TECHNISCHE UNIVERSITÄT MÜNCHEN
TUM School of Management

Empirical Insights on Executive Compensation, Human Capital Creation, and Accounting Misconduct

Matthias Christopher Regier

Vollständiger Abdruck der von der TUM School of Management der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.) genehmigten Dissertation.

Vorsitz: Prof. Dr. Christoph Kaserer
Prüfer: 1. Prof. Dr. Gunther Friedl
2. Prof. Dr. Jürgen Ernstberger

Die Dissertation wurde am 16.12.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Management am 15.03.2023 angenommen.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor Gunther Friedl for his encouragement, guidance, and challenging feedback. I enjoyed much leeway to pursue intriguing and innovative research endeavors, change directions, mobilize resources, and independently design this journey. He creates an environment where visionary ideas and critical thoughts regarding all academic matters can be openly discussed, which nourishes personal as well as organizational development.

I further thank my colleagues and doctoral fellows, in particular, Daniel Beck, Tobias Beibl, Carl-Philipp Beichert, Maximilian Blaschke, Markus Brunner, Markus Buchner, Andrea Doetsch, Björn-Eric Förster, Markus Frank, Yanis Gamarra, Gunther Glenk, Christian Göttsch, Christoph Heilmann, Antonia Hoffmann, Julia Holzmann, Anton Kelnhofer, Lukas Kemmer, Alexandra Knoth, Verena Limbrunner, Julian Ludat, David Matthäus, Marc Mehrer, Lambert Neumayr, Fridolin Pflugmann, Moritz Rombach, Peter Schäfer, Magnus Schauf, Lukas Schloter, Eline Schoonjans, Alexander Schult, Leonard Schwellnus, Verena Springer, Sarah Steinbach, Christian Stoll, Magnus Strobel, Andrea Stübner, Benedikt Tratt, Johannes Trenkle, Johannes Viehhauser, Friedrich Walcher, Dominik Weise, Marcus Witter, and others, for mutual inspiration, inclusive collaboration, academic and personal feedback, patience and indulgence, as well as energy and fun.

My dissertation greatly benefited from constructive discussion and rigorous review by members of my committee Christoph Kaserer and Jürgen Ernstberger, coauthors Ethan Rouen, Dan Amiram, Zahn Bozanic, and Johannes Roscher, as well as countless fellow researchers whom I met at doctoral seminars, conferences, or during my time as visiting researcher at Harvard Business School.

My family and friends deserve a special thank you for their enduring support throughout the years. I am indebted to my wife Hannah Lisiecki for sticking with me during the ups and downs throughout the years. Thank you for your endless love and friendship.

Finally, I am grateful for (financial) support of the TUM School of Management, the TUM Graduate School, the TUM Fördergesellschaft Controlling e.V., and the Research Excellence Program.

Abstract

Agency and corporate investment, economics of regulation, and various stakeholders' interests in firms functioning well, are important aspects of firms' corporate governance. This dissertation addresses three research questions related to these aspects. The first asks whether regulation-induced lengthening of the duration in executive compensation improves investment efficiency. I find that firms affected by the regulation reduce their abnormal investment relative to unaffected firms. The treatment effect is economically significant, as the reduction in abnormal investment amounts to about 10% of mean investment. It appears that the effects are greatest in firms that had a low degree of compensation committee independence prior to the regulation. These results are informative for many stakeholders that are affected by executives' incentives and their investment decisions. The second research question asks whether and when market participants realize firms' investments into human capital. Analysts and shareholders seem to capture firm-year-specific investments into human capital that are implicit in their personnel expenditures with delay. Most notably, long-short portfolios based on a human capital investment proxy produce annualized value-weighted or equal-weighted abnormal returns of between 3.5% and 7.8%. This may provide regulators and other stakeholders with arguments to reconsider the accounting for input resource expenditures that create intangible capital. The third research question asks whether U.S.-trained financial accounting misconduct prediction models can be applied *far*-out-of-sample. Established models trained with U.S. misconduct cases are highly predictive for a European sample containing 21 misconduct cases relating to 59 misconducted firm-years. Straightforward Logistic Regression models as well as sophisticated RUSBoost Ensemble Learning would have assigned high probabilities of misconduct to a large portion of the cases in years where the deceitful schemes were likely implemented. These results have implications for various stakeholders, whereby the strongest implication is that non-U.S. enforcement institutions learn how a preselection of the overseen firms prior to engaging in more rigorous examinations could be designed.

Abstract in German

Unternehmerische Investition unter Agency, Regulierungsökonomik, und die Interessen verschiedener Stakeholder am Funktionieren der Firma sind wichtige Aspekte von Corporate Governance. Diese Dissertation adressiert drei Forschungsfragen zu diesen Aspekten. Die erste testet, ob durch Regulierung induzierte Verlängerung der Fristigkeiten in der Vorstandsvergütung zu effizienteren Investitionen führt. Ich zeige, dass von der Regulierung betroffene Firmen ihre abnormalen Investitionen relativ zu nicht betroffenen Firmen reduzieren. Dieser Treatment Effekt ist ökonomisch signifikant, da die Reduktion in den abnormalen Investitionen etwa 10% der mittleren Investitionen entspricht. Der Effekt scheint am stärksten in Firmen zu wirken, die vor der Regulierung ein wenig unabhängiges Vergütungskomittee hatten. Diese Ergebnisse sind relevant für viele Stakeholder die von den Anreizen für Vorstände und deren Investitionsentscheidungen betroffen sind. Die zweite Forschungsfrage befasst sich damit, ob und wann Marktteilnehmer realisieren, wenn Firmen in Humankapital investieren. Analysten und Anteilseigner scheinen die in Personalausgaben impliziten Firmenjahr-spezifischen Investitionen in Humankapital nur mit Verzögerung zu realisieren. Besonders Long-Short Portfolios basierend auf Schätzungen der Humankapitalinvestitionen sind mit annualisierten wert- oder gleichgewichteten abnormalen Renditen von zwischen 3,5% und 7,8% assoziiert. Das wappnet Regulierer und andere Interessengruppen mit Argumenten, die Berichterstattung für die Ausgaben für Inputressourcen die immaterielles Kapital generieren zu überdenken. Die dritte Forschungsfrage untersucht, ob mit U.S. Daten trainierte Modelle für die Vorhersage von Bilanzmanipulation fernab out-of-sample angewandt werden können. Etablierte Modelle trainiert mit historischen U.S. Fällen haben einen hohen Vorhersagegehalt für 21 Europäische Fälle mit 59 betroffenen Firmenjahren. Unkomplizierte Logistische Regressionsmodelle sowie fortschrittliches RUSBoost Ensemble Learning hätte für einen großen Teil der Fälle hohe Wahrscheinlichkeiten für Fehlverhalten in den (nachträglich vermuteten) Jahren der Implementierung des Betrugsschemas vorhergesagt. Diese Ergebnisse haben Implikationen für diverse Stakeholder, insbesondere Durchsetzungs- und Vollstreckungsinstitutionen außerhalb der USA, die lernen, wie eine Vorauswahl der beaufsichtigten Firmen vor weitergehenden Überprüfungen gestaltet werden kann.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation for and juxtaposition of the three essays included in the dissertation

