem gegenüber verfahren andere Interessengruppen mit einer möglichst neutralen Expertise zum Themenkomplex als bestmögliche Strategie zur Politikbeeinflussung (vgl. Eising und Spohr, 2017, S. 328). Solche strategischen Überlegungen könnten die wahren Präferenzen der Akteure allerdings verzerren, was bei der Untersuchung berücksichtigt werden muss. In diesen relativen Vergleichen wird deutlich, dass Angehörige der **Zivilgesellschaft** weniger Einfluss ausüben können. Ein solches Ungleichgewicht findet auch Mahoney (2007), die diese Diskrepanz in den Vereinigten Staaten stärker ausgeprägt sieht als in der Europäischen Union (EU) (vgl. ebd., S. 51). Bei Gesetzesvorhaben in Deutschland haben zivile Akteure nur bei der Äußerung von sehr bedeutsamen Einwänden eine signifikante Chance, Änderungen bei Gesetzesvorhaben auszulösen (vgl. Cross et al., 2021, S. 370).

Expertise scheint ein relevanter Faktor bei der Ausübung von Einfluss zu sein, da bei empirischen Untersuchungen Fachexperten bei ihren behandelten Themengebieten mehr Einfluss ausüben können als andere Akteure (vgl. McLeay et al., 2000, S. 94; Perry und Nöelke, 2005, S. 17). Da beide Gruppen zur Bewerbung ihrer präferierten Politik einen Anreiz für eine verzerrte Präsentation von privaten Informationen haben, kann mithilfe von Experten mit wahrgenommener unabhängiger Autorität solchen Verzerrungen entgegengewirkt werden (vgl. Tandy und Wilburn, 1996, S. 93; Cross et al., 2021, S. 358). Eine solche Gruppe können etwa **Wissenschaftler** darstellen. Diese können den Entscheidungsträgern etwa bei der Einordnung fachspezifischer Daten oder Forschung beratend zur Seite stehen (vgl. Tandy und Wilburn, 1996, S. 93). Allerdings ist die Teilnahmeintensität von Akademikern an solchen Gesetzgebungsverfahren sehr gering (vgl. Tandy und Wilburn, 1996, S. 101; Yen et al., 2007, S. 59; Mellado und Parte, 2021, S. 17), weshalb Parlamentarier nur selten in den Genuss einer solchen Unterstützung kommen. Geringe Erwartungen an einen möglichen Einfluss auf die finale Entscheidung, Zeit- und Ressourcenmangel, sowie eine unbefriedigende akademische Belohnung werden als Gründe für die geringe Partizipation seitens der Wissenschaftler genannt (vgl. Tandy und Wilburn, 1996, S. 105–107). McLeay et al. (2000) bestätigen in ihrer Untersuchung zur Implementierung neuer Rechnungslegungsstandards im deutschen Handelsrecht den relativ geringen Einfluss dieser Gruppe; dabei stellen sie jedoch auch fest, dass durch Interaktionen mit Unternehmen ihre Einflussmöglichkeiten bedeutsam zunehmen (vgl. ebd., S. 94f). Trotz der geringen Teilnahmeanreize wird ihr Expertenwissen zunehmend wichtiger, je komplexer die angestoßenen Gesetzgebungsverfahren sind; Cross et al. (2021) stellen in dieser Hinsicht sogar eine Abhängigkeit der Parlamentarier fest (vgl. ebd., S. 371).

Eine andere zu dieser Thematik in der Literatur weitgehend unbeachtete Expertengruppe sind die **Steuerberater**. In ihrer speziellen Position existiert jedoch eine Ambivalenz zwischen unabhängigen Experten und Akteuren mit wirtschaftlichen Interessen: Durch ihre ausgeprägten Kenntnisse über steuerliche Themen könnten sie mit ihren hilfreichen Praxisinformationen die Entscheider in für ihre Mandanten oder sogar für ihre eigenen geschäftlichen Interessen bevorzugte Richtungen lenken (vgl. Elschner und Hardeck, 2022, S. 312), beispielsweise für geringere Steuersätze im ersteren Fall oder für wachsende Dokumentationspflichten mit erhöhtem Beratungsbedarf im letzteren Fall. Zumindest unter Zuhilfenahme der Ergebnisse von Untersuchungen zum Verhältnis von Wirtschaftsprüfern zu ihren Mandanten lässt sich jedoch trotz Abhängigkeitsverhältnisses (vgl. Puro, 1984, S. 636) keine kundenbezogene Positionierung in Lobbyfragen feststellen (vgl. Puro, 1984, S. 640f; Allen et al., 2018, S. 314), sondern vielmehr eine konzeptionelle Unterstützung der Standardsetzer (vgl. Allen et al., 2018, S. 325). Auch die mögliche Motivation von Wirtschaftsprüfern zur Teilnahme für eine potenzielle Reputationssteigerung (vgl. Bamber und McMeeking, 2016, S. 62) könnte für die Partizipation von Steuerberatern eine ähnliche Erklärung sein. Da diese tendenziell auch moralische Hintergründe berücksichtigen (vgl. OECD, 2008, S. 14; Radcliffe et al., 2018, S. 51) und durch ihre verantwortungsvollen Tätigkeiten zur Unterstützung bei der Tax Compliance von Unternehmen einen wichtigen Beitrag für die Gesellschaft leisten (vgl. OECD, 2008, S. 14; Elschner und Hardeck, 2022, S. 312), ist eher

mit einer neutraleren Positionierung der Steuerberater zu rechnen.

2.2.3 Steuergesetze

Lobbyismus bei Themen mit steuerlicher Relevanz hat eine große gesellschaftliche Bedeutung, sowohl für die Steuerzahler auf der einen und der Gesellschaft als Ganzes auf der anderen Seite. Dabei sind die verfügbaren Informationen, wie Steuerlobbyismus genau ausgeführt wird, relativ spärlich, obwohl dieses Feld sogar als "heiliger Gral" des Lobbyismus angesehen wird (vgl. Barrick und Alexander, 2014, S. 43f). Erfolgsversprechend für die Unternehmen ist es allemal: So haben Firmen mit politischen Verbindungen einen durchschnittlich niedrigeren effektiven Steuersatz als Firmen ohne solche Berührungspunkte (vgl. Faccio, 2010, S. 913). Kim und Zhang (2016) stellen fest, dass Unternehmen mit politischen Verbindungen im Vergleich zu unpolitischen Firmen tendenziell eine aggressivere Steuerplanung betreiben (vgl. ebd., S. 96). Ähnliche Vorteile sehen Richter et al. (2009), die bei der Erhöhung der Lobbyausgaben von börsennotierten U.S.-Unternehmen eine Reduzierung ihrer effektiven Steuersätze im darauffolgenden Jahr ermitteln (vgl. ebd., S. 898, 901). In einer der wenigen Studien, die die Auswirkungen von Lobbyismus auf steuerliche Gesetzgebungsprozesse erörtern (laut Elschner und Hardeck, 2022, S. 305), stellen Mulligan und Oats (2016) einen signifikanten Einfluss durch Steuerfachleute großer amerikanischer multinationaler Unternehmen (MNEs) auf Steuergesetze und -regelungen fest, die für ihre betroffenen Unternehmen eine vorteilige Auswirkung erzielen sollen (vgl. ebd., S. 71-73). Auch wegen der möglichen Gefahr einer Legitimitätsunterminierung durch einen größeren Einfluss von Unternehmen werden Rufe nach mehr Transparenz bei Steuerlobbyismus laut (vgl. van de Vijer, 2022, S. 58).

In der nach eigenen Worten ersten empirischen Untersuchung zu dieser Thematik stellen Elschner und Hardeck (2022) jedoch gegensätzlich fest, dass Unternehmen im Vergleich zu den übrigen Gruppen zumindest bei der Signalisierung von Präferenzen am wenigsten Einfluss ausüben, diesen aber mit der Bereitstellung nützlicher Informationen im Gegensatz zur Zivilgesellschaft signifikant verbessern können (vgl. Elschner und Hardeck, 2022, S. 321). In beiden Fällen können jedoch unabhängige Steuerexperten aufgrund eines geringeren anzunehmenden Bias mehr Einfluss ausüben als der Rest (vgl. ebd.). Generalisierende Aussagen sind durch diese gemischten Ergebnisse demnach nicht möglich. Somit wächst die Relevanz der hier intendierten Untersuchung im deutschen Kontext.

2.3 Einfluss durch Anhörungen / Stellungnahmen

Das Konstrukt der Anhörungen bzw. der damit verbundenen Möglichkeit zur Einreichung schriftlicher Stellungnahmen ist in demokratischen Systemen weitverbreitet und oft Gegenstand wissenschaftlicher Untersuchungen. Die offensichtlichste Funktion solcher Anhörungen ist es, den Entscheidungsträgern Informationen von verschiedenen Gruppen zu liefern (vgl. Hillman und Hitt, 1999, S. 834; Brasher, 2006, S. 585). Dabei sind diese je nach politischem Kontext unterschiedlich bedeutsam: Während etwa Anhörungen zur Bürgerbeteiligung in der Lokalpolitik mehrheitlich als ineffektive Formalie ohne nennenswerten Einfluss angesehen werden (vgl. Checkoway, 1981, S. 578; Lando, 2003, S. 81), genießen Anhörungen und Gelegenheiten zur schriftlichen Stellungnahme bei Prozessen zur Entwicklung von Rechnungslegungsstandards eine große Aufmerksamkeit in der Literatur über Einflussausübung (vgl. Baudot und Walton, 2014, S. 320). Sie gelten als die bekannteste und "archetypische" Methode zur Partizipation an politischen Prozessen (vgl. Furlong und Kerwin, 2005, S. 362). Als erstes begannen in diesem Kontext Watts und Zimmerman (1978) damit, die Präferenzen der Verfasser von schriftlichen Stellungnahmen (sog. "Comment Letters") aus ihren Beiträgen zu extrahieren (vgl. Kwok und Sharp, 2005, S. 81). Solche erlauben einen Einblick auf die vertretenen Positionen und die Verteilung von Lobbyaktivitäten verschiedener Akteure auf veröffentlichte Entwürfe, wobei daraus lediglich Erkenntnisse über die Präferenzen und die Lenkungsrichtung abgeleitet werden sollten und nicht über die schwierig einzuschätzende Intensität der Lenkungsversuche (vgl. Puro, 1984, S. 632; Sutton, 1984, S. 86; Kwok und Sharp, 2005, S. 80; Bisogno et al., 2015, S. 312). Die Analyse schriftlicher Stellungnahmen ist ein nützliches Instrument zur Messung von Lobbystrategien, da sie öffentlich verfügbar sind und auch unsichtbare Aktivitäten der Verfasser reflektieren (vgl. Mellado und Parte, 2021, S. 17), was sie zum meistverwendeten Proxy für Lobbyaktivitäten in der Rechnungslegungsliteratur macht (vgl. ebd., S. 19). Bei der inhaltlichen Untersuchung schriftlicher Stellungnahmen existieren zwei generelle Ansätze: Während bei einer quantitativen Untersuchung numerische Aspekte berücksichtigt werden, konzentrieren sich qualitative Untersuchungen auf die Bedeutung hinter dem geschriebenen Wort, um so reichhaltigere Einsichten zu gewinnen (vgl. Yen et al., 2007, S. 61). Da dahinter jedoch individuelle Interpretationsentscheidungen stehen, sind diese Methoden nicht so objektiv wie quantitative Ansätze.

Unabhängig der Methodenwahl konnten jedoch in einigen Beiträgen diverse Feststellungen getroffen werden, auch wenn im Vergleich zu Untersuchungsansätzen bei Regierungen empirische Studien zu parlamentarischer Interessenvermittlung einen kleineren Teil ausmachen (vgl. Eising und Spohr, 2017, S. 315). Jorissen et al. (2012) stellen eine steigende Anzahl von Kommentareinreichungen fest, sobald bei veröffentlichten Regelungsentwürfen ein großer Einfluss auf Bilanzwerte oder Veröffentlichungspflichten von Unternehmen zu erwarten ist (vgl. ebd., S. 712–714). Dies deckt sich mit der in der Literatur anerkannten Erkenntnis einer größeren Teilnahme von Stakeholdern, sofern eine größere wirtschaftliche Betroffenheit zu erwarten ist (vgl. Mellado und Parte, 2021, S. 7f). Viele Organisationen sehen eine Teilnahme am Entscheidungsprozess als sehr wichtig für ihre Gesamtlobbystrategie an (vgl. Furlong und Kerwin, 2005,

S. 356). Dabei muss es nicht zwangsläufig um eine direkte Beeinflussung gehen. Stellungnahmen können auch für defensives Lobbyieren verwendet werden, um die Ansichten anderer Gruppen zu relativieren und den Status quo möglichst zu erhalten (vgl. Furlong und Kerwin, 2005, S. 365). Der Grad an Einflussnahme unterscheidet sich dabei nach den Interessengruppen: So kann ein signifikanter Einfluss durch große Wirtschaftsprüfungsgesellschaften ermittelt werden (vgl. Allen et al., 2018, S. 325). Bamber und McMeeking (2016) stellen jedoch einen im Vergleich zu anderen Wirtschaftsprüfern geringeren Einfluss der sogenannten Big Four[2] in Entwicklungsprozessen zu Rechnungslegungsstandards fest (vgl. ebd., S. 69). Somit könnte auch in dieser Untersuchung die Zugehörigkeit geladener Steuerberater zu Kanzleien einer bestimmten Größe ausschlaggebend auf Einflussergebnisse sein.