This dissertation provides empirical insights into three different aspects of today's accounting scholarship. These aspects are 1) the effects of regulation of executive compensation on firms' investment behavior, 2) capital markets' perception of firms' human capital creation, and 3) prediction of financial accounting misconduct. I briefly motivate each research question in the following.[1] [2]

This dissertation's first essay (i.e., **Essay 1** in Chapter 2) empirically connects mandated incentive duration with firms' investment decisions. Beyond accounting research, compensation design is one of the corporate governance tools that draws the most attention and creates the most controversy about corporations among the public (e.g., Edmans et al., 2017b). While public outcry regarding executive compensation often focuses on how much is paid, the possibly even bigger problem is how the compensation is paid (Jensen and Murphy, 1990). One important aspect of *how* to pay is *when* to pay, meaning that the duration of the incentive system is a critical aspect (e.g., Bebchuk and Fried, 2010; Gopalan et al., 2014; Ladika and Sautner, 2020). The research question is whether mandated (i.e., exogenously imposed) longer incentive duration alters the efficiency of firms' investments.[3]

Asking such a research question builds on a couple of implicit theoretical underpinnings. A central one is that executives are better informed about their firms' investment opportunities than shareholders are and may even use their informational advantage to pursue private benefits (e.g., Berle and Means, 1932; Jensen and Meckling, 1976). More-

[1] The essays included in my dissertation rely on individual literature backgrounds and therefore build, at times, on different implicit conceptual presumptions that I also outline on a high level in the following.

[2] I refer to the three research questions addressed in this dissertation as *essays* in this overall introduction as well as in the final conclusion. Within each essay's chapter, I keep them in their working paper/published paper style referring to themselves as *papers*.

[3] Investment efficiency is a key determinant of a firm's economic productivity (Biddle and Hilary, 2006). It means that firms undertake only projects and all projects with a positive net present value (NPV) (Modigliani and Miller, 1958). Inefficient investment may relate to over-investment, which corresponds to investing in negative NPV projects, and under-investment, which corresponds to passing up positive NPV opportunities (Biddle et al., 2009).

over, the executive compensation system – contracted between executives and shareholders of the firm (or their representatives) to act as a governance mechanism – need not necessarily arrive at the optimal design (e.g., Bebchuk and Fried, 2010).[4] This may result in non-optimal incentives and investment decisions. Testing a regulation's effect on the investment decisions further implies that there is a common understanding of how to measure corporate investment and that it is possible to capture deviations from the optimal investment level in empirical analyses (e.g., Biddle and Hilary, 2006).[5]

While also analyzing firms' investments, **Essay 2** departs from established measures of corporate investment used in the first essay and tries to further our understanding of investment in the modern intangible firm. It starts with acknowledging that the possibly most important kind of investment of this modern firm – namely investments in human capital – is difficult to understand, make, and measure (e.g., Zingales, 2000). Firms differ in how well they invest in personnel in ways that provide future benefits (e.g., Flamholtz, 1971). However, human capital related assets usually do not show up on the balance sheets, instead, the expenditures must be expensed as incurred. One of the most important complications might be that firms do not own their employees, so when firms invest into them in a way that creates intangible assets with future benefits, there is a risk that they might simply leave (e.g., Blair, 2003). The resulting paucity of disclosures related to human capital in current accounting systems may create an information gap that distorts valuation of the firm (Zingales, 2000).

We therefore develop a novel attempt at measuring the deviation in firms human capital creation across firms and years.[6] We then estimate how long it takes market participants to impound these differences into stock prices. We thereby ask whether markets timely grasp firms' investments in human capital. If not, there could be grounds

[4]In broad terms, one may argue that executives generally prefer to receive compensation earlier rather than later. However, it is widely common to implement some kind of longer term (e.g., multi-year) compensation plan to provide incentives for long-term value creation that may be difficult to measure and reward in the short-term.

[5]This research question assumes that frictions regarding corporate investments cancel each other out on some aggregate level, so that investment efficiency as such and changes therein become measurable on firm-year-level. For instance, researchers may assume the aggregate investment of all firms in an industry to be efficient, which allows to estimate positive and negative deviations on the level of the firm.

[6]Essay 2 presented in the third chapter of this dissertation is based on a joint research project with Ethan Rouen. I refer to the two of us when using plural pronouns in the context of this essay.

for a need to reconsider established reporting frameworks (e.g., Leuz and Wysocki, 2016) and make them more suitable for the intangible economy. This research question thereby implicitly assumes that markets behave efficiently in determining risks and returns in some aggregate.

The first two essays mainly provide insights for corporate governance actors like executives, supervisory board members, and shareholders, with the regulator as a secondary addressee. While also informing different market participants and corporate stakeholders, **Essay 3** primarily speaks to regulators. Moreover, the first two essays deal with different aspects of corporate investment and take the reported figures for granted while disregarding that such reporting may be misconducted. The third essay, on the contrary, is not concerned with whether firms are investing well but solely focuses on misconduct in firms' financial accounting. Such misconduct undermines capital markets' role of efficiently allocating corporate resources and also impairs stakeholders' trust in corporations (Amiram et al., 2018). For instance, in the recent German high-profile case of Wirecard, the company had a market capitalization of more than 20 billion Euros at some point before its misconduct-related bankruptcy.

However, most jurisdictions around the world do not have strong institutions in place to identify and enforce deceitful accounting.[7] [8] The United States are likely the world leader in such enforcement (e.g., Brown et al., 2014) and have seen by far the most detected misconduct cases over the last decades, which allows training prediction models to improve enforcement processes (e.g., Bao et al., 2020; Dechow et al., 2011). Adopting the perspective of a non-U.S. institution, we ask whether U.S.-trained misconduct prediction models are applicable *far*-out-of-sample (i.e., in our European sample). While prediction models generally assume that misconduct characteristics (or their manifestation in ac-

[7]An implicit assumption in this essay is that the main goal of enforcement institutions is to detect or prevent misconduct for the benefit of the functioning of the overall economy (e.g., Amiram et al., 2018). We disregard the fact that any enforcement institution is itself a bureaucratic agency that may be subject to many incentives that are not aligned with effective detection of misconducted behavior. First, undertaking efforts to detect financial accounting misconduct behavior is (extremely) costly. Second, regulators may be self-interested, captured, uninformed, or even ideological. We refer to Hail et al. (2018) for a discussion of important frictions in the interplay of misconduct scandals and regulation.

[8]Essay 3 presented in the fourth chapter of this dissertation is based on a joint research project with Dan Amiram, Zahn Bozanic, and Johannes Roscher, so that I refer to the four of us when using plural pronouns in the context of this essay.

counting figures) somewhat persist over time, with our approach we additionally assume (and to some extent show) that such characteristics have common global properties.

In the next section of the introductory chapter, I discuss how the three essays introduced above belong to the broader area of firms' corporate governance.[9] A very uniting shared characteristic of the essays is that all three sets of insights are based on European firms. I provide the corresponding reasoning in the third introductory section. The fourth section briefly presents each essay's main methodology and results so that the fifth can discuss the dissertation's contribution. Finally, the last introductory section outlines the structure of the following chapters.

1.2 The dissertation in the broader context of corporate governance

1.2.1 Connection to various corporate governance aspects

The essays of this dissertation touch upon various aspects of firms' corporate governance. This can be illustrated when employing a rather broad meta understanding of corporate governance as, for instance, in Shailer (2018), where *corporate governance encompasses the roles of governments and regulators, capital market participants and existing capital providers (i.e., shareholders and debtholders), executive labor markets, as well as other stakeholders like suppliers and customers, interest groups, or internal participants (e.g., employees) in influencing the processes, structures, or mechanisms to control (i.e., those responsible for managing the firm) and direct (i.e., strategically guide) the firm.*[10]

One of the fundamental aspects of corporate governance considerations is to ensure the efficient allocation of corporate capital when ownership and control of the capital are separated (e.g., Jensen and Meckling, 1976) and there is asymmetric information (e.g., Myers and Majluf, 1984). Many governance mechanisms therefore directly (e.g., through the functioning of firms' investment committees) or indirectly (e.g., through designing ex-

[9] I refer to the overall introduction, the three essays, and the overall conclusion as *Chapters* 1 to 5. Within each chapter, I refer to different *Sections* regardless of their hierarchy level.

[10] The understanding of corporate governance used in the first essay is similar but more empirically operational with a stronger focus on specific (measurable) mechanisms or characteristics in place in a single firm.

ecutives' compensation systems to set the right incentives) relate to corporate investment. This dissertation heavily relates to corporate investment along the following dimensions. Given that executives can have incentives for too little or too much investment, **Essay 1** is concerned with the optimal level of investment. **Essay 2** improves our understanding of firms' intangible investment in human capital, given that such investment is difficult to make and measure. **Essay 3** relates to the prediction of financial accounting misconduct. Since one of the main goals of the accounting system is to provide accurate information necessary to facilitate capital allocation, and misconduct can hamper this system, this essay is also related to investment on a high level. Due to the breadth of different investment contexts that this dissertation handles, I dedicate a section to agency and corporate investment addressed by the three essays.

Corporate governance further includes all kinds of regulatory aspects, many of which tackle problems that may stand in the way of efficiently allocating capital to firms' investment opportunities. Governance regulation and enforcement begin with very general investor protection laws (e.g., La Porta et al., 2000). They comprise a series of more specific legislation, such as mandates for firms to report certain financial figures (**Essay 2**) or regulation of how to compensate executives (**Essay 1**). Eventually, to ensure good governance and protect investors, governments also implement a series of processes to enforce accounting rules and detect misconduct (**Essay 3**). Given the breadth of regulatory aspects in general as well as in my essays, I dedicate another section to a discussion of the economics of regulation in the context of this dissertation.

Finally, corporate governance measures also protect various stakeholders other than investors who are interested in firms functioning well. Next to shareholders, firms' (potential) employees are affected by firms' investment in the development of human capital and market participants' recognition of such investment (**Essay 2**). In addition to employees, quite a number of stakeholders like suppliers, customers, and also the general public are concerned about financial accounting misconduct as well as the prediction and detection thereof (**Essay 3**). In particular, firms' employees and the general public take substantial interest in executive compensation and are affected by whether firms invest too little or too much (**Essay 1**). I also dedicate a section to a discussion of how my dissertation relates to other stakeholders' interests.

1.2.2 On agency and corporate investment

Already recognized by Berle and Means (1932) and Jensen and Meckling (1976), a central theme in modern finance and accounting literature is that executives of (public) firms enjoy some discretion and may pursue objectives that do not necessarily coincide with those of the principals (i.e., the capital providers) after signing the contract (i.e. moral hazard) (Jensen, 1986). Executives may have incentives to grow their empires beyond the optimal size and, for instance, over-invest in projects with negative net present value, especially when they control abundant financial resources. Alternatively, they may seek to lead an entrenched and quiet life by maintaining the status quo and overpaying their workers, in particular the bureaucratic apparatus (Bertrand and Mullainathan, 2003). Shareholders, therefore, have reason and require (corporate governance) mechanisms to monitor and control executive behavior regarding capital investment.

Essay 1 analyzes how firms' investment deviates from the optimal level. I hypothesize that the length of the duration in the executive compensation is a core tool to mitigate misaligned incentives. Practitioners often believe that merely granting equity-based compensation provides incentives to forgo projects when they do not enhance value (i.e., over-investment). Yet, when the executive decides upon a project that destroys value but helps build an empire, she considers both the value reduction of current stock (options) and the aggregate future benefits stemming from steering a larger firm. Bebchuk and Fried (2003) therefore argue that it is actually the duration length of the incentive system that needs to be matched to the horizon of the underlying investment problem. This consideration ties well into the motivation for the executive pay regulation that I utilize as a shock to the duration length, as the regulator apparently considers this a primary tool to address inefficient corporate investment (Deutscher-Bundestag, 2009).

A specific problem that the German government intended to address was that of excessive risk-taking. However, executives' tendency to take too little risk may even be the bigger problem. Given that shareholders are far more diversified than executives of single firms, executives may be too risk-averse from the shareholders' perspective (e.g., Ross, 2004). Career concerns (Holmstrom and Costa, 1986), in particular related to bankruptcy risk (Rose-Ackerman, 1991), may also make executives too conservative. Further, long-term investments with risky outcomes, such as intangible investments, may be difficult for accounting systems to be impounded into current earnings (e.g., Lev, 2019), leading to myopic disincentives to make such investments (e.g., Stein, 1988). This makes it important to assess whether shareholders timely understand such investments, allowing executives to commit to them. **Essay 2** does so for the intangible investment in human capital.

Finally, capital markets will only work and meet their role in allocating resources to investment if financial accounting rules are enforced and misconduct is effectively prosecuted (e.g., Amiram et al., 2018). As **Essay 3** engages with the enforcement of such misconduct, this is broadly related to the governance of corporate investment as well. Many of the identified European cases tabulated in Appendix C.4 relate to firms that falsify (i.e., inflate) the profitability or disguise (i.e., understate) the obligations associated with investments they made or want to continue to make.