Inwiefern im Deutschen Bundestag die Expertengruppen Gesetzgebungsprozesse durch Anhörungen beeinflussen können, wurde bisher nur selten untersucht. In der Analyse aller eingereichten Stellungnahmen des Jahres 2011 stellen Eising und Spohr (2017) fest, dass Unternehmen mit einer starken Opposition gegen die Entwurfsposition am meisten Einfluss ausüben können, wohingegen bei Angehörigen anderer Gruppen eine solche Strategie keine Auswirkungen zeigt; letztere sollten vielmehr mit neutraler Expertise dienen (vgl. ebd., S. 328). Dieses Ungleichgewicht wird mit den unternehmerischen Sanktionsmöglichkeiten gegenüber dem (nationalen) Gesetzgeber begründet (vgl. ebd.), da die Unternehmen mit potenziellen Investitionsreallokationen ungewollte makroökonomische Auswirkungen provozieren könnten. Die Autoren konstatieren darüber hinaus aber eine größere Bedeutung aller Expertengruppen, die über mögliche Bestätigungen bereits bekannter Informationen hinausgeht (vgl. ebd.). Dagegen kommen in einer quantitativen Analyse, bei der Einfluss durch die Intensität an Textänderungen gemessen wird, Cross et al. (2021) zu dem Schluss, dass nur signifikante große Einwände der Zivilgesellschaft zu Textergänzungen führen (vgl. ebd., S. 369f). Dagegen werden bei einer mehrheitlichen Opposition von wirtschaftlichen Akteuren nur durch ihre Anzahl bereits Textergänzungen getätigt (vgl. ebd.). Allerdings bleibt fraglich, was die jeweiligen Textänderungen qualitativ bedeuten, weshalb keine generalisierende Aussage zur Effektivität solcher Einflussausübung möglich ist.

Bei der Untersuchung des Einflusses von Interessengruppen auf die Ausgestaltung von Besteuerungsregeln mithilfe schriftlicher Stellungnahmen existiert bis dato nur die Ausarbeitung von Elschner und Hardeck (2022). Im internationalen Kontext des BEPS-Projekts erkennen die Autoren, im Gegensatz zu den zuvor genannten Untersuchungen, bei der Signalisierung von Präferenzen einen im Vergleich zu Steuerexperten und Zivilgesellschaft geringeren Einfluss von Unternehmen (vgl. ebd., S. 322). Erst bei der Bereitstellung nützlicher Informationen werden geringere Erfolgschancen der Zivilgesellschaft festgestellt, wohingegen Unternehmen und Steuerexperten mit ihrer Expertise mehr Einfluss ausüben können (vgl. ebd., S. 328). Rücksicht auf persönlich bedingte Verzerrungen sowie unabhängige Expertise scheinen Treiber einer erfolgreichen Einflussausübung zu sein.

Im weiteren Verlauf wird nun die Arbeit von Elschner und Hardeck (2022) auf einen deutschen Kontext adaptiert, um Erkenntnisse über die Beeinflussung deutscher Steuergesetzgebung durch unterschiedliche Expertengruppen zu gewinnen.

3 Institutioneller Hintergrund

3.1 Ansatz im deutschen Gesetzgebungsprozess

Um die Adaption angemessen nachvollziehen zu können, ist ein genaueres Verständnis über das klassische Gesetzgebungsverfahren in der Bundesrepublik nötig. Neben der Bundesregierung haben der Bundesrat und der Bundestag das Recht, eine Gesetzesinitiative einzubringen. Da in den meisten Fällen die Bundesregierung Gesetzgebungsverfahren initiiert und auch die betrachteten Gesetze allesamt durch diese eingebracht worden sind, wird sich im Folgenden auf diesen Fall beschränkt.

Die Erarbeitung eines ersten Gesetzentwurfs erfolgt im für Besteuerungsangelegenheiten zuständigen Fachreferat des Bundesministeriums der Finanzen (BMF).[3] Dieser Referentenentwurf zirkuliert anschließend zur inhaltlichen und rechtlichen Prüfung in anderen wichtigen Ministerien und wird dabei mehrmals überarbeitet. Erst beim finalen Beschluss des Gesetzes im Bundeskabinett wird der Entwurf zum für die Untersuchung relevanten Regierungsentwurf. Nachdem der Bundesrat Gelegenheit zur Stellungnahme zu dem Regierungsentwurf hatte, wird dieser an den Bundestag weitergeleitet. Nach seiner 1. Lesung im Gesamtplenum wird der Entwurf u.a. an den fachlich zuständigen Finanzausschuss überwiesen, der in Steuersachen als federführender Ausschuss gilt. In diesem Stadium wird in einer Detailberatung der Ausschussmitglieder der Entwurf genau geprüft und weiterbearbeitet. Dabei kann von den Mitgliedern auch eine öffentliche Anhörung durchgeführt werden (dazu sogleich in 3.2). Am Ende der Detailberatung entsteht ein Bericht inklusive einer Beschlussempfehlung mit dem exakten Wortlaut des Gesetzes sowie einer Gegenüberstellung mit jeglichen Änderungen gegenüber dem zugrundeliegenden Regierungsentwurf. Die Beschlussempfehlung des Finanzausschusses ist anschließend Gegenstand der Schlussabstimmung nach der 2./3. Lesung im Bundestag. Da die Empfehlung regelmäßig die Annahme des Gesetzentwurfs in Ausschussfassung vorsieht (vgl. Siefken, 2021, S. 128) und ohne Intervention des Bundesrats keine weiteren Änderungen am Entwurf mehr möglich sind, ist der

[2] Als Big Four werden die vier größten Wirtschaftsprüfungsgesellschaften der Welt zusammengefasst.

[3] Schon in dieser Phase werden Gespräche mit verschiedenen Stakeholdern geführt, die bereits dort mit ihren Anmerkungen Einfluss geltend machen können. Da es in dieser Untersuchung jedoch nur um den Einfluss auf das parlamentarische Verfahren geht, bleibt diese Perspektive im weiteren Verlauf unbeachtet.

Gesetzentwurf in Ausschussfassung in der Regel auch das endgültige Gesetz. Somit ist die Detailberatung im Ausschuss die letztmalige Möglichkeit für materielle Änderungen am Gesetz und damit für die Expertengruppen auch die letztmalige Gelegenheit zur effektiven Geltendmachung ihrer Expertise. Auch wenn das Gesetz im Anschluss noch die verwaltungstechnischen Stationen des Gesetzgebungsprozesses[4] durchläuft, können bei der Unterschiedsbetrachtung zwischen Regierungs- und Ausschussentwurf somit die letztmaligen (Nicht-)Änderungen am endgültigen Gesetz gänzlich nachvollzogen werden. Diese Tatsache wird für die spätere Untersuchung maßgeblich sein.

3.2 Expertenanhörungen in Bundestagsausschüssen

Im Zuge der Detailberatung im Ausschuss kann zur Heranziehung von Expertenwissen, das bei den Ausschussmitgliedern u.U. nicht bzw. nicht in ausreichendem Maße vorhanden ist, nach § 70 Abs. 1 GO-BT zum betroffenen Gesetzesvorhaben eine öffentliche Anhörung anberaumt werden, zu der Interessenvertreter, Sachverständige und andere Auskunftspersonen geladen werden können. Diese Norm ist die bedeutsamste Vorschrift für institutionalisierte Einflussnahme durch außenstehende Interessenvertreter (vgl. Gerig, 2014, S. 249). Sobald mindestens ein Viertel der Ausschussmitglieder für die Durchführung einer solchen Anhörungssitzung stimmt, sind ihre benannten Sachverständigen zu hören (§ 70 Abs. 2 S. 1 GO-BT). Erst seit dem Jahr 2023 ist dabei in der Tagesordnung der Veranstaltung bei den eingeladenen Akteuren zusätzlich bekanntzumachen, durch welche Fraktionen diese zur Sitzung eingeladen worden sind und welche möglichen finanziellen Interessensverknüpfungen zum behandelten Gesetzesvorhaben bestehen (§ 70 Abs. 4 S. 3 + Abs. 6 S. 3 GO-BT). Im Vorfeld der Anhörung gibt es die freiwillige Möglichkeit für die Auskunftspersonen jeweils eine schriftliche Stellungnahme einzureichen (§ 70 Abs. 6 S. 2 GO-BT), welche auch veröffentlicht wird.

Sinn und Zweck solcher Anhörungssitzungen werden im Schrifttum unterschiedlich eingeschätzt. Bei der Beurteilung der Einflussintensität von Auskunftspersonen auf die Entscheidungsfindung wird diesen eher eine begrenzte Wirkung zugeschrieben (vgl. Blum, 2004, S. 105; Knauff, 2016, S. 196 Fn.10); auch wegen der begrenzten Möglichkeiten der Parlamentarier. Doch obwohl die Bundesregierung die Schlüsselposition im Gesetzgebungsprozess innehat und die Leitlinien bestimmt, hat sie nicht die volle Kontrolle über den gesamten Prozess (vgl. Bräuninger et al., 2017, S. 530). Im internationalen Vergleich sind sogenannte Arbeitsparlamente wie der Deutsche Bundestag machtvolle Institutionen, die in ihren Ausschüssen zumindest auf der Detailebene größere Möglichkeiten für Gesetzesänderungen haben (vgl. Eising und Spohr, 2017, S. 318; Cross et al., 2021, S. 359). Die eingeladenen Akteure berücksichtigen diese Umstände und ihre damit verbliebenen Spielräume in ihren Ausführungen (vgl. Blum, 2004, S. 105). In einem Extremfall wurde durch die Anhörung im Finanzausschuss das zugrundeliegende Gesetzesvorhaben sogar gestoppt (vgl. Siefken, 2021, S. 123).

Anhörungssitzungen sind dabei für die Abgeordneten ein wichtiges Mittel, um für die Bildung ihrer eigenen Ansicht neben der Ministeriumsperspektive auch etwas über die fachlichen und sozialen Hintergründe des Gesetzentwurfs und die dabei tangierten Interessen zu erfahren (vgl. Blum, 2004, S. 105; Eising und Spohr, 2017, S. 316). Das so eingebrachte "Rationalisierungsmoment" durch die externen Akteure bietet einen Kontrast zu den rein politischen Motiven des bisherigen Legislativprozesses, indem diese die Abgeordneten mit relevanten Informationen versorgen und auf aus ihrer Sicht notwendige Änderungen beim Gesetzentwurf hinweisen (vgl. Blum, 2004, S. 107; Gerig, 2014, S. 249). Allerdings muss einschränkend konstatiert werden, dass jede vertretene Fraktion dazu angereizt ist, Auskunftspersonen einzuladen, die mit ihnen übereinstimmende Positionen teilen (vgl. Kropp und Wagner, 2010, S. 828; Siefken und Schüttemeyer, 2013, S. 169–171). Dies kann zwar ihre jeweiligen Ansichten mithilfe einer vermeintlich unabhängigen Expertenmeinung validieren und damit mehr Gewicht verleihen, konterkariert damit jedoch teilweise den eigentlichen Zweck einer solchen Anhörungssitzung. Dennoch trägt sie dazu bei, die Transparenz und Legitimität politischer Entscheidungen zu stärken, weil Experten, Interessenverbände und zum Teil die Abgeordneten selbst ihre Positionen öffentlich präsentieren und rechtfertigen (vgl. Eising und Spohr, 2017, S. 316). Damit dienen die Anhörungen nicht nur zur Einholung von Expertise, sondern auch zur Einflusseinräumung für Interessenverbände und anderen Institutionen der öffentlichen Meinung (vgl. Kloepfer, 2005, S. 405). Dies macht sie zu einem effizienten Weg, Informationen von verschiedenen Gruppen zu diffundieren und im Dialog mit den Parlamentariern einen Gesamteindruck vom Gesetz und dem zugrundeliegenden Interessenspektrum zu erhalten (Blum, 2004, S. 107; Siefken und Schüttemeyer, 2013, S. 171). Ob dieser Weg von den unterschiedlichen Expertengruppen bei der Steuergesetzgebung auch unterschiedlich effektiv verwendet werden kann, soll nun Gegenstand der durchzuführenden Untersuchung sein.

4 Forschungsdesign

4.1 Hypothesenentwicklung und Operationalisierung

Zur Beantwortung der Forschungsfrage ("Welchen Expertengruppen folgen die Parlamentarier des Deutschen Bundestages bei der Implementierung von neuer Steuergesetzgebung?") sollen durch Zuhilfenahme von eingereichten schriftlichen Stellungnahmen zu den stattgefundenen Expertenanhörungen unter Differenzierung der Zugehörigkeit zur jeweiligen Expertengruppe Rückschlüsse auf erfolgreiche Einflussnahmen gezogen werden. Nach dem Studium relevanter Literatur (siehe Kapitel 2) ist

[4] Gegenzeichnung durch Bundesregierung, Unterschrift des Bundespräsidenten sowie Verkündung im Bundesgesetzblatt; jedoch keinerlei materiellen Änderungen mehr möglich.

Tabelle 1: Berücksichtigte Gesetzesvorhaben

	Name des Gesetzes	Verkündung	Anzahl	Sitzung	DIP-ID
A	Gesetz zum automatischen Austausch von Informationen über Finanzkonten in Steuersachen und zur Änderung weiterer Gesetze	30.12.2015	8	18/57	68566
B	Investmentsteuerreformgesetz	26.07.2016	10	18/78	73042
C	BEPS-Umsetzungsgesetz 1	23.12.2016	13	18/91	76153
D	Steuerumgehungsbekämpfungsgesetz	24.06.2017	11	18/107	78818
E	Gesetz gegen schädliche Steuerpraktiken im Zusammenhang mit Rechteüberlassungen	04.07.2017	12	18/109	79562
F	Gesetz zur Einführung einer Pflicht zur Mitteilung grenzüberschreitender Steuergestaltungen	30.12.2019	6	19/61	253781
G	Abzugsteuerentlastungsmodernisierungsgesetz	08.06.2021	8	19/131	272828
H	Fondsstandortgesetz	10.06.2021	7	19/126	272820
I	Steuervermeidungsabwehrgesetz	30.06.2021	6	19/141	276106
J	ATAD-Umsetzungsgesetz	30.06.2021	5	19/137	275938
K	Gesetz zur Umsetzung der DAC 7-Richtlinie	28.12.2022	8	20/28	290716

Hinweise zu Spalten: *Name des Gesetzes* beinhaltet den gängigen Namen des Gesetzes, welches im maßgeblichen Gesetzgebungsprozess durch die Bundesregierung eingebracht wurde. *Verkündung* weist das Datum der Verkündung des maßgeblichen Gesetzes aus. *Anzahl* zeigt die Anzahl der eingereichten Stellungnahmen für die jeweilige Anhörung an, die im Rahmen des Gesetzgebungsverfahrens im Finanzausschuss durchgeführt worden ist. *Sitzung* bezeichnet die Legislaturperiode sowie die Nummer der durchgeführten Sitzung des Finanzausschusses, in welcher die Anhörung stattgefunden hat. *DIP-ID* enthält die ID, welche das jeweilige parlamentarische Gesetzgebungsverfahren in einer Übersicht des Dokumentations- und Informationssystem für Parlamentsmaterialien (DIP) identifiziert.

keine klar zu favorisierende Expertengruppe erkennbar, die gegenüber anderen Gruppen in diesem Kontext im Vorteil sein sollte. In Untersuchungen über die Einflussstärke etwa bei der Etablierung neuer Rechnungslegungsstandards oder Gesetzgebungsprozesse an sich sieht die bisherige Literatur eher die Unternehmen gegenüber den anderen Gruppen im Vorteil. Vergleichsweise große Ressourcen sowie damit einhergehend die Möglichkeit zur besseren Bereitstellung von relevanter Expertise sind die dafür zugrundeliegenden Gründe. Doch bei der Betrachtung von Angelegenheiten der Besteuerung im internationalen Kontext sind Unternehmen nur zum Teil im Vorteil, denn die maßgebliche Wertschätzung von unabhängiger Expertise lässt vor allem Angehörige der Wissenschaft oder teilweise die steuerberatenden Berufe eine Favoritenrolle einnehmen. Angesichts dessen wird für die zu prüfende Hypothese nur die zu widerlegende Nullhypothese verwendet. Die für die folgende Untersuchung maßgebliche Hypothese lautet daher:

> **H_0:** *Bei der Steuergesetzgebung im Deutschen Bundestag berücksichtigen Parlamentarier die Ausführungen jeder Expertengruppe im selben Maße.*

Sollte es für eine oder mehrere Expertengruppen signifikante Unterschiede geben, wird die Nullhypothese abgelehnt, sodass von einem signifikanten Zusammenhang zwischen dem Lobbyerfolg und der betroffenen Gruppe ausgegangen werden kann.

Um den Umfang der Untersuchung im Rahmen einer Masterarbeit zu belassen, wird sich auf ertragsteuerliche Gesetzgebungsverfahren beschränkt, die vorwiegend Auswirkungen auf international tätigende Unternehmen bzw. auf die Behandlung von internationalen Geschäftsvorfällen haben. Dieses Besteuerungsfeld genießt seit den Anfängen des BEPS-Projekts eine große politische und gesellschaftliche Relevanz, weil dort in den vergangenen Jahren das für die letzten Jahrzehnte gültige globale Steuersystem grundlegende Veränderungen erfahren hat (vgl. OECD, 2015, S. 5f). Für die steuerliche Einkommensallokation und somit die Gewinnbesteuerung großer MNEs durch die einzelnen Staaten bedeutet das signifikante Änderungen, die national hauptsächlich durch die hier betrachteten Gesetze umgesetzt werden. Daneben ist das Feld relativ homogen, sodass große Messabweichungen in Abhängigkeit des Besteuerungsfeldes eher nicht zu erwarten sind.

Aus diesem Grund wird auch auf die Berücksichtigung der heterogenen Jahressteuergesetze verzichtet, die ein Sammelbecken für jegliche steuerlichen Änderungen aus allen Besteuerungsfeldern bilden, welche am jeweiligen Jahresende zusammen beschlossen werden, um im folgenden Jahr bereits Gültigkeit zu erlangen. Zeitlich wird die Untersuchung auf Gesetze beschränkt, die in den Legislaturperioden des 18., 19. und teilweise des 20. Deutschen Bundestages erfolgreich verabschiedet worden sind, was einen Zeitraum von Oktober 2013 bis Juli 2023 und somit fast zehn Jahren umfasst. Zusätzlich werden nur Vorhaben betrachtet, bei denen eine öffentliche Anhörungssitzung im Finanzausschuss stattgefunden hat, bei denen die geladenen Teilnehmer schriftliche Stellungnahmen einreichen konnten. Somit verbleiben elf Gesetzesvorhaben mit insgesamt 94 eingereichten Stellungnahmen, die in Tabelle 1 nachvollzogen werden können.

Weil jeder Gesetzgebungsprozess individuelle und in der Anzahl variierende Schwerpunkte setzt, sollen in dieser Untersuchung für jedes Vorhaben die jeweils fünf relevantesten Aspekte identifiziert werden. Dabei ist die Identifikation von Relevanz relativ schwierig, denn in den dürftig verfügbaren Nachrichten zu den Anhörungssitzungen existieren lediglich subjektive Aussagen der vertretenen Regierungsfraktionen über Erfolge und Ausführungen der Oppositionsfraktionen über die Misserfolge des behandelten Gesetzvorhabens. Aufgrund dessen wird die Relevanz der Schwerpunk-

te an der Häufigkeit der Nennungen in den Stellungnahmen bestimmt, um möglichst viele Einzelaussagen der Teilnehmer berücksichtigen zu können. Für jeden der insgesamt 55 Schwerpunkte wird sein jeweiliger Stand im Regierungsentwurf mit dem Stand der Ausschussfassung gegenübergestellt, was in dem jeweiligen Bericht des Finanzausschusses durch den Wort-zu-Wort Vergleich des Regierungsentwurfs mit dem Ausschussentwurf nachvollzogen werden kann. Diese Gegenüberstellung ist in Anhang 1 aufgeführt und ist maßgeblich für die Beurteilung des Lobbyerfolgs der Expertengruppen (dazu sogleich in 4.2).

Damit speist sich die Gesamtstichprobe von insgesamt 470 Datenpunkten (= N) aus den 94 Stellungnahmen und den jeweils zu den ausgewählten Schwerpunkten erhobenen fünf Aussagen. Im Zuge der Vorbereitung dieser Untersuchung haben sich die folgenden fünf Expertengruppen differenzieren lassen, die den jeweiligen Autoren zugeordnet werden können: Unternehmen, Steuerberater, Zivilgesellschaft, Wissenschaft und Autoritätspersonen des Staates, die durch ihren Beruf einen hoheitlichen Auftrag erfüllen (kurz: Autoritätsgruppe). Da bei Gesetz C die Autoren 12 und 13 lediglich für eine gesonderte Anhörung zu einem Vorhaben nur nationalen Interesses eingeladen worden sind, welches nichts mit dem eigentlich betrachteten Vorhaben gemein hat und dennoch im selben Gesetzgebungsprozess behandelt wird (bspw. Kindergelderhöhung u.ä.), werden diese Aussagen aus der Analyse ausgeklammert. Dies reduziert die Gesamtstichprobe auf 460 Datenpunkte, sollte aber hinsichtlich des untersuchten Besteuerungsfelds adäquatere Ergebnisse liefern.

4.2 Kodierungen

Der große Arbeits- und Zeitaufwand bei einer solchen qualitativen Inhaltsanalyse macht den Prozess für die Sicherung der Ergebnisse schwierig und komplex, weswegen ein angemessenes Kodierungsschema für die Reliabilität der Untersuchung unerlässlich ist (vgl. Yackee und Yackee, 2006, S. 131; Mellado und Parte, 2021, S. 15). In dieser Arbeit sind zwei Kodierungsverfahren zu unterscheiden: Zum einen die Zuordnung der Autoren der Stellungnahmen zu den Expertengruppen und zum anderen die Festlegung eines vorhandenen **Lobbyerfolgs**.

Letztere ist bereits Gegenstand in einer Vielzahl verschiedener Untersuchungen. In den frühen Stadien dieses Forschungsstrangs war es üblich, für die Bestimmung eines Erfolgs Präferenzen der Akteure nur als zustimmend oder ablehnend zu kategorisieren und neutrale Positionen auszusortieren (bspw. Watts und Zimmerman, 1978, S. 122; Saemann, 1999, S. 10). Aufgrund der Notwendigkeit eines erheblichen Urteilsvermögens ist eine solche Kodierung in der Regel zu vereinfachend, weshalb auch neutrale Aussagen, wie z.B. Meinungslosigkeit oder eine Mischung aus Zustimmung und Ablehnung, als gültige Positionen berücksichtigt werden müssen (vgl. Puro, 1984, S. 634f). Eine solche dreiteilige Einordnung ist in der Literatur zurzeit die verbreitetste Methodik (vgl. Mellado und Parte, 2021, S. 14). In überwiegend systematischen Szenarien wird sogar eine differenzierte Kategorisierung in fünf verschiedene Zustimmungsebenen durchgeführt: Von (1) vollständiger Ablehnung und (2) teilweiser Ablehnung über (3) Neutralität hin zu (4) teilweiser Zustimmung und (5) vollständiger Zustimmung erlaubt eine solche Skalierung differenziertere Ergebnisse und Erfolgszuordnungen für die Einflussnahme (bspw. Cross et al., 2021, S. 360). Eine ähnliche Kategorisierung haben teilweise Elschner und Hardeck (2022), die bei der Betrachtung der Änderungsrichtung nach jeweils (1) starker oder (2) moderater Verschärfung bzw. Entlastung (4/5) oder (3) Neutralität differenzieren (vgl. ebd., S. 317). Angesichts der unsystematischen Kommentarformate und der verschiedenen behandelten Aspekte der internationalen Steuergesetzgebung zwischen unterschiedlichen Gesetzesvorhaben wird in dieser Untersuchung auf ein solch differenziertes Spektrum verzichtet[5] und auf die gängige Dreiteilung zurückgegriffen:

Da in diesem Kontext die Änderungs**richtung** einer Änderung für den Lobbyerfolg maßgeblich ist, wird im Folgenden zwischen Rufen nach entlastenden, neutralen sowie verschärfenden Regelungen unterschieden. Als entlastend wird eine Änderung betrachtet, wenn im Vergleich zum Regierungsentwurf der Ausschuss die zu erwartenden administrativen oder monetären Aufwendungen eines betroffenen Steuerpflichtigen reduziert. Der maßgebliche Steuerpflichtige entspricht der Definition des Steuerpflichtigen im jeweils betrachteten Gesetz, welches durch den Regierungsentwurf geändert werden soll. Diese sind durch die Fokussierung der Stichprobe auf internationale Besteuerungsregelungen überwiegend Unternehmen. Im Umkehrschluss bedeutet eine zu erwartende Erhöhung solcher Aufwendungen durch die Änderungen im Ausschuss eine Klassifizierung als Verschärfung. Als neutral werden Aspekte gehandhabt, die keine Änderungen im Ausschuss erfahren oder nur solche Änderungen, die als Klarstellung dienen. Daher wird von einer erfolgreichen Beeinflussung ausgegangen und damit einem Datenpunkt ein Lobbyerfolg zugewiesen, wenn die von den Autoren präferierte Änderungsrichtung mit der Änderungsrichtung der Ausschussfassung übereinstimmt bzw. bei einer neutralen Position der Autoren zum jeweiligen Aspekt in der Ausschussfassung keine oder eine neutrale Änderung erfolgt. Allerdings wird bei einer neutralen geforderten Klarstellung nur von einem Lobbyerfolg ausgegangen, wenn diese mit der erfolgten Klarstellung einhergeht.[6] Dieses Vorgehen ist im Einklang mit den Vorgehensweisen von McLeay et al. (2000, S. 86) , Bamber und McMeeking (2016, S. 66) und Elschner und Hardeck (2022, S. 315). Die Vehemenz der ausformulierten Forderung oder der Grad der Übereinstimmung spielen in dieser Betrachtung keine Rolle, da die Ausgestaltung der Formulierung von Änderungspräferenzen regelmäßig taktischen Überlegungen der Verfasser unterliegen und damit nicht zwangsläufig ihre

[5] Bei der Kodierung der Stellungnahmen waren zu den jeweiligen Aspekten nur überaus selten starke Verschärfungen/Entlastungen festzustellen, weshalb die gröbere Betrachtung keinen nennenswerten Unterschied hervorbringen sollte.