1.2.3 On the economics of regulation

Disclosure regulation is probably one of the most significant fields of regulations that concern corporate governance. This is reflected by intense debates and development effort for reporting standards like International Financial Reporting Standards (IFRS), U.S. Generally Accepted Accounting Principles (GAAP), as well as local GAAPs or beyond (e.g., disclosure mandates as specific as mine safety performance in section 1503 of the Dodd–Frank Act or as broad as sustainability reporting in Directive 2014/95/EU). While there can also be voluntary disclosure, firms only voluntarily disclose if 1) they assume investors to know that the firm has a certain information, 2) it is impossible to lie, and 3) the disclosure is not costly, whereas voluntary disclosure will be less than complete as soon as one of the assumptions is violated (Grossman and Hart, 1980). Going beyond what is relevant for investors, the disclosure of most information is usually a public good with externalities for many stakeholders like, for instance, competitors, that firms cannot internalize or that may even hurt them when it comes to disseminating proprietary information. One can therefore expect that firms disclose less than what is socially optimal (Zingales, 2009), providing room for disclosure regulation.

Quantifying the costs and benefits of disclosure regulation is, however, quite difficult, as it requires causally measuring several effects of a disclosure regulation which is rarely possible.[11] Particularly in new terrain, like disclosure around intangible investment into human capital, association studies are needed to build the knowledge inventory of potential economic outcomes (Leuz and Wysocki, 2016). In **Essay 2**, we, therefore, analyze a disclosure practice that already exists. We assess whether different market participants timely impound the information content that is implicit in the disclosure practice. Our study allows to directionally show that the existing disclosure is meaningful and could be made mandatory under other GAAP systems and that it could be improved to allow better dissemination of the underlying information. In addition to this benefit for capital providers, there might be benefits for other stakeholders like employees. However, we are unable to quantify the costs of such disclosure. While merely disclosing total personnel expenditures (as it is the current practice under IFRS that we utilize) should not be so costly, detailed disclosure of firms' efforts investing in human capital could be far more costly to provide and may have significant value for employee interest groups or competitors that firms may not be able to internalize.

One major reason for wide-ranging disclosure mandates is that transparency can already often incentivize desirable behavior (Leuz and Wysocki, 2016). Sometimes transparency does not suffice, prompting regulators to adopt regulation that explicitly stipulates or prohibits corporate practices. The regulation that I analyze in **Essay 1** is a good example where transparency mandates had been in place for some years, but the regulator ultimately decided to implement a mandate that interferes with actual behavior. German listed firms already had to disclose their executive compensation system at the time of

[11] Optimal disclosure regulation analysis would actually require regulators and scientists to collaborate on random trials (e.g., on a random subset of firms) of potential regulation (or even deregulation) (Leuz and Wysocki, 2016).

the mandate that I utilize as a shock to the compensation systems in terms of duration length.[12] Similar to mandated disclosure, such an instance can provide a suitable setting to causally measure the effect of the mandate on corporate behavior. I argue this to be the case and lever the setting for measuring the effect on the efficiency of firms' investments. As such a strong intervention may also interfere with efficient contracting and lead to a decrease in investment efficiency (i.e., the outcome that I analyze), I motivate that the effects may be positive, negative, or insignificant.

Eventually, the effectiveness of most regulations is limited by the (expected) effectiveness of their enforcement. In accounting, enforcement can be interpreted broadly as mechanisms to ensure obedience to security laws, with financial reporting enforcement being an important aspect thereof (e.g., Ernstberger et al., 2012). Wide global differences in jurisdictions' enforcement are related to cultural backgrounds, legal origins, and the development of auditing professions (e.g., Brown et al., 2014; Gray, 1988; La Porta et al., 2000). **Essay 3** addresses some of the differences in enforcement efficacy related to financial accounting misconduct, which is a rare event that is difficult to identify, leaving most jurisdictions far behind the efficacy of the U.S. enforcers. We, therefore, assess whether non-U.S. enforcers can utilize the U.S. historic data of identified cases to train models that could improve their own enforcement processes.

1.2.4 On other stakeholders' interests

Finally, corporate governance measures also protect various stakeholders other than investors who are interested in firms functioning well.[13] The human capital creation that is analyzed in **Essay 2** is also important for current and future employees. There may likely be benefits for employees when firms improve human capital investment disclosure. However, this does not mean that such investments are solely beneficial for employees, as

[12] I am referring to the 'Management Board Remuneration Disclosure Law' (Vorstandsvergütungs-Offenlegungsgesetz (VorstOG)) implemented in 2005 and the 'Appropriate Director Compensation Act' (Vorstandsvergütungs-Angemessenheitsgesetz (VorstAG)) implemented in 2009.

[13] Literature often refers to these other stakeholders as non-financial stakeholders (e.g., John and Senbet, 1998). Extending the perspective on corporate governance to include such non-financial claims is not new (e.g., Freeman and Reed, 1983), but became very popular in recent years.

building up firm-specific human capital may make them dependent on the specific firm.[14] From the perspective of the employees, investing in human capital that is particularly valuable for the firm and less valuable in other firms requires them to have mechanisms that protect their stakes like life-time employment or severance pay and internal job ladders (e.g., Blair, 2003). Therefore, while we do not assess this, it seems reasonable that employees and their interest groups like unions should benefit from more detailed human capital investment disclosure. At the same time, there may be arguments why firms would not be able to internalize these benefits as it might weaken their power in contracting with employees or provide information that competitors could be able to exploit.

Essay 3 takes into account numerous other stakeholders, particularly suppliers, customers, and the general public by describing a tool that could substantially improve enforcement of financial accounting misconduct. Enforcement of accounting rules protects other claim holders from being exploited by shareholders. Shareholders of a limited liability corporation might walk away from the liabilities when profits are insufficient to cover them and leave the liquidation value to involuntary creditors like unpaid suppliers, unsupplied customers, or the general public that deals with the pollution that the corporation leaves behind. Shareholders essentially hold a put option with a strike price of zero that allows them to get rid of the corporation without covering the societal costs it leaves behind (e.g., MacMinn and Han, 1990). Although insolvency is neither the only possible outcome of financial accounting misconduct nor the only reason to enforce it, misconduct often delays insolvency, increasing societal costs.[15] Protecting suppliers, customers, and the general public therefore is a major motivation for the regulator to enforce accounting rules.

Finally, even though **Essay 1** centers around a core agency conflict between shareholders and executives, it also touches upon the interests of other stakeholders. First, employees and the general public are interested in executive compensation as a corporate governance tool (Edmans et al., 2017b; Jensen and Murphy, 1990). In Germany, it is

[14] An alternative term often used in this context is *relationship-specific capital* (e.g., Jaggia and Thakor, 1994).

[15] This is also visible in the summaries of misconduct cases tabulated in Appendix C.4.

even common to have strong employee representation on the compensation committee.[16] Second, whether a firm invests efficiently, sooner or later, directly or indirectly affects many stakeholders of the firm. For instance, signals sent by an inefficiently investing firm may lead to the same behavior in its suppliers and customers.