[6] Vgl. zum besseren Verständnis die Kodierungsschemata in Anhang 3 sowie die Aufführung relevanter Textpassagen für jeden Datenpunkt im elektronischen Anhang 1.

wahren Interessen widerspiegeln; beispielsweise durch die Signalisierung von starker Opposition, um auch nur kleine Zugeständnisse im Gesetzgebungsverfahren zu erlangen (vgl. Eising und Spohr, 2017, S. 328).

Das andere Kodierungsverfahren betrifft die Zuordnung zu den **Expertengruppen**. In Abhängigkeit ihrer Eigenschaften und Affiliationen zum Einreichungszeitpunkt werden sämtliche Autoren der Stellungnahmen zu einer der fünf Expertengruppen zugerechnet: Als *Unternehmen* gelten Verfasser, die überwiegend wirtschaftliche Interessen für ihr Unternehmen oder ihre Mitglieder in ihrem Verband verfolgen. Unter *Steuerberater* fallen sämtliche Akteure, die den steuerberatenden Berufen angehören und/oder im Namen einer Steuerberatungsgesellschaft, Steuerberaterkammer oder eines zugehörigen Verbands agieren. Unter *Zivilgesellschaft* fallen alle unabhängigen und nicht profitorientierten Vereinigungen, die keine privatwirtschaftlichen oder staatlichen Akteure sind. Als *Wissenschaftler* gelten alle Autoren mit einer Lehrstuhlzugehörigkeit an einer Hochschule. Zuletzt werden alle Autoritätspersonen des Staates, die durch ihren Beruf einen hoheitlichen Auftrag erfüllen, der *Autoritäts*gruppe zugeordnet. In Letztere könnten auch die Wissenschaftler fallen, die im weiteren Sinne die Wissenschaftsfreiheit als Teil einer offenen und freiheitlichen Gesellschaft in der Regel in staatlichen Strukturen ausüben. Dies wird in späteren Robustheitstests (vgl. 5.3) berücksichtigt.

4.3 Schätzungsansatz mit analytischem Modell

Der so entstandene Datensatz wird nun zur Beantwortung der Forschungsfrage bzw. zur Prüfung der Hypothese herangezogen. Es werden Probit-Regressionen zu folgender Regressionsgleichung[7] durchgeführt:

$$\begin{aligned} POS_{ij} = \alpha &+ \beta_1 \text{TAX_PROF}_i + \beta_2 \text{CIVIL_SOC}_i \\ &+ \beta_3 \text{SCIENCE}_i + \beta_4 \text{AUTHORITY}_i \\ &+ \sum_{k=1}^{K} \gamma_k X_{ij} + \varepsilon_{ij} \end{aligned}$$

Die abhängige Variable POS_{ij} stellt einen Lobbyerfolg durch Autor i für den relevanten Aspekt j dar, indem sie den Wert 1 annimmt und bei einem fehlenden Lobbyerfolg entsprechend eine 0. Die unabhängigen Variablen bilden die jeweilige Expertengruppe ab, wobei die Unternehmen (BUSINESS_i) als Basisgruppe fungieren. Die Binärvariablen nehmen für die zugewiesene Expertengruppe für Autor i nach dem β_i eine 1 an und für die übrigen korrespondierend eine 0. Daneben werden noch im Summenzeichen die zu prüfenden Kontrollvariablen abgebildet, die durch X_{ij} dargestellt werden.

[7] Die Voraussetzungen zur Nutzung dieser Regression wurden geprüft und erfüllt.

Für die Anwendung von Kontrollvariablen ergeben sich diverse Möglichkeiten: Um einen möglichen Einfluss der Autorenrelevanz zu prüfen, wird als Annäherung die Größe der Institution mit der ordinalen Variable *SIZE_RANK* verwendet. In Anlehnung an Mahoney (2007, S. 41) und Elschner und Hardeck (2022, S. 316) wird die Größe dabei nach drei Kategorien differenziert (groß, mittel und klein). Dabei kann bei der Einteilung aufgrund der Gruppeneigenschaften, -diversität und die damit einhergehende schwierige Vergleichbarkeit nicht auf triviale Größenkennzahlen zurückgegriffen werden (vgl. ebd.). Die Einteilung erfolgt vielmehr nach groben Kriterien: So gelten Verfasser als klein, wenn sie als Einzelperson oder für ein Unternehmen auftreten. Eine mittlere Größe wird bei Gruppierungen angenommen, die eine nicht unerhebliche Anzahl an Mitgliedern oder Mitarbeitern aufweisen können. Als groß gelten Institutionen, die eine Mehrheit ihrer zugehörigen Gruppe hinter sich versammeln können, indem sie beispielsweise mehrere Verbände vereinen oder innerhalb ihrer Profession als großer Akteur gelten.

Neben der Relevanz kann auch die Reputation des Autors bzw. der kommentierenden Institution durch die wahrgenommene Expertise für den Lobbyerfolg ausschlaggebend sein. In diesem Kontext wird näherungsweise davon ausgegangen, dass eine höhere Reputation vorhanden ist, wenn der betroffene Akteur öfter zu den untersuchten Expertenanhörung eingeladen wird: *INVITATIONS* beinhaltet daher die Anzahl an Einladungen zu den untersuchten Anhörungen. Daneben wird mit *ANSWERS* die Häufigkeit der Nennung des Sachverständigen im Wortprotokoll der Anhörung gemessen. Es ist anzunehmen, dass Sachverständige eher häufiger befragt werden oder auch auf die Ausführungen des Befragten im späteren Anhörungsverlauf referenziert wird, wenn die Abgeordneten ihnen eine relevante Expertise auf ihrem Feld zutrauen. Daneben könnte die angestrebte Änderungsrichtung der Kommentierenden einen relevanten Einflussaspekt darstellen. Hierfür wird mit der nominalen Variable *CHANGE* die jeweilige präferierte Änderungsrichtung (entlastend, neutral oder verschärfend) bei dem betroffenen Aspekt abgebildet (ähnlich zu Elschner und Hardeck, 2022, S. 317). Außerdem könnte es einen Einfluss auf den Lobbyerfolg haben, wenn der Autor bereits zum maßgeblichen Referentenentwurf eine Stellungnahme abgegeben hat; dies wird mit der Binärvariable *FIRST_DRAFT* abgebildet. Des Weiteren wird mit *PAGES* geprüft, inwiefern die Länge der abgegebenen Stellungnahmen mit einem positiven Lobbyerfolg zusammenhängt. Es ist davon auszugehen, dass längere Ausarbeitungen detailreichere Ausführungen beinhalten, die mehr nützliche Informationen für die Abgeordneten bedeuten. Wie bereits dargelegt sind solche wichtig, um von den Parlamentariern berücksichtigt zu werden.

Somit variieren *ANSWERS, FIRST_DRAFT* und *PAGES* auf Anhörungsebene, *SIZE_RANK* und *INVITATIONS* auf Autorenebene und *CHANGE* auf Aspektebene.

Tabelle 2: Deskriptive Statistiken

Variable	N	Mittelwert	Median	Std.-Abweichung	Min.	Max
POS	460	0,33	0	0,469	0	1
Expertengruppen						
BUSINESS	460	0,28	0	0,451	0	1
TAX_PROF	460	0,20	0	0,397	0	1
CIVIL_SOC	460	0,14	0	0,349	0	1
SCIENCE	460	0,22	0	0,413	0	1
AUTHORITY	460	0,16	0	0,370	0	1
Kontrollvariablen						
CHANGE						
Relaxing	267	0,48	0	0,501	0	1
Neutral	267	0,43	0	0,496	0	1
Tightening	267	0,09	0	0,287	0	1
ANSWERS	460	10,67	10	5,223	3	29
FIRST_DRAFT	460	0,39	0	0,489	0	1
PAGES	460	12,60	9	10,031	2	62
SIZE_RANK						
SIZE_RANK1	460	0,48	0	0,500	0	1
SIZE_RANK2	460	0,23	0	0,420	0	1
SIZE_RANK3	460	0,29	0	0,456	0	1
SIZE_MEMBERS	460	7,26	7,74	5,450	0	15,58
INVITATIONS	460	3,85	3	2,885	1	9

5 Ergebnisse

5.1 Deskriptive Statistiken und qualitative Analyse

Tabelle 2 fasst die deskriptiven Statistiken aller in dieser Untersuchung verwendeten Variablen zusammen.

In Tabelle 3 wird innerhalb der untersuchten Stichprobe die Anzahl der eingereichten Stellungnahmen nach der jeweiligen Expertengruppe aufgeschlüsselt und deren Anteile an der Gesamtzahl dargelegt. Außerdem werden die aus den Stellungnahmen erwachsenen Positionen zu den relevanten Aspekten sowie die davon als Lobbyerfolg eingeschätzten Fälle inklusive einer Erfolgsrate aufgeführt.

Tabelle 3: Stellungnahmen nach Expertengruppen inkl. Erfolgsrate

	Stellungnahmen		Positionen		
	Anzahl	Anteil	Gesamt	$POS = 1$	Erfolgsrate
BUSINESS	26	28,26%	130	36	27,69%
TAX_PROF	18	19,57%	90	30	33,33%
CIVIL_SOC	13	14,13%	65	27	41,54%
SCIENCE	20	21,74%	100	31	31,00%
AUTHORITY	15	16,30%	75	26	34,67%
Summe	92	100%	460	150	32,61%

Aus der Tabelle geht hervor, dass Unternehmen die meistrepräsentierte Gruppe darstellen: Trotz einer Differenzierung in fünf Expertengruppen konnten sie mit 26 Einreichungen mehr als ein Viertel (28%) aller Stellungnahmen an den Finanzausschuss senden. Demgegenüber waren Vertreter der Zivilgesellschaft und der Autoritätsgruppe mit jeweils 13 (14%) bzw. 15 Einreichungen (16%) vergleichsweise unterrepräsentiert. Diese Feststellung für die Zivilgesellschaft ist zwar im Einklang mit der Unterrepräsentierung in den Ergebnissen von Elschner und Hardeck (2022, S. 321), steht jedoch unter verschiedenen Vorzeichen: Während sie in jenem Papier aufgrund eines anzunehmenden Defizits an Expertise von sich aus weniger Kommentare abgab, sind in dieser Untersuchung die Akteure von einer Einladung der dem Ausschuss angehörenden Abgeordneten abhängig. Daher wäre ein möglicher Grund für die Unterzahl die Einräumung einer im Verhältnis nur eingeschränkten Expertise durch die Abgeordneten. Allerdings müssen hier auch parteipolitische Hintergedanken berücksichtigt werden, die in diesem Kontext jedoch nicht untersucht werden können. Die beiden übrigen Gruppen Steuerberater und Wissenschaft weisen mit 18 bzw. 20 Einreichungen (20% bzw. 22%) eine ausgewogene Beteiligung auf.

Bei der Betrachtung positiver Lobbyversuche haben die Unternehmen auch in absoluten Zahlen (36) im Vergleich zu den übrigen Gruppen die meisten Erfolge, während Vertreter der Zivilgesellschaft und der Autoritätsgruppe erneut auf den hinteren Plätzen (27 bzw. 26) zu verorten sind. Allerdings verändert sich die Erfolgsbetrachtung bei der adäquateren Betrachtung des relativen Erfolges grundlegend: Mit einer Erfolgsrate von fast 42% sind Akteure der Zivilgesellschaft im Vergleich zu den übrigen Gruppen am erfolgreichsten darin, die relevanten Aspekte in ihrem Sinne zu beeinflussen. Demgegenüber können in diesem Setting Unternehmen trotz ihres Vorteils in der Anzahlbetrachtung nur mit einer Erfolgsrate von knapp 28% keine andere Expertengruppe übertreffen. Die übrigen Gruppen sind dazwischen zu verorten: Während die Steuerberater mit exakt einem Drittel ihrer Positionen erfolgreich sind, hat die Autoritätsgruppe eine marginal höhere bzw. die Wissenschaftler eine marginal niedrigere Erfolgsrate (31% bzw. 35%) zu verzeichnen. Bei Betrachtung der Gesamterfolgsrate von ca. 33% ist jedoch zu betonen, dass keine Expertengruppe sonderlich weit von dieser Rate entfernt ist, was noch relevant für die Regressionsanalysen sein wird. Der größtmögliche Unterschied besteht zwischen Zivil-

gesellschaft und den Unternehmen mit einer Differenz von fast 14%-Punkten. Die deskriptiven Ergebnisse sind teilweise im Einklang mit den deskriptiven Ergebnissen von Elschner und Hardeck (2022, S. 321), da dort die Zivilgesellschaft zumindest bei der Signalisierung ihrer Positionen ebenfalls die höchste Erfolgsrate aufweist und in dieser Arbeit keine Differenzierung hinsichtlich einfacher Signalisierung und nützlicher Information durchgeführt wird.

Tabelle 4: Präferierte Änderungsrichtung nach Expertengruppe

	N	Relaxing		Neutral		Tightening	
		abs.	rel.	abs.	rel.	abs.	rel.
BUSINESS	73	51	69,9%	22	30,1%	0	0%
TAX_PROF	66	46	69,7%	20	30,3%	0	0%
CIVIL_SOC	36	2	5,6%	21	58,3%	13	36,1%
SCIENCE	58	25	43,1%	28	48,3%	5	8,6%
AUTHORITY	34	5	14,7%	23	67,6%	6	17,6%
Summe	267	129	48,31%	114	42,7%	24	8,99%

In der gesamten Stichprobe existieren 267 Datenpunkte, bei denen sich die Autoren zum jeweiligen Aspekt explizit geäußert haben und somit auch eine Änderungsrichtung festgestellt werden konnte. Tabelle 4 schlüsselt die Datenpunkte nach präferierter Änderungsrichtung und korrespondierender Expertengruppe auf. In der Gesamtbetrachtung sind neutrale und entlastende Änderungen in etwa auf einem vergleichbaren Niveau vertreten, wobei letztere in der Mehrzahl sind (48% vs. 43%). Forderungen nach verschärfenden Änderungen sind hingegen mit nur 9% der Änderungspräferenzen eindeutig in der Minderheit. Dabei sind starke Differenzen hinsichtlich der gewünschten Richtung erkennbar, sofern man die einbezogenen Expertengruppen untereinander vergleicht. So wünschen Unternehmen und Steuerberater gleichermaßen in jeweils ca. 70% der Fälle im Vergleich zum Regierungsentwurf eine entlastende Anpassung des Aspekts. In den übrigen Fällen (jeweils ca. 30%) führen diese Gruppen neutrale Argumente an. Keine der beiden Expertengruppen präferiert in der Stichprobe jemals eine verschärfende Änderung durch den Ausschuss. Dies ist zumindest für Unternehmen auch nicht überraschend, denn diese haben ein nachvollziehbares Interesse zur Gewinnmaximierung und damit einhergehend eine Minimierung von Steuer- oder Offenlegungsaufwendungen. Bemerkenswert ist jedoch die Ähnlichkeit des Gesamtverhaltens von Unternehmen und Steuerberater. Da die bisherige Literatur mögliche Verzerrungen durch Steuerberater für möglich, aber nicht ausgeprägt hält, sondern eher von einer unabhängigen Expertise ausgegangen wird, erschweren nun die vorliegenden Ergebnisse die Vertretung dieser Sichtweise. Möglicherweise könnte für ein deutsches Setting bei der internationalen Besteuerung der Bund zwischen Unternehmen und Steuerberatern stärker sein als angenommen. Dies könnte Gegenstand für weitergehende Untersuchungen in anderen Arbeiten sein.

Die Hälfte aller verschärfenden Änderungsforderungen werden durch die Zivilgesellschaft erhoben. Diese hat in etwas mehr als jedem dritten Fall (36%) einer gewünschten Änderung nach strengeren Regelungen verlangt. Im Kontrast zu den vorher genannten Gruppen sind Rufe nach Entlastungen bei den Vertretern der Zivilgesellschaft eine Seltenheit (6%). Vielmehr weisen die übrigen Gruppen mehrheitlich eine Tendenz zu neutralen Aussagen auf, wobei die Autoritätsgruppe mit fast 68% ihrer Aussagen den größten Anteil an Neutralität vertritt, gefolgt von Zivilgesellschaft (58%) und Wissenschaft (48%). Letztere Gruppe präferiert jedoch ebenfalls relativ häufig entlastende (43%) und relativ selten verschärfende Änderungen (9%) in der Stichprobe. Einzig die Autoritätsgruppe ist bei der Betrachtung von entlastenden und verschärfenden Präferenzen relativ ausgeglichen (15% vs. 18%).

5.2 Ergebnisse der Regressionsanalysen

Tabelle 5 zeigt die Ergebnisse der Probit-Schätzungen für die Regressionsgleichung, die den Lobbyerfolg in Abhängigkeit der Expertengruppe abbildet. Dort werden in Spalten abgetrennt fünf unterschiedliche Variationen von angewendeten Kontrollvariablen und der Einbeziehung von sogenannten Nullaussagen dargelegt. Eine Nullaussage gilt als solche, sofern der Autor sich zum jeweils betrachteten Aspekt gar nicht äußert, weswegen in einem solchen Fall generell nicht von einem erfolgreichen Lobbyversuch ausgegangen wird. Bei einem Ausschluss der Nullaussagen reduziert sich zwar die Anzahl der Gesamtbeobachtungen, doch die Stichprobe enthält dann nur noch tatsächlich stattgefundene Lobbyversuche. Zunächst wird das einfache Modell ohne Kontrollvariablen betrachtet (1). Die Schätzung zeigt in diesem Fall eine im Vergleich zu Unternehmen höhere Erfolgswahrscheinlichkeit für alle übrigen Expertengruppen an. Während Wissenschaftler nur marginal erfolgreicher eingeschätzt werden, hat die Zivilgesellschaft die höchste Wahrscheinlichkeit für einen Lobbyerfolg.

Dies ist im Einklang mit den Ergebnissen aus Kapitel 5.1. Im Gegensatz zu den übrigen Gruppen ist das Ergebnis der Zivilgesellschaft zumindest auf dem 10%-Signifikanzniveau auch signifikant (p-Wert = 0,053).

Steuerberater und die Autoritätsgruppe haben eine ähnlich hohe Erfolgswahrscheinlichkeit und reihen sich zwischen Wissenschaft und Zivilgesellschaft ein. Mit einem Pseudo R^2 von 0,007 wird die Gesamtgüte des ersten Modells mit den folgenden Variationen verglichen. Eine isolierte Betrachtung des Pseudo R^2 ohne Relation zu ähnlichen Modellen mit identischem Datensatz ist für die Beurteilung der Modellgüte unzureichend, wobei Vergleiche von Modellen mit unterschiedlicher Datengrundlage unzulässig sind (vgl. Long und Freese, 2001, S. 191).

Unter Zuhilfenahme erster Kontrollvariablen wie die Anzahl der Nennungen des Autors in der Anhörung, der öffentlichen Beteiligung an Diskussionen am Referentenentwurf und der Seitenanzahl der eingereichten Stellungnahme wird in (2) ein weitergehendes Modell betrachtet, welches die Expertise und relevante Informationen gesondert berücksichtigt. Im Vergleich zu (1) hat sich die Erfolgswahrscheinlichkeit für die Wissenschaftler fast an die der Steuerberater angeglichen. Letztere hat sich fast gar nicht verändert. Eine Erhö-

Tabelle 5: Probit-Schätzungen der Regressionsgleichung

Variable	(1)	(2)	(3)	(4)	(5)
Expertengruppen					
TAX_PROF	0,161	0,186	-0,115	-0,216	-0,300
	(0,180)	(0,187)	(0,224)	(0,245)	(0,274)
CIVIL_SOC	0,378*	0,469**	0,629**	0,655*	-0,019
	(0,196)	(0,229)	(0,312)	(0,350)	(0,420)
SCIENCE	0,096	0,181	0,129	0,094	-0,260
	(0,176)	(0,223)	(0,290)	(0,331)	(0,380)
AUTHORITY	0,198	0,278	0,707**	0,681**	-0,089
	(0,189)	(0,223)	(0,320)	(0,323)	(0,382)
Kontrollvariablen					
CHANGE					
Relaxing					-1,648***
					(0,217)
Tightening					-1,191***
					(0,335)
ANSWERS		0,002	-0,019	-0,014	0,009
		(0,013)	(0,017)	(0,017)	(0,019)
FIRST_DRAFT		0,098	0,050	-0,010	0,054
		(0,160)	(0,202)	(0,249)	(0,274)
PAGES		0,004	-0,011	-0,009	-0,011
		(0,007)	(0,009)	(0,009)	(0,010)
SIZE_RANK					
SIZE_RANK2				-0,303	-0,429
				(0,337)	(0,381)
SIZE_RANK3				-0,296	-0,455
				(0,371)	(0,425)
INVITATIONS				-0,058	-0,064
				(0,048)	(0,055)
Konstante	-0,592***	-0,753***	0,327	0,692	1,780***
	(0,117)	(0,207)	(0,287)	(0,439)	(0,531)
Nullaussagen	Ja	Ja	Nein	Nein	Nein
Beobachtungen	460	460	267	267	267
Pseudo R^2	0,007	0,009	0,052	0,056	0,252

Anmerkungen: Die Tabelle zeigt die sich aus Probit-Schätzungen ergebenden Regressionskoeffizienten der Regressionsgleichung für variierende Kontrollvariablen. Referenzgruppen: *BUSINESS*, *Neutral* und *SIZE_RANK1*. Abhängige Variable: *POS*. Die zugehörigen Standardabweichungen sind jeweils darunter in Klammern. *, ** und *** repräsentieren jeweils Signifikanzniveaus von 10%, 5% und 1%. Alle Variablen sind in Anhang 2 definiert.

hung gab es ebenfalls bei der Autoritätsgruppe und bei der Zivilgesellschaft, wobei letztere nun auch auf dem 5%-Niveau eine signifikant höhere Erfolgswahrscheinlichkeit hat. Die Ergebnisse der übrigen Gruppen sind weiterhin nicht signifikant. Auch die Einflüsse der hinzugezogenen Kontrollvariablen haben keinen statistisch relevanten Effekt. Die Gesamtgüte des Modells hat sich mit einem Pseudo R^2 von 0,009 etwas verbessert.

Dasselbe Modell wird nun unter Ausschluss der Nullaussagen in (3) erneut geschätzt. Hieraus erwachsen essenzielle Unterschiede in den zu interpretierenden Ergebnissen: Die zunächst erfolgreicher einzustufenden Steuerberater werden nun schlechter als Unternehmen eingeschätzt. Die Angehörigen der Wissenschaft büßen auch etwas Erfolgswahrscheinlichkeit ein, bleiben aber vor den Unternehmen. Allerdings sind beide Koeffizienten nicht signifikant. Statistische Signifikanz (5%-Niveau) ist jedoch neben den Angehörigen der Zivilgesellschaft nun auch bei der Autoritätsgruppe festzustellen. Beide Gruppen haben im Vergleich zu (1) und (2) auch ihre jeweils höchste Einflussintensität, wobei die Autoritätsgruppe in diesem Fall sogar erfolgreicher als die Zivilgesellschaft eingeschätzt wird. Obwohl die Kontrollvariablen weiterhin insignifikante Koeffizienten aufweisen, erhöht sich das Bestimmtheitsmaß des Modells auf ein Pseudo R^2 von 0,052.

Die signifikant größeren Chancen für erfolgreiches Lobbying von Zivilgesellschaft und der Autoritätsgruppe sollen nun in (4) mit weiteren Kontrollvariablen hinsichtlich der Relevanz der maßgeblichen Akteure untersucht werden. Unter Berücksichtigung eines kategorialen Größenparameters und der Anzahl der Einladungen innerhalb der Stichprobe ändern sich die Ergebnisse nur marginal. Während Steuerberater gegenüber den Unternehmen nochmals an Einflussintensität einbüßen und Wissenschaftler nur noch geringfügig erfolgreicher als letztere einzuschätzen sind (beides abermals nicht signifikant), nähern sich die Koeffizienten von Autoritätsgruppe und Zivilgesellschaft ein wenig an, bleiben aber weiterhin hoch und signifikant. Allerdings ist der Schätzer für die Zivilgesellschaft nur noch für das 10%-Niveau signifikant, während die Autoritätsgruppe innerhalb des 5%-Niveaus am erfolgreichsten bleibt. Die Effekte der bisherigen Kontrollvariablen bleiben marginal und nicht signifikant. Die

Koeffizienten der neu hinzugezogenen Größenkriterien suggerieren im Vergleich zu großen Institutionen schlechtere Erfolgschancen für mittlere und kleine Institutionen. Allerdings sind diese sowie die Anzahl an Einladungen im Modell ebenfalls nicht signifikant. Die Gesamtmodellgüte verbessert sich erneut auf ein Pseudo R^2 von 0,056.

Bei der Berücksichtigung der präferierten Änderungsrichtung als Kontrollvariable im Gesamtmodell (5) ändern sich die bisherigen Ergebnisse jedoch grundlegend: Sämtliche Expertengruppen werden nun gegenüber Unternehmen mit einer schlechteren Erfolgswahrscheinlichkeit eingeschätzt. Während Zivilgesellschaft und Autoritätsgruppe minimal unter den Unternehmen liegen, werden Wissenschaft und Steuerberater wesentlich schlechter eingeschätzt. Dieser Umstand würde vorherige Annahmen eines positiven Einflusses von unabhängiger Expertise konterkarieren, welcher angesichts dieser Ergebnisse in diesem Fall wohl nicht relevant wäre. Allerdings ist nun keiner der Koeffizienten für die Expertengruppen signifikant, weshalb hieraus keine endgültigen Schlüsse gezogen werden sollten. Hochsignifikant (<1%) ist vielmehr der Einfluss der Änderungsrichtung: Sowohl entlastende als auch verschärfende Änderungspräferenzen der Autoren haben weitaus schlechtere Chancen für erfolgreiches Lobbyieren als neutrale Positionen, wobei Rufe nach entlastenden Änderungen noch geringere Erfolgschancen aufweisen als verschärfende Rufe. Die übrigen Kontrollvariablen zeigen bis auf das Größenkriterium weiterhin keinen nennenswerten Effekt. Die Wahrscheinlichkeit für mittlere und kleine Institutionen verschlechtert sich jedoch nochmals. Allerdings sind bis auf die Änderungsrichtung sämtliche andere Kontrollvariablen nicht signifikant. Die Güte des Gesamtmodells hat sich bei Betrachtung des Pseudo R^2 (= 0,252) im Vergleich zu den übrigen Modellen deutlich verbessert.