1.3 The use of European data in the three empirical designs

The empirical accounting literature in general, as well as in the particular fields of this dissertation, heavily relies on data from U.S. firms. The specific research questions of this dissertation, however, are best addressed by leveraging European settings. In this section, I lay out which European data is well-suited to provide empirical insights into the respective aspect.

Both the bulk and the frontier of the empirical insights on the determinants and (probably more importantly) the effects of executive compensation are based on U.S. (listed) firms' data (e.g., Edmans et al., 2017b). One major reason for this is the high level of transparency on executive compensation that has been required for a long time (i.e., in annual proxy statements) along with the databases collected accordingly. Standard & Poor's database ExecuComp containing compensation details for the executive team of S&P 1500 firms since the early 1990s is the most extensive and standardized collection used to analyze executives' annual fixed compensation, bonus payments, granted options, and other features.[17] More recently, Institutional Shareholder Services' Incentive Lab was established as the most comprehensive database, containing details on performance metrics and goals as well as payout structures on all incentive awards for largely the same set of firms since the late 1990s.[18] However, next to comprehensive panel data, effect studies require suitable exogenous variation in compensation design to get closer to

[16] For instance, this has been established in BASF SE, Deutsche Post AG, and Volkswagen AG for many years according to their annual reports.

[17] For instance, Gillan et al. (2018) count more than 1,000 published articles using this database, more than half of which made it into the leading accounting and finance journals.

[18] In Europe, for instance, transparency on executive compensation came a lot later with regulation being enacted in many countries following a recommendation by the European Union (EU) in 2004 (2004/913/EC). Accordingly, database collection only started later. Yet until today, there is substantially less transparency so that databases like Institutional Shareholder Services' Incentive Lab can only offer a lot less coverage and detail for European firms than they can offer for their U.S. flagship.

causal inference.[19] Such variation may well be observed outside the U.S. firm universe.

In **Essay 1**, I exploit a country-level mandate to increase the duration of executive compensation that serves as a quasi-exogenous shock to the compensation design. Hence, subsequent effects are reasonably likely to be caused by the change in the compensation. Such a far reaching interference with private contracting (that comes with substantial risks of creating unanticipated and undesirable frictions) would not be likely in the U.S. market. However, Germany enacted 'unprecedented interference with private compensation schemes' (Hitz and Müller-Bloch, 2015, p. 659), while other European countries did not and can serve as counterfactuals.[20] Leveraging such a setting requires adapting the dominant U.S.-based operationalizations of the dependent variable (i.e., investment efficiency) in the empirical literature to the European multi-country and multi-currency setting.[21] Being sensitive to both the opportunities and the institutional specifics of European data is a dominant theme across the essays of this dissertation.

Essay 2 replies to calls for reconsidering accounting for intangible investments (e.g., Lev, 2019), most importantly investments in human capital (e.g., Rouen, 2019). It extends the literature on how stock markets recognize firms' intangible investment, i.e., expenditures with future benefits other than capital expenditures. Accounting standards usually do not treat these expenditures as investments but require firms to expense them as incurred. The most important literature stream is that on future value estimation of several different functional input resource expenditures, i.e., of research & development (e.g., Lev and Sougiannis, 1996), advertising (Chan et al., 2001), and selling, general & administrative (Banker et al., 2019). The empirical insights in this area are again largely

[19] Any empirical endeavor attempting to estimate the effect of some compensation design on some outcome is subject to concerns about endogeneity of the independent construct. Firms' choice of the compensation design is not random, but may be influenced by any other unobservable. Researchers therefore try to exploit settings in which an exogenous force enables treatment and control firms to be identified.

[20] To identify a quasi-exogenous change in the length of the incentive duration in the U.S., one could turn to the setting used in Ladika and Sautner (2020). However, this one-time non-lasting change would not be useful to investigate the effect on investment efficiency that is commonly assessed over a couple of years.

[21] Firms' investment efficiency, the dependent variable that I analyze with respect to the compensation duration mandate, is a concept where again the bulk of the literature and the dominant operationalizations are based on U.S. firms. One exception is, for instance, Chen et al. (2011)'s empirical insights into private firms in emerging markets.

based on U.S. data. While solely separating the reporting in functional terms as in the items above is common under U.S. GAAP, outside that universe (e.g., under IFRS), it is also common to differentiate by nature (i.e., personnel vs. material) (e.g., Baker et al., 2003).[22] Although broader and possibly less precise, this allows capturing firms' human capital creation implicit in personnel expenditures more holistically.[23] By looking at personnel expenditures, we can include all areas of the firm where such expenditures may have the character of an investment. For instance, also expenditures for blue-collar staff that may create future benefits are part of this, while they are usually part of the cost of sales and not included in any of the above functional reporting items.

There are several reasons to use European data for extending established methodologies by estimating the future value of personnel expenditures. First, IFRS requires disclosure of personnel expenditures even if firms disaggregate by function and almost all European listed firms must report according to IFRS since 2005.[24] [25] Second, Europe provides an opportunity to include firms from different countries that belong to a largely coherent economic area. For instance, employee mobility is strong across internal borders (e.g., European Commission, 2019) and less strong beyond European borders. Third, factor returns in European stock markets are generally comparable to those of the U.S. market so that one can reasonably analyze market recognition of our new proxy in a similar way as in the U.S. and relative to earlier results in this literature.[26] Therefore, bringing an established methodology to a European setting again implies a unique opportunity to extend the literature in that field.

[22]This is due to the impact of several local GAAP traditions such as the German or the French GAAP that shaped the considerations around the development of IFRS.

[23]For the sake of simplicity, throughout this introductory chapter, I only use the term personnel *expenditures* even though the proxy is based on reported personnel *expenses*, most of which relate to expenditures that are expensed as incurred.

[24]U.S. GAAP does not require such disclosure and only few firms report it voluntarily. Two decades ago, Ballester et al. (2002) found that only 10% of publicly listed U.S. firms consistently report labor-related costs. Even today, only 15% make this cost item transparent (O'Byrne and Rajgopal, 2022).

[25]The European Union prompted its member states to adopt IFRS in 2005 and other larger European economies like Norway or Switzerland that are not part of the European Union enacted simultaneous adoption rulings.

[26]For instance, Fama and French (2017) show that factor return patterns for their (more developed) European countries are generally comparable to the North American ones. While Foye et al. (2013) show that emerging Eastern European countries have some more differences in the patterns, they make up only a small fraction of our data.