Angesichts der Dominanz der Veränderungsvariable wurde mithilfe von Varianzinflationsfaktoren (VIF) sowie einer Korrelationsmatrix die Möglichkeit eines Multikollinearitätsproblems insb. für die letzte Modellspezifikation geprüft (nur im Code), wobei die Ergebnisse bei allen Spezifikationen nicht für ein solches Problem sprechen und das Modell deshalb als ökonomisch und ökonometrisch sinnvolles Referenzmodell für weitere Robustheitstests verwendet wird.

In der Gesamtbetrachtung bietet sich eine zweiteilige Ergebnisinterpretation an: Ohne Kontrolle der Änderungsrichtung ist eine signifikant höhere Wahrscheinlichkeit eines Lobbyerfolgs in Expertenanhörungen gegeben, wenn die Aktivität von der Autoritätsgruppe ausgegangen ist. Im Vergleich zu Unternehmen haben Angehörige dieser Gruppe eine 24% größere Erfolgswahrscheinlichkeit (marginaler Effekt). Bei optionaler Berücksichtigung eines schwächeren Signifikanzniveaus (10%) ist ein ähnliches Ergebnis auch bei Vertretern der Zivilgesellschaft festzustellen, die gegenüber den Unternehmen ebenfalls eine 24% größere Wahrscheinlichkeit für einen Lobbyerfolg ausweisen (marginaler Effekt). Allerdings verschwinden sämtliche signifikanten Ergebnisse für die Expertengruppen, sobald auch für die Änderungsrichtung kontrolliert wird. Diese scheint einen weitaus größeren Einfluss für einen möglichen Lobbyerfolg zu spielen als die korrespondierenden Expertengruppen, da die zugehörigen Regressionskoeffizienten der Variablen als einzige (hoch-)signifikant sind. Sowohl entlastende als auch verschärfende Änderungspräferenzen vermindern die Erfolgswahrscheinlichkeit im Vergleich zu neutralen Positionierungen deutlich. Unter Berücksichtigung marginaler Effekte haben entlastende bzw. verschärfende Änderungsforderungen im Vergleich zu neutralen Aussagen eine 58% bzw. 44% geringere Wahrscheinlichkeit für einen Lobbyerfolg. Die insignifikanten Koeffizienten der Gruppenvariablen lassen die Ablehnung der Nullhypothese nicht zu, weshalb innerhalb dieser Untersuchung weiterhin von einer gleichmäßigen Berücksichtigung der Expertengruppen durch die Parlamentarier ausgegangen werden muss.

5.3 Robustheitstests

Um die Ergebnisbetrachtung aus dem vorherigen Unterkapitel zu validieren, werden verschiedene Robustheitstests durchgeführt und ihre Ergebnisse in den Tabellen 6 und 7 dargestellt.

In (1) wird mit *SIZE_MEMBERS* ein weiteres Maß für Relevanz erhoben, indem hier der natürliche Logarithmus der Mitgliederanzahl der kommentierenden Institution aufgenommen wird (vgl. Elschner und Hardeck, 2022, S. 335). Die bisherige Größenkontrolle nach drei Kategorien könnte zu grob ausfallen, was mit dieser kardinalen Variable geprüft wird. An den Ergebnissen zu (1) ist jedoch keinerlei nennenswerte Änderung zum Referenzmodell abzulesen, sodass die ordinale Größeneinteilung bestehen bleiben kann.

Spezifikationen (2) und (3) variieren die Kodierung von *POS*, um bei einer möglichst lockeren (2) und möglichst strengen (3) Kodierung des Lobbyerfolgs die Bandbreite der Kodierungsmöglichkeiten zu überprüfen. Trotz festem Kodierungsschema sind nicht alle Sachverhalte dazu geeignet, streng dichotom betrachtet zu werden, weshalb anhand der betrachteten Aussage indirekte Präferenzen antizipiert wurden. Solche Fälle werden in diesen zwei Betrachtungen nun nicht mehr persönlich abgewogen, sondern durchgängig als Lobbyerfolg (2) bzw. kein Lobbyerfolg (3) kategorisiert. Damit wird versucht, der Restsubjektivitätsproblematik entgegenzuwirken, indem ein Einteilungsspektrum abgebildet wird. Im Datensatz wurden 44 Datenpunkte identifiziert, die in die Kodierungsvariation von (2) und (3) einbezogen werden.[8] Bei Sichtung der Ergebnisse zeigt sich, dass bei beiden Extremfällen keine signifikanten Änderungen bei unabhängigen Variablen und Kontrollvariablen auftreten. Zwar sind Angehörige der Autoritätsgruppe und Zivilgesellschaft jeweils teilweise wieder erfolgreicher einzuschätzen als Unternehmen, doch dieser Effekt ist angesichts fehlender Signifikanz weiterhin vernachlässigbar. Bemerkenswert ist die abnehmende Effektstärke bei einer vorliegenden

[8] Vgl. auch Tabellenblatt “Robustheitstests 2 + 3” im elektronischen Anhang 1.

Tabelle 6: Probit-Schätzungen unter Berücksichtigung von Robustheitstests

Variable	(1)	(2)	(3)	(4)	(5)
Expertengruppen					
TAX_PROF	-0,285	-0,253	-0,315	-0,346	-0,312
	(0,271)	(0,273)	(0,258)	(0,263)	(0,273)
CIVIL_SOC	-0,067	0,061	0,042	-0,015	-0,046
	(0,443)	(0,429)	(0,371)	(0,396)	(0,415)
SCIENCE	-0,105	-0,332	-0,299	-0,244	
	(0,366)	(0,381)	(0,349)	(0,353)	
AUTHORITY	-0,005	0,081	-0,116	-0,140	-0,172
	(0,379)	(0,393)	(0,340)	(0,362)	(0,325)
Kontrollvariablen					
CHANGE					
Relaxing	-1,631***	-1,628***	-0,992***	-1,852***	-1,664***
	(0,215)	(0,220)	(0,193)	(0,207)	(0,215)
Tightening	-1,192***	-1,135***	-0,652**	-1,402***	-1,189***
	(0,335)	(0,348)	(0,314)	(0,327)	(0,335)
ANSWERS	0,005	0,011	-0,016	0,007	0,009
	(0,019)	(0,019)	(0,019)	(0,019)	(0,019)
FIRST_DRAFT	0,095	0,073	0,222	0,107	0,057
	(0,269)	(0,277)	(0,259)	(0,260)	(0,274)
PAGES	-0,013	-0,011	-0,006	-0,015	-0,012
	(0,010)	(0,010)	(0,009)	(0,010)	(0,010)
SIZE_RANK					
SIZE_RANK2		-0,378	0,028	-0,315	-0,430
		(0,379)	(0,357)	(0,355)	(0,382)
SIZE_RANK3		-0,383	0,126	-0,444	-0,528
		(0,423)	(0,381)	(0,393)	(0,394)
SIZE_MEMBERS	0,044				
	(0,041)				
INVITATIONS	-0,063	-0,045	-0,014	-0,064	-0,064
	(0,059)	(0,056)	(0,049)	(0,053)	(0,055)
Konstante	1,233***	1,698***	0,660	2,042***	1,826***
	(0,368)	(0,527)	(0,470)	(0,495)	(0,525)
Nullaussagen	Nein	Nein	Nein	Nein	Nein
Beobachtungen	267	267	267	322	267
Pseudo R^2	0,252	0,254	0,114	0,325	0,252

Anmerkungen: Die Tabelle zeigt die sich aus Probit-Schätzungen ergebenden Regressionskoeffizienten der Regressionsgleichung für das Referenzmodell unter Anwendung diverser Robustheitstests. Referenzgruppen: *BUSINESS*, *Neutral* und *SIZE_RANK1*. Abhängige Variable: *POS*. Die zugehörigen Standardabweichungen sind jeweils darunter in Klammern. *, ** und *** repräsentieren jeweils Signifikanzniveaus von 10%, 5% und 1%. Alle Variablen sind in Anhang 2 definiert.

Änderungspräferenz in der Konstellation mit strenger Erfolgskodierung und die schwächere (aber ausreichende) Signifikanz des Koeffizienten für verschärfende Änderungen. Die gleichzeitige Halbierung des Maßes für Modellgüte ist eine auffallende Gegebenheit, welche aufgrund der veränderten Datengrundlage jedoch nicht relevant ist.

In Variation (4) wird die Lobbyerfolgsdefinition für Nulläußerungen bei Betrachtung von Aspekten ohne Änderung durch den Finanzausschuss modifiziert: Bisher wurde davon ausgegangen, dass eine Nulläußerung (= keine Äußerung) zu einem Aspekt immer keinen Lobbyerfolg bedeutet, da der Autor in diesem Fall keine aktive Anstrengung für den Erfolg unternommen hat. Angesichts einer fehlenden Änderung eines Aspekts und einer hierzu fehlenden Äußerung könnte jedoch auch eine implizite bzw. nicht geäußerte Zustimmung des Autors vorliegen, weshalb er sich dazu aus Platzgründen o.ä. nicht mehr äußern wollte. Unter dieser Variation wird in solchen Fällen nun doch ein Lobbyerfolg angenommen. Diese Anpassung erhöht die betrachteten Beobachtungen um 55.[9] Da diese bei der Änderungsrichtung nun stets als erfolgreich und neutral einzuordnen sind, überrascht die Verschlechterung der Erfolgswahrscheinlichkeit für verschärfende bzw. entlastende Änderungspräferenzen nicht. Dennoch ändert sich die Effektstärke und die Signifikanz der einzelnen Expertengruppen weiterhin nicht nennenswert, sodass die bisherige Lobbyerfolgsdefinition beibehalten wird. Das erhöhte Pseudo R^2 ist auf die veränderte Datengrundlage zurückzuführen.

In Spezifikation (5) werden die Wissenschaftler ebenfalls der Autoritätsgruppe zugeordnet, da erstere im weiteren Sinne auch der Definition der letzteren Gruppe entsprechen (vgl. Ende von Kapitel 4.2). Eine Reduzierung der Gruppenanzahl könnte darüber hinaus ebenfalls Auswirkungen auf die Ergebnisse erzielen. Allerdings sind solche Auswirkungen in dieser Untersuchung nicht zu beobachten: Regressionskoeffizienten und Signifikanzniveaus verbleiben bei einem ähn-

[9] Vgl. auch Tabellenblatt “Robustheitstest 4” im elektronischen Anhang 1.

Tabelle 7: Probit-Schätzungen unter Berücksichtigung von Robustheitstests

Variable	(6)	(7)	(8)
Expertengruppen			
TAX_PROF	-0,527	-0,301	-0,534
	(0,436)	(0,274)	(0,438)
CIVIL_SOC	-0,063	-0,070	-0,121
	(0,427)	(0,458)	(0,468)
SCIENCE	-0,300	-0,250	-0,290
	(0,388)	(0,382)	(0,389)
AUTHORITY	-0,137	-0,076	-0,122
	(0,392)	(0,387)	(0,398)
TAX_PROF x Relaxing	0,304		0,311
	(0,459)		(0,460)
CIVIL_SOC x Tightening		0,176	0,199
		(0,654)	(0,656)
Kontrollvariablen			
CHANGE			
Relaxing	-1,738***	-1,655***	-1,748***
	(0,259)	(0,218)	(0,261)
Tightening	-1,214***	-1,269***	-1,302***
	(0,338)	(0,443)	(0,447)
ANSWERS	0,009	0,009	0,010
	(0,019)	(0,019)	(0,019)
FIRST_DRAFT	0,065	0,060	0,072
	(0,274)	(0,275)	(0,275)
PAGES	-0,011	-0,011	-0,011
	(0,010)	(0,010)	(0,010)
SIZE_RANK			
SIZE_RANK2	-0,412	-0,430	-0,414
	(0,380)	(0,381)	(0,380)
SIZE_RANK3	-0,424	-0,461	-0,430
	(0,426)	(0,425)	(0,427)
INVITATIONS	-0,063	-0,065	-0,064
	(0,055)	(0,055)	(0,055)
Konstante	1,831***	1,786***	1,840***
	(0,540)	(0,532)	(0,541)
Nullaussagen	Nein	Nein	Nein
Beobachtungen	267	267	267
Pseudo R^2	0,253	0,252	0,254

Anmerkungen: Die Tabelle zeigt die sich aus Probit-Schätzungen ergebenden Regressionskoeffizienten der Regressionsgleichung für das Referenzmodell unter Anwendung diverser Robustheitstests. Referenzgruppen: *BUSINESS*, *Neutral* und *SIZE_RANK1*. Abhängige Variable: *POS*. Die zugehörigen Standardabweichungen sind jeweils darunter in Klammern. *, ** und *** repräsentieren jeweils Signifikanzniveaus von 10%, 5% und 1%. Alle Variablen sind in Anhang 2 definiert.

lichen Level. Die Gruppenvariation aus dem Referenzmodell kann demnach beibehalten werden.

Zuletzt werden in (6) bis (8) Interaktionsterme auf ihren Einfluss auf das Referenzmodell überprüft. Die Ergebnisse in Kapitel 5.1 legen nahe, dass die präferierte Änderungsrichtung von der betrachteten Expertengruppe abhängt. Die Kontrolle von zwei exemplarischen Gruppen mit der jeweils mehrheitlich verwendeten Änderungsrichtung bietet somit relevante Erkenntnisse. In diesem Fall werden die Steuerberater mit vorwiegend entlastenden Änderungspräferenzen sowie die Zivilgesellschaft mit überwiegend verschärfenden Änderungswünschen hinzugezogen. Im Folgenden werden beide Interaktionsterme zunächst einzeln (6 + 7) und danach zusammen mit dem Referenzmodell verbunden. Die übrigen Gruppen können im Modell nicht mit ihrer mehrheitlichen Änderungspräferenz kontrolliert werden, da sie als Referenzgruppe dienen (Unternehmen) oder mehrheitlich neutrale Aussagen getätigt haben, die ebenfalls als Referenzänderungsrichtung im Modell berücksichtigt sind.

Der Einbezug von Referenzvariablen führt in dieser Untersuchung stets zu Multikollinearitätsproblemen, weswegen solche Konstellationen nicht weiterverfolgt werden. Die übrigen Variationen suggerieren in allen Fällen mit ihren Regressionskoeffizienten eine erhöhte Erfolgswahrscheinlichkeit für Steuerberater bzw. die Zivilgesellschaft, wenn diese entlastende bzw. verschärfende Änderungen fordern, wobei die Steuerberater jeweils erfolgreicher einzuschätzen sind. Doch wie in den übrigen Konstellationen sind diese Ergebnisse ebenfalls nicht signifikant, sodass im Referenzmodell richtigerweise auf Interaktionsterme verzichtet wird.

Unter Berücksichtigung sämtlicher Robustheitstests wird

die Nullhypothese weiterhin beibehalten. Für keine der betrachteten Expertengruppen konnte ein signifikanter Unterschied in ihrer Wahrscheinlichkeit für einen Lobbyerfolg festgestellt werden. Die präferierte Änderungsrichtung bleibt auch hier unabhängig der betrachteten Expertengruppe das wesentlich bedeutsamere Kriterium für die Determinierung von Lobbyerfolg in der Steuergesetzgebung des deutschen Parlaments. Präferierte Entlastungen bzw. Verschärfungen sind dabei nach wie vor weniger erfolgreich als neutral einzuordnende Aussagen.

5.4 Limitationen

Jegliche Ergebnisse innerhalb dieser Ausarbeitung sind unter Berücksichtigung diverser Limitationen zu betrachten. Zunächst ist in diesem Zuge die Problematik hinsichtlich der fehlenden Kausalitätsbegründung bei der Betrachtung von Korrelationen zu nennen. Auch wenn die gewählte Methodik eine Differenzierung nach Expertengruppen bei der Betrachtung von Lobbyerfolg erlaubt, legitimiert eine vorhandene Korrelation nicht die Feststellung eines begründeten Zusammenhangs zwischen den beiden Aspekten (vgl. Elschner und Hardeck, 2022, S. 333) bzw. keines Zusammenhangs bei fehlender Korrelation. Sie geben nur erste Anhaltspunkte zur Beurteilung eines großen wissenschaftlichen Phänomens. Dabei sind auch Endogenitätsprobleme in Lobbyuntersuchungen generell nicht auszuschließen, da die Entscheidung zur Einladung und Beiwohnung der Anhörung nicht zufällig geschieht, sondern von den Abgeordneten und den Sachverständigen selbst abhängt (vgl. de Figueiredo und Richter, 2014, S. 170; Elschner und Hardeck, 2022, S. 333). Daneben könnten die erwarteten Aussagen von konkurrierenden Gruppen die Aussagen der jeweils anderen Gruppe ex-ante beeinflussen und somit das eigentliche Ergebnis verzerren (anlehnend an de Figueiredo und Richter, 2014, S. 170). In diesem Zuge wird der Aufruf von Barrick und Brown (2019, S. 77) betont, die bei dem Planungsprozess der Anhörungen und dem Selektionsverfahren der Sachverständigen insbesondere mit Bezug zu Besteuerungsaspekten noch großen Forschungsbedarf sehen.

Die bedeutsamste Limitation bleibt jedoch die Kodierung der Datenpunkte durch nur einen einzigen Kodierer, was durch die Prüfungsbedingungen dieser Masterarbeit begründet wird. Trotz bestmöglicher Anstrengung zur objektiven Beurteilung der Sachverhalte kann bei einem Einzelnen eine gewisse Restsubjektivität bei einer strengen Binärbetrachtung von uneindeutigen Sachverhalten nicht außer Acht gelassen werden. Um diesem Umstand ein Stück weit entgegenzuwirken, wurde in einem Robustheitstest eine mögliche Kodierungsbandbreite auf Unterschiede zum Referenzmodell geprüft. Eine angemessenere Herangehensweise in ähnlichen wissenschaftlichen Papieren sieht regelmäßig mindestens zwei voneinander unabhängige Kodierer vor, die jeweils für sich die Sachverhalte nach den vorher aufgestellten Kodierungsregeln kodieren und ihre Ergebnisse miteinander abgleichen. Zur anschließenden Beurteilung der Intercoder-Reliabilität wird etwa Krippendorffs Alpha herangezogen (Krippendorff, 1980, S. 221f), welches sich nahe 1 befinden sollte, um hohe bzw. ausreichende Übereinstimmung und damit objektive Kodierungen zu signalisieren. Etwaige Kodierungskonflikte werden zum Schluss von einer dritten Person final gelöst, um eine möglichst unbefangene Beurteilung zu erreichen. Darüber hinaus kann bei der Selektion relevanter Aspekte, an denen die Beurteilung des Lobbyerfolgs anknüpft, ebenfalls eine gewisse Restsubjektivität nicht ausgeschlossen werden. Trotz des objektiven Auswahlkriteriums der Nennungshäufigkeit der Aspekte kann nicht verhindert werden, dass manche Aspekte für bestimmte Gruppen relevanter sind als die ausgewählten Aspekte. Daran anknüpfend kann die Begrenzung auf fünf Aspekte pro Gesetzesvorhaben auch keine Rücksicht auf den Grad an Relevanz nehmen: So kann beispielsweise Aspekt 1 gesamtgesellschaftlich oder für den einzelnen Autor selbst wesentlich (un-)bedeutsamer sein als Aspekt 2, wodurch eine gelungene Einflussnahme beim ersten Aspekt als ein größerer (kleinerer) Lobbyerfolg anzusehen wäre als ein Erfolg bei Aspekt 2. Demnach könnten Gewichtungen von Lobbyerfolgen Perspektiven für weitergehende Forschung sein.

Des Weiteren sind die von den Autoren genannten Aspekte abhängig von dem bisherigen Verlauf des einzelnen Gesetzgebungsverfahrens und damit verbundenen Einflussnahmen, die in dieser Betrachtung nicht berücksichtigt werden können. Über verschiedene Lobbykanäle können die Interessengruppen auch auf einer informellen Ebene mit den politischen Entscheidungsträgern kommunizieren und sich im Voraus miteinander austauschen (vgl. Hoffmann und Zülch, 2014, S. 712). Solche Kontakte sind zwar politisch gewollt und für beide Seiten von nützlicher Natur, doch durch die fehlende Einbeziehung der Öffentlichkeit sehr intransparent. Ein hier stattfindender Austausch über weitere relevante Aspekte lenkt den Gesetzgebungsprozess möglicherweise bereits vorher in bestimmte Richtungen, die von den Sachverständigen im späteren Verlauf der Expertenanhörung dann nicht mehr angesprochen werden müssen und innerhalb dieser Untersuchung somit als nicht relevant erachtet werden. Die Intransparenz von Lobbyaktivitäten bleibt somit eine Schwierigkeit für die Forschung über politische Einflussnahme auch in der deutschen Steuergesetzgebung.

Darüber hinaus ist der Legislativprozess vor allem in Themengebieten der internationalen Besteuerung durch vorhergehende intergouvernementale Konsultationen innerhalb der OECD bzw. der EU geprägt. Innerhalb dieser Organisationen fand und findet der Großteil der Erarbeitung des künftigen internationalen Steuersystems statt, wodurch häufig umfangreiche Änderungen bereits feststehen, bevor das jeweilige Vorhaben in den nationalen Gesetzgebungsprozess eingebracht wird. Dadurch verschieben sich auch Lobbyaktivitäten auf eine solche internationale Ebene, die damit mittelbar auch die deutsche Umsetzung tangiert. Die Umsetzung selbst hat dann in einem festgelegten Rahmen zu erfolgen, sodass der nationale Handlungsspielraum im Parlament sowie die Möglichkeiten für inländische Interessengruppen bei diesen Gesetzen zum Teil sehr begrenzt sind.

In der untersuchten Stichprobe haben beispielsweise zehn von elf Gesetzgebungsverfahren einen Bezug zu einer umzusetzenden EU-Richtlinie, Anweisungen der Europäischen Kommission oder einem Urteil des Europäischen Gerichtshofs, wodurch der Großteil der relevanten Aspekte in das obige Szenario einzuordnen ist. Ein angedachter Split der Daten zur Prüfung der Aspekte, die nicht durch europäische Einflüsse tangiert werden, war aufgrund einer zu geringen Anzahl an Datenpunkten statistisch nicht möglich. Diese Limitation beeinflusst damit auch den möglichen Lobbyerfolg der Sachverständigen, die Änderungen außerhalb des international festgesetzten Rahmens gar nicht erwirken können. Eine ähnliche Untersuchung zu rein nationaler Steuergesetzgebung könnte somit für die zukünftige Forschung interessant sein, um weitergehende Erkenntnisse zu inländischer Interessenvertretung hinsichtlich der Differenzierung von Expertengruppen bei der Steuergesetzgebung zu erhalten. Auch wird damit gezeigt, dass Untersuchungen hinsichtlich politischer Einflussnahme auf europäischer Ebene ein wichtiges Forschungsfeld darstellen.

6 Fazit

Lobbybemühungen in der Steuergesetzgebung sind und bleiben ein dynamisches Tätigkeitsfeld auf deutschem sowie internationalem Parkett. Die Literatur sieht sowohl im Allgemeinen als auch bei Gesetzgebungsprozessen im Speziellen verschiedene Anknüpfungspunkte und Herangehensweisen für erfolgreiche politische Einflussnahme unterschiedlicher Interessengruppen. Als eines der bedeutsamsten Lobbygebiete bleibt die Steuergesetzgebung ein vielbeachtetes Feld für unterschiedliche Stakeholder, die bei einer erfolgreichen Einflussnahme mittelbare Auswirkungen auf die öffentlichen und ihre eigenen Finanzen auslösen. In einer Vielzahl von Publikationen konnte der individuelle Erfolg an der Zugehörigkeit des Lobbyisten zu einer bestimmten Interessengruppe festgemacht werden, wobei in der Gesamtbetrachtung jedoch kein klarer Favorit determiniert werden kann. Sowohl der größere Teil an Ressourcen bei den Unternehmen als auch die unabhängige Expertise von Berufsträgern und Angehörigen der Wissenschaft spielen zentrale Rollen bei der größeren Lobbyerfolgswahrscheinlichkeit der jeweiligen Gruppe in den zugrundeliegenden Untersuchungen. Allerdings ist bisher unklar geblieben, ob diese Erkenntnisse auch auf den parlamentarischen Kontext hinsichtlich der Steuergesetzgebung in Deutschland zutreffen.

Diese Arbeit hat mit der durchgeführten Untersuchung versucht, zur Schließung der vorhandenen Forschungslücke beizutragen. Die Analyse setzt dabei an den Expertenanhörungen an, die die Mitglieder des Finanzausschusses im Bundestag zu komplexen Gesetzesvorhaben einberufen können. Die Stichprobe wurde auf den Zeitraum der jetzigen und den zwei vorangegangenen Legislaturperioden sowie auf Gesetzesvorhaben mit Bedeutung für die internationale Besteuerung beschränkt. Die hierbei eingereichten Stellungnahmen unterschiedlicher Sachverständiger bildeten den Ausgangspunkt für die Untersuchung, inwiefern ein möglicher Lobbyerfolg mit der korrespondierenden Expertengruppe zusammenhängt. Qualitative Analysen suggerieren zumindest innerhalb der Stichprobe einen größeren Lobbyerfolg für Angehörige der Zivilgesellschaft und den geringsten Erfolg bei Unternehmen. Dies würde bisherigen Erkenntnissen der Literatur nicht entsprechen, welche die Zivilgesellschaft regelmäßig im Nachteil zu den übrigen Gruppen eingeordnet hat. Allerdings wurden mithilfe von Probit-Regressionen quantitative Koeffizienten geschätzt, die zwar Differenzen zwischen den betrachteten Gruppen signalisieren, aber für keine Gruppe einen signifikanten Unterschied feststellen lassen. Nur bei der Vernachlässigung der präferierten Änderungsrichtung der Sachverständigen sind signifikant höhere Erfolgschancen für Angehörige der Autoritätsgruppe und zum Teil auch für die Zivilgesellschaft erkennbar. Angesichts der Dominanz der Änderungsrichtung als Kontrollvariable wäre ein solcher Ausschluss jedoch ökonomisch und ökonometrisch nicht sinnvoll. Vielmehr sollten die geringeren Erfolgswahrscheinlichkeiten hervorgehoben werden, die entlastende und verschärfende Änderungspräferenzen im Vergleich zu neutralen Positionierungen aufweisen. Auch diverse Robustheitstests haben keine Auswirkungen auf die genannten Erkenntnisse. Dieses Ergebnis bestätigt im weiteren Sinne die Ergebnisse von Mahoney (2007, S. 52), die bei Lobbyisten mit Änderungswünschen gegenüber ihren Kollegen, die keine Änderung präferieren, geringere Erfolgschancen feststellt.