Essay 3 builds on yet another empirical literature stream that predominantly relies on U.S. data. The literature on how to predict financial accounting misconduct uses data from U.S. firms as they play in the largest among the most developed financial markets (e.g., Rajan and Zingales, 1998; Wurgler, 2000), where they are subject to the toughest enforcement (e.g., Brown et al., 2014; Prentice, 2005). To identify cases of misconduct, most research relies on data from the U.S. Securities and Exchange Commission (SEC) which issues Accounting and Auditing Enforcement Releases (AAERs). Over the last decades, the SEC issued up to a high double-digit number of cases per year (e.g., Bao et al., 2020). While we should assume that many cases of misconduct still remain undetected (i.e., false negatives), the AAER database can reasonably be used to train and test models for misconduct prediction. Researchers developed and refined numerous models with early contributions dating back more than two decades (e.g., Summers and Sweeney, 1998) and very recent contributions showing continued interest in the problem (e.g., Brown et al., 2020). While these U.S.-based predictability findings facilitate early detection in U.S. firms, they may be even more helpful for weaker enforcement entities that do not have an own long history of many identified misconduct cases required to train such models.

To test whether U.S.-trained models can be applied in other markets, we turn to firms in the major European economies for the following reasons. First, these firms are generally exposed to similar institutional settings in terms of auditing of financial statements and enforcement of compliance that are, however, weaker.[27] Second, similar to the first two essays, it is beneficial that almost all listed firms in those countries have a common set of GAAP, namely IFRS. Third, our vision of a supranational enforcement entity pooling resources to detect or deter financial accounting misconduct in the overseen firms is very plausible in the European setting with the newly established European Securities and Markets Authority (ESMA) that could take on such a task.

[27] For instance, Brown et al. (2014)'s proxies for *audit* and *enforcement* show that (with the exception of the United Kingdom) our eight major European countries have weaker institutional settings in particular regarding enforcement in the years 2002, 2005, and 2008.

1.4 Main methodologies and results

1.4.1 On Essay 1: Does Longer Duration of Executive Compensation Foster Investment Efficiency?

To shed light on whether mandated longer duration of executive compensation fosters investment efficiency, I compare German firms with firms based in other European countries or a matched subset of these firms in a difference-in-differences design.[28] I use different proxies for investment and operationalize firms' abnormal investment with the residuals of commonly used investment model regressions that follow the accelerator theory (e.g., Biddle et al., 2009; Chen et al., 2013, 2011) as well as Tobin's q theory (e.g., Li et al., 2018; McNichols and Stubben, 2008).[29] I find strong support for the notion that mandated longer duration increases investment efficiency across measures of investment and models of expected investment. My results reveal statistically significant reductions in abnormal investment following the adoption of the VorstAG act. Treated firms show a reduction in abnormal investment that amounts to about 10% of mean investment, which is also economically significant.

In cross-sectional tests, I find evidence that mandated longer duration has the largest effect on investment efficiency in firms with non-independent compensation committees or a weak corporate governance score. My results are robust to employing a German set of non-treated control firms alleviating concerns that country-level factors other than the VorstAG regulation drive the effects in my main analyses. Finally, I also differentiate reductions in over- and under-investment in additional analyses which reveal that reductions in over-investment drive the treatment effect. Longer duration therefore seems to hamper over-investment but may not be effective in providing incentives for under-investing executives to invest more. This could imply that it is easier to force an over-investing executive to internalize later negative effects of the investment decisions via long-term

[28] Firms in the STOXX Europe 600 index comprise the main sample.

[29] There is no single widely accepted proxy for firms' investment, it is therefore common to use different ones and put the results next to each other. The same holds for the models applied to estimate the efficiency of the investment. As I mainly use two different investment proxies and two different investment models, every analysis is basically performed four times. Reporting consistent results across these four alternatives helps build confidence in that the operationalizatios of the constructs are actually measuring what they are supposed to measure.

compensation contracts. While the methodology for this differentiation is not yet established, these results demand caution as to assigning universal benefits to a longer duration of executive compensation.

1.4.2 On Essay 2: The Stock Market Valuation of Human Capital Creation

In this essay, we build a proxy of firms' human capital creation implicit in their personnel expenditures (PE). To do so, we adapt a methodology to identify at the firm-year level how successful a firm is at investing in personnel (i.e., the human capital investment) through regressing operating income on several earlier years of PE.[30] First, we validate our proxy by showing that it is associated with firm characteristics that are likely to be related to the importance of human capital creation. Next, we examine the association between the creation of human capital and contemporaneous stock price. The results suggest that the stock market, to some extent, differentiates between the current operating expense component of PE and the future value of PE, which is treated as an intangible asset. However, this does not tell us whether market participants fully and timely recognize this future value of the intangible asset included in PE.

Our main analyses therefore concern the predictive power of our proxy for sell-side analysts' earnings forecast errors and firms' future stock returns. First, we find a significant positive relationship between the magnitude of the proxy and absolute as well as signed forecast errors. These results suggest that analysts fail to incorporate the full value of the investment component of PE into their forecasts, and are too pessimistic, on average. Next, we build portfolios based on our proxy for human capital creation. Analyzing these portfolios reveals statistically and economically meaningful results. Value-weighted (equal-weighted) long-short investment strategies based on our proxy return annualized abnormal alphas of 6.5% to 7.8% (3.5% to 4.8%) in the subsequent year. These results suggest that the market fails to fully impound the human capital development embedded in PE. The results are also robust to numerous alternative specifications, including assigning portfolios based on industry, excluding firms from countries with illiquid currencies,

[30] I use versions of the term *human capital creation/investment (implicit in PE)* in this introduction and do not differentiate between the specific variables estimated in the essay.

and using different factor models. Lastly, we find that the abnormal portfolio returns decrease monotonically over time, with (still) statistically significant value-weighted returns of 5.1% in the second year after portfolio formation, and insignificant returns of 3.1% in the third year.

1.4.3 On Essay 3: ***Far*-Out-of-Sample Accounting Misconduct Prediction – Application to Non-U.S. Cases**

The methodology in this essay is best understood when taking on the perspective of a non-U.S. enforcement authority (e.g., a European authority) that tries to improve the identification of misconducted financial accounting by levering the U.S. history of identified cases for predictions. To test whether this works, we choose the more straightforward models among the established ones and focus on utilizing readily available financial statement items. We use a common set of 28 raw items (that firms report and database providers collect around the globe) in either ratio-based Logistic Regression (LR) (Dechow et al., 2011), or in a more advanced learning algorithm called RUSBoost Ensemble Learning (Bao et al., 2020). The direction of the main results is consistent across models and specifications. The European *far*-out-of-sample prediction performance is similar to - or even slightly higher than - the U.S. benchmark prediction performance.[31]

Since we use two different databases to collect the data for the predictors for the U.S. training (Compustat) and the European testing (Compustat Global) samples, we address inconsistencies in the two by comparing the figures for firms covered by both. When we remove raw items with large deviations, we observe the following: As one would expect, the benchmark performance for the U.S. sample mostly declines, because the reduced models are likely less good than the optimal models identified by prior research. For the ratio-based LR models, the prediction performance for the European out-of-sample

[31] For instance, at a cutoff at the top 10% of predicted misconduct probabilities (i.e., this means that the top 10% of firms regarding misconduct probability would be flagged accordingly), the European out-of-sample results for the LR models have Sensitivity metrics of between 14 and 22% (i.e., this is the percentage of actual misconduct observations flagged correctly), slightly higher than the results for the U.S. benchmark prediction, which is between 12% and 21%. The results for the RUSBoost approach are even stronger for both the U.S. benchmark prediction and our *far*-out-of-sample application with Sensitivity figures around and above 30%.

application stays about the same with both slight improvements and declines. For the raw item-based RUSBoost models, however, the European prediction performance even improves. These results build strong confidence in that the *far*-out-of-sample application generally is a promising approach that can be further adapted in line with specific data availabilities.