Diese Erkenntnis bestätigt den primären Zweck solcher Expertenanhörungen: Die Einholung fachlicher Expertise zur besseren Einordnung komplexer Sachverhalte für die Abgeordneten. Der Raum zur Bekanntgabe individueller Präferenzen durch die Interessengruppen ist lediglich ein sekundäres Ziel solcher Anhörungen. Außerdem würden ungleiche Berücksichtigungen unterschiedlicher Gruppen Fragen der demokratischen Legitimation des gesamten Prozesses aufwerfen, die nun zumindest im Kontext dieser Untersuchung nicht betrachtet werden müssen. Die Forschungsfrage kann somit in die Richtung beantwortet werden, dass Parlamentarier des Deutschen Bundestages zumindest in diesem Kontext jeder Gruppe gleichermaßen folgen bzw. keine Gruppe bevorzugt behandeln. Betrachtet werden müssen jedoch vielmehr die mit dieser Untersuchung einhergehenden Limitationen: Vorangegangene intransparente Einflussnahmen, einschränkende Handlungsmöglichkeiten durch europäische Richtlinien und Gerichtsurteile sowie ein verbleibendes Restsubjektivitätsproblem durch die Modalitäten dieser Arbeit schränken die Aussagekraft der Ergebnisse ein.

Gleichzeitig eröffnen diese auch Möglichkeiten für weitergehende Forschung in diesem Feld. Unter Zuhilfenahme eines oder mehrerer zusätzlicher Autoren und der Analyse eines nationalen Besteuerungsrahmens ohne intergouvernementale Einflüsse können weitergehende und qualitativ bessere Erkenntnisse über die politische Einflussnahme im deutschen Parlament gewonnen werden. Auch die Untersuchung der Entstehung von Besteuerungsregelungen auf der europäischen Ebene kann ein weiterer Anknüpfungspunkt für fortgeschrittene Forschungsvorhaben mit der hier angewandten Methodik sein.

Literatur

Allen, A., Ramanna, K., & Roychowdhury, S. (2018). Auditor Lobbying on Accounting Standards. *Journal of Law, Finance, and Accounting*, *3*(2), 291–331.

Aplin, J. C., & Hegarty, W. H. (1980). Political Influence: Strategies Employed by Organizations to Impact Legislation in Business and Economic Matters. *Academy of Management Journal*, *23*(3), 438–450.

Bamber, M., & McMeeking, K. (2016). An Examination of International Accounting Standard-Setting Due Process and the Implications for Legitimacy. *British Accounting Review*, *48*(1), 59–73.

Barrick, J. A., & Alexander, R. M. (2014). Tax Lobbying and Corporate Political Activity: How Do Firms Seek Tax Relief? [Working Paper]. https://doi.org/10.2139/ssrn.2500959

Barrick, J. A., & Brown, J. L. (2019). Tax-Related Corporate Political Activity Research: A Literature Review. *Journal of the American Taxation Association*, *41*(1), 59–89.

Baudot, L., & Walton, P. (2014). Influences on the Standard-Setting and Regulatory Process. In C. van Mourik & P. Walton (Hrsg.), *The Routledge Companion to Accounting, Reporting and Regulation* (1. Aufl., S. 318–338). Routledge.

Bertrand, M., Bombardini, M., & Trebbi, F. (2014). Is It Whom You Know or What You Know? An Empirical Assessment of the Lobbying Process. *American Economic Review*, *104*(12), 3885–3920.

Bisogno, M., Santis, S., & Tommasetti, A. (2015). Public-Sector Consolidated Financial Statements: An Analysis of the Comment Letters on IPSASB's Exposure Draft No. 49. *International Journal of Public Administration*, *38*(4), 311–324.

Blum, P. (2004). *Wege zu besserer Gesetzgebung - sachverständige Beratung, Begründung, Folgeabschätzung und Wirkungskontrolle, Gutachten I zum 65. Deutschen Juristentag Bonn 2004*. C.H. Beck.

Brasher, H. (2006). Listening to Hearings, Legislative Hearings and Legislative Outcomes. *American Politics Research*, *34*(5), 583–604.

Bräuninger, T., Debus, M., & Wüst, F. (2017). Governments, Parliaments and Legislative Activity. *Political Science Research and Methods*, *5*(3), 529–554.

Checkoway, B. (1981). The Politics of Public Hearings. *Journal of Applied Behavioral Science*, *17*(4), 566–582.

Cross, J. P., Eising, R., Hermansson, H., & Spohr, F. (2021). Business Interests, Public Interests, and Experts in Parliamentary Committees: Their Impact on Legislative Amendments in the German Bundestag. *West European Politics*, *44*(2), 354–377.

de Figueiredo, J. M., & Richter, B. K. (2014). Advancing the Empirical Research on Lobbying. *Annual Review of Political Science*, *17*, 163–185.

Eising, R., & Spohr, F. (2017). The More, the Merrier? Interest Groups and Legislative Change in the Public Hearings of the German Parliamentary Committees. *German Politics*, *26*(2), 314–333.

Elschner, C., & Hardeck, I. (2022). Assessing the Influence of Different Interest Groups on International Tax Policy: Evidence from the BEPS Project. *Contemporary Accounting Research*, *39*(1), 304–338.

Faccio, M. (2010). Differences Between Politically Connected and Nonconnected Firms: A Cross-Country Analysis. *Financial Management*, *39*(3), 905–928.

Furlong, S. R., & Kerwin, C. M. (2005). Interest Group Participation in Rule Making: A Decade of Change. *Journal of Public Administration Research and Theory*, *15*(3), 353–370.

Georgiou, G. (2004). Corporate Lobbying on Accounting Standards: Methods, Timing and Perceived Effectiveness. *Abacus*, *40*(2), 219–237.

Gerig, M. (2014). Der rechtliche Rahmen für Lobbyisten. *Zeitschrift für Rechtspolitik*, *47*(8), 247–249.

Grossman, G. M., & Helpman, E. (2001). *Special Interest Politics*. MIT Press.

Hall, R. L., & Deardorff, A. V. (2006). Lobbying as Legislative Subsidy. *American Political Science Review*, *100*(1), 69–84.

Hillman, A. J., & Hitt, M. A. (1999). Corporate Political Strategy Formulation: A Model of Approach, Participation, and Strategy Decisions. *Academy of Management Review*, *24*(4), 825–842.

Hoffmann, S., & Zülch, H. (2014). Lobbying on Accounting Standard Setting in the Parliamentary Environment of Germany. *Critical Perspectives on Accounting*, *25*(8), 709–723.

Jorissen, A., Lybaert, N., Orens, R., & van der Tas, L. (2012). Formal Participation in the IASB's Due Process of Standard Setting: A Multi-Issue/Multi-Period Analysis. *European Accounting Review*, *21*(4), 693–729.

Keim, G. D., & Zeithaml, C. P. (1986). Corporate Political Strategy and Legislative Decision Making: A Review and Contingency Approach. *Academy of Management Review*, *11*(4), 828–843.

Kim, C. F., & Zhang, L. (2016). Corporate Political Connections and Tax Aggressiveness. *Contemporary Accounting Research*, *33*(1), 78–114.

Kloepfer, M. (2005). § 42 Öffentliche Meinung, Massenmedien, B. Öffentliche Meinung und politische Willensbildung des Volkes [12 Bände]. In J. Isensee & P. Kirchhof (Hrsg.), *Handbuch des Staatsrechts. Band III: Demokratie - Bundesorgane* (3. Aufl., S. 395–406, Bd. 3). C.F. Müller.

Knauff, M. (2016). Strukturfragen des neuen Vergaberechts. *Neue Zeitschrift für Baurecht und Vergaberecht*, *16*(4), 195–198.

Krippendorff, K. (1980). *Content Analysis - An Introduction to Its Methodology*. Sage.

Kropp, C., & Wagner, J. (2010). Knowledge on Stage: Scientific Policy Advice. *Science, Technology, & Human Values*, *35*(6), 812–838.

Kurbjuweit, D., & Medick, V. (2022). "Im Übrigen gilt das strucksche Gesetz" [10.12.2022]. *DER SPIEGEL*, *2022*(50), 28–31.

Kwok, W. C., & Sharp, D. (2005). Power and International Accounting Standard Setting: Evidence from Segment Reporting and Intangible Assets Projects. *Accounting, Auditing & Accountability Journal*, *18*(1), 74–99.

Lando, T. (2003). The Public Hearing Process: A Tool for Citizen Participation, or a Path Toward Citizen Alienation? *National Civic Review*, *92*(1), 73–82.

Long, J. S., & Freese, J. (2001). *Regression Models for Categorial Dependent Variables Using Stata* (1. Aufl.). Stata Press.

Mahoney, C. (2007). Lobbying Success in the United States and the European Union. *Journal of Public Policy*, *27*(1), 35–56.

McLeay, S., Ordelheide, D., & Young, S. (2000). Constituent Lobbying and Its Impact on the Development of Financial Reporting Regulations: Evidence from Germany. *Accounting, Organizations and Society*, *25*(1), 79–98.

Mellado, L., & Parte, L. (2021). Insights from Lobbying Research on the Accounting Standard-Setting Process Through Comment Letter Submissions. In M. H. Bilgin & H. Danis (Hrsg.), *Eurasian Studies in Business and Economics* (S. 3–22, Bd. 16/1). Springer.

Mulligan, E., & Oats, L. (2016). Tax Professionals at Work in Silicon Valley. *Accounting, Organizations and Society*, *52*, 63–76.

OECD. (2008). Study Into the Role of Tax Intermediaries. https://www.oecd.org/tax/administration/39882938.pdf

OECD. (2015). Explanatory Statement, OECD/G20 Base Erosion and Profit Shifting Project. https://www.oecd.org/tax/beps-explanatory-statement-2015.pdf

Olson, M. (1965). *The Logic of Collective Action, Public Goods and the Theory of Groups*. Cambridge University Press.

Parsons, T. (1963). On the Concept of Influence. *Public Opinion Quarterly*, *27*(1), 37–62.

Perry, J., & Nöelke, A. (2005). International Accounting Standard Setting: A Network Approach [Artikel 5]. *Business and Politics*, *7*(3), 1–32.

Puro, M. (1984). Audit Firm Lobbying Before the Financial Accounting Standards Board: An Empirical Study. *Journal of Accounting Research*, *22*(2), 624–646.

Radcliffe, V. S., Spence, C., Stein, M., & Wilkinson, B. (2018). Professional Repositioning During Times of Institutional Change: The Case of Tax Practitioners and Changing Moral Boundaries. *Accounting, Organizations and Society*, *66*, 45–59.

Richter, B. K., Samphantharak, K., & Timmons, J. F. (2009). Lobbying and Taxes. *American Journal of Political Science*, *53*(4), 893–909.

Saemann, G. (1999). An Examination of Comment Letters Filed in the U.S. Financial Accounting Standard-Setting Process by Institutional Interest Groups. *Abacus*, *35*(1), 1–28.

Schnakenberg, K. E. (2017). Informational Lobbying and Legislative Voting. *American Journal of Political Science*, *61*(1), 129–145.

Schneider, G., & Baltz, K. (2003). The Power of Specialization: How Interest Groups Influence EU Legislation. *Rivista di Politica Economica*, *93*(1), 253–288.

Siefken, S. T. (2021). No Paradise of Policy-Making, The Role of Parliamentary Committees in the German Bundestag. In S. T. Siefken & Rommetvedt (Hrsg.), *Parliamentary Committees in the Policy Process* (1. Aufl., S. 116–136). Routledge.

Siefken, S. T., & Schüttemeyer, S. S. (2013). The German Bundestag and External Expertise: Policy Orientation as Counterweight to Deparliamentarisation? In S. Blum & K. Schubert (Hrsg.), *Policy Analysis in Germany* (1. Aufl., S. 161–180). Policy Press.

Smith, R. A. (1995). Interest Group Influence in the U.S. Congress. *Legislative Studies Quarterly*, *20*(1), 89–139.

Sutton, T. G. (1984). Lobbying of Accounting Standard-Setting Bodies in the U.K. and the U.S.A.: A Downsian Analysis. *Accounting, Organizations and Society*, *9*(1), 81–95.

Tandy, P. R., & Wilburn, N. L. (1996). The Academic Community's Participation in Standard Setting: Submission of Comment Letters on SFAS Nos. 1-117. *Accounting Horizons*, *10*(3), 92–111.

van de Vijer, A. (2022). Morality of Lobbying for Tax Benefits: A Kantian Perspective. *Journal of Business Ethics*, *181*(1), 57–68.

Watts, R. L., & Zimmerman, J. L. (1978). Towards a Positive Theory of the Determination of Accounting Standards. *The Accounting Review*, *53*(1), 112–134.

Yackee, J. W., & Yackee, S. W. (2006). A Bias Towards Business? Assessing Interest Group Influence on the U.S. Bureaucracy. *Journal of Politics*, *68*(1), 128–139.

Yen, A. C., Hirst, D. E., & Hopkins, P. E. (2007). A Content Analysis of the Comprehensive Income Exposure Draft Comment Letters. *Research in Accounting Regulation*, *19*, 53–79.