1.5 Contribution to public, academic, and regulatory debates

Similar to the dominance of U.S.-based empirical insights in many fields of accounting literature, scholarly contribution is typically targeted at a U.S.-based audience. The contribution, as described in each essay's introduction, is tailored accordingly. Slightly deviating from this and in light of the dissertation's focus on European settings, I present the contribution targeted at European market participants in the following.

Essay 1 makes three contributions. First, I contribute to the debate on whether long-term incentive duration can improve firm outcomes. On the one hand, both theorists and empiricists argue for and show benefits of longer duration (e.g., Bebchuk and Fried, 2010; Flammer and Bansal, 2017; Jochem et al., 2018; Larcker, 1983). On the other hand, there are arguments for keeping incentives short-term under certain conditions (e.g., Evans et al., 2017; Schroth, 2018). While the potential for classical principal-agent problems between shareholders (i.e., principals) and executives (i.e., agents) is straightforward, recent survey responses indicate that there is even disagreement between shareholders and the supervisory board members that they install to act on their behalf: In a UK sample, it appears that most shareholders believe that more long-term incentives would lead to better investment decisions. Supervisory board members, however, view incentives to be sufficiently long-term and raise concerns about further lengthening them (Edmans et al., 2021). My study allows an estimate of the effect of an exogenously triggered lengthening in incentive duration and shows that it reduces firms' abnormal investment. I do generally interpret this as a positive effect on real investment behavior. However, there seems to be little evidence for reductions in under-investment, which demands caution in claiming universal benefits of longer duration.

Second, I contribute insights into the interplay of corporate governance characteristics and executive pay duration. Pay duration may generally be a substitute or a complement to alternate forms of corporate governance (Gopalan et al., 2014). This means that either duration is designed to substitute for control mechanisms in firms where the general governance system is weak or that already better-governed firms also put in place better-fitted pay duration. More specifically, the independence of the supervisory board may play a major role in determining the right incentive pay (e.g., Knyazeva et al., 2013; Li and Wang, 2016) and providing the right monitoring (e.g., Kumar and Sivaramakrishnan, 2008) for optimal investment decisions. My cross-sectional tests indicate that longer duration matters in firms with low compensation committee independence as well as weak overall corporate governance. This rather suggests a complementary relationship between governance and incentive duration, i.e., better governed firms implement better designed duration. This insight may be particularly important for European markets where the governance systems are quite diverse, and where supranationally driven convergence is an important topic (e.g., Cernat, 2009).

Third, any study that estimates the effects of a regulatory action provides information for regulators and political decision makers. The literature on the consequences of governance regulation in general and the effects of compensation regulation in particular is growing strongly. There is a contemporaneous debate on the EU level regarding whether and how to reform executives' and supervisory boards' duties towards sustainable corporate governance. One may say that a group of (empirical) researchers is currently at war with the European Commission's 'Study on directors' duties and sustainable corporate governance'. The study recommends far-reaching regulatory action, while the scholars demand a more thorough review and analysis of existing governance research before implementing regulatory action (e.g., Roe et al., 2020). I add to this debate the notion that mandated longer duration leads, on average, to more efficient investment decisions, which is in line with shareholders' long-term interests. My analyses suggest that the generalization of this notion to other jurisdictions may depend on prevalent governance characteristics, particularly whether there are any independent board members responsible for compensation contracting. Moreover, the expected effects in other jurisdictions may depend on whether over-investment is a major issue. This means that the supranational EU regulator may prefer to respect single member states' institutional differences when regulating the design of executive compensation systems.

Essay 2 also makes three contributions. First and most broadly, we provide evidence of the value of human-capital-related disclosures to market participants. There is little evidence of the relation between employee expense and future firm performance. Most closely related, Schiemann and Guenther (2013) show in a UK sample that employee expenses improve earnings predictability. We develop a proxy to extract the future value of the expenditure from the total expense and show that there is significant variation in the ability of firms to generate future value from their investment in employees through PE. We then show that our proxy is mispriced by market participants, most importantly, investors seem to not incorporate it very timely as there are significant risk-adjusted returns in future years.

Second, we make several regulatory contributions. Firstly, as we show abnormal returns in European IFRS-reporting firms, we inform regulators that there is room for more detailed or differentiated disclosure around firms' investments in human capital.[32] [33] One direction for such disclosure that our results point towards could be information regarding the role of training employees in the firm. Secondly, the contemporaneous convergence project between the Financial Accounting Standards Board and the International Accounting Standards Board discusses whether it is more informative to disaggregate costs by their function or by their nature, including a debate as to whether disclosure of nature of expense items like PE should also be mandatory under by-function systems.[34] Our

[32]The various regulators that one may think of are, among others, the single countries in our sample, supranational entities like the European Union, and most importantly the International Accounting Standards Board that constantly develops IFRS further.

[33]Our results indicate that there is some information content regarding firms' human capital investments that is not timely incorporated into stock prices by merely disclosing total PE. This suggests that there could be *benefits* for the functioning of the market when improving transparency on firms' PE in a way that markets learn more about the actual human capital investments. There may further be *benefits* for other stakeholders (e.g., employees) that the regulator may take into account but that are beyond the scope of our essay. However, we are by no means able to speak to *costs* of such disclosure that regulators would or should consider (e.g., Leuz and Wysocki, 2016).

[34]See, for instance, material on the Financial Accounting Standards Board and International Accounting Standards Board Joint Meeting on Primary Financial Statements in June 2018.

results suggest that there is some information contained in PE, which may generalize to the U.S. GAAP universe. Thirdly, the latter presumption may further be informative to the U.S. SEC, which recently passed an amendment to its Regulation S-K requiring firms to provide a description of the importance of their human capital resources to the underlying business.[35]

Third, we make a methodological contribution by expanding the emerging literature on the impact of firms' ability to generate future value from input resource expenditures. This literature so far focuses on functionally reported input resources like R&D (Chan et al., 1990; Eberhart et al., 2004; Lev and Sougiannis, 1996), advertising (Chan et al., 2001), and SG&A (Banker et al., 2019), none of which is mandatory to report under IFRS. In terms of outcomes of firms' ability, scholars analyze the effects on executive compensation and cost decisions (Banker et al., 2011; Chen et al., 2012; Huson et al., 2012) and market valuations (e.g., Banker et al., 2019). Until now, this research has paid little attention to the nature of the expense, broadly, and PE, specifically. This literature also only relies on evidence for U.S. firms. We examine an intuitive, widely reported input resource that can be analyzed in an IFRS cross-country setting. This offers opportunities for future research on the impact of the ability to generate intangible assets in the rest of the world.

Essay 3 makes one major and two minor contributions. Its first and foremost contribution to accounting scholarship lies in its regulatory implication. We are the first to show that U.S.-trained misconduct prediction models may be applied *far*-out-of-sample to preselect firms and track down deceitful behavior. This insight offers a potential tool and workflow to any market outside the reach of U.S. enforcement. One may even go so far as to describe this essay's contribution as *untheoretical*. From a theoretical perspective, it is rather clear that the suggested approach should work. However, there may be countless practical obstacles, such as different reporting frameworks, currencies, databases, and others. Not least does the fact that no enforcement institution yet uses such an approach

[35]See the SEC's proposal to modernize Item 101(c) of Regulation S-K (August 2019) to include a description of the firm's human capital resources and objectives.

provide an argument for perceived barriers.[36] The contribution therefore lies in showcasing that the impediments can be overcome as well as providing a recipe on how to do it.

While our sample of European IFRS reporters covers different countries with different currencies and legal systems, it is arguably not so distant from the U.S. reporting environment as other markets in the world.[37] Our robustness tests, however, indicate that the established prediction approaches, in particular most recent advancements like RUSBoost Ensemble Learning, are quite robust to many design choices and data limitations. Our results regarding the out-of-sample applicability should therefore generalize to other (more institutionally distant) jurisdictions. Regardless of the latter, our results could be most valuable to European regulators as the essay provides a workflow description that could directly be rolled out by an enforcer overseeing any subset of the firms in our sample.

The essay's second contribution lies in improving our understanding of the factors determining financial accounting misconduct across different countries. We offer two interpretations for the strength of our results in terms of *far*-out-of-sample prediction performance, where the European predictions usually even outperform the U.S. benchmark predictions. Firstly, our European cases could be more extreme than the average U.S. case as they were (eventually) identified by weaker institutions. It could be that we only have the most extreme correct positive cases in our European sample along with many false negatives that never got detected and are weaker on average. Secondly, it could be that European misconduct firms somewhat adopted earlier U.S. misconduct schemes, making it easier to detect them when training with U.S. data. Both explanations are in line with the notion that misconduct has common properties at least across the U.S. and Europe, if not across the globe.

We further see a third contribution in offering our pioneering work on *far*-out-of-sample application to all kinds of other prediction contexts in financial accounting literature. For

[36] To the best of our knowledge, no enforcement institution yet uses this approach, however, enforcers may also use such an approach without publicly spreading information about it.

[37] While large markets (with many misconduct cases) like China or Japan would also be of particular interest for our approach, the author team does not have the institutional knowledge to collect comparable cases in those markets.

instance, our suggested workflow may be used to improve predictions of other rare events like corporate bankruptcy (e.g., Agarwal and Taffler, 2008; Charitou et al., 2004; Jones et al., 2017) or bank failure (e.g., Jin et al., 2011; Tam and Kiang, 1992). Interestingly, those two literature streams are again strongly based on U.S. data, together with UK data, even though both issues should be important across the globe.

1.6 Structure of the dissertation

The remainder of this dissertation is organized in the following manner. In Chapter 2, I present evidence on the effects of mandated longer incentive duration of executive compensation on investment efficiency. Chapter 3 analyzes the stock market valuation of human capital creation. Chapter 4 reports the results of applying accounting misconduct prediction models to non-U.S. cases. Finally, in Chapter 5, I provide a short summary of each essay and discuss some limitations.

2 Does Longer Duration of Executive Compensation Foster Investment Efficiency?

Abstract

In this paper, I examine whether longer duration of executive compensation influences investment decisions. I exploit a regulation designed to foster long-term orientation in executive compensation as an exogenous trigger to lengthen executives' incentive duration. I find that treated firms reduce their abnormal investment relative to control firms, implying an increase in investment efficiency. These results are robust to different measures of investment, several models of expected investment, and different plausible control groups. The treatment effect is economically significant, as the reduction in abnormal investment amounts to about 10% of mean investment. It appears that a mandated longer duration has the greatest effect on investment efficiency in firms that had a low degree of compensation committee independence before the shock. Further, it seems that the lower abnormal investment stems to a greater extent from reductions in over-investment.

Keywords: Executive compensation, Incentive duration, Investment, Investment efficiency, Regulation
JEL classification: J33, M12, M48, M52, G31

Authors: Matthias Regier
Lead author: Matthias Regier

Status: Accepted at European Accounting Review
https://doi.org/10.1080/09638180.2021.1989317

Acknowledgements Accepted by Christian Hofmann. I gratefully acknowledge the helpful comments given by two anonymous reviewers, Alexander Brüggen (discussant), Jürgen Ernstberger, Gunther Friedl, Manuel Herkenhoff, Sydnee Manley (discussant), Garen Markarian, Ethan Rouen, Peter Schäfer, Eline Schoonjans, and seminar participants at 2019 International Workshop on Accounting and Regulation, 2019 Conference on Performance Measurement and Management Control, 2019 AAA Midwest Region Meeting (Best Student Paper Award), 2020 Annual Conference for Management Accounting Research, and 2021 EAA Annual Congress. Additional materials are available in an online Supplement at the journal's Taylor and Francis website. All remaining errors are mine.

2.1 Introduction

In this paper, I examine whether and how mandating companies to implement longer durations of executive compensation affects investment efficiency.[38] Executive compensation is one of the corporate governance tools that draws major attention and creates major controversy about corporations among the general public (Edmans et al., 2017b). With regard to compensation, the duration of the incentive is a crucial aspect (e.g., Bebchuk and Fried, 2010; Gopalan et al., 2014) as it has a direct effect on executives' investment decisions. While research finds that firms reduce investment when there are strong short-term incentives (Edmans et al., 2017a; Ladika and Sautner, 2020) and increase investment when executive compensation is more long-term (Flammer and Bansal, 2017; Larcker, 1983), it is indeed the case that neither less investment need be less efficient nor more be more efficient. I therefore set out to empirically investigate the link between longer duration of executive compensation and the efficiency of firms' investment.

Evaluating this link is difficult because firms typically design their compensation scheme independently, whenever shareholders or their representatives on the supervisory board negotiate with executives.[39] I exploit a unique setting, in which an exogenous regulation enables treatment and control firms to be identified. The introduction of the German 'Appropriate Director Compensation Act' (Vorstandsvergütungs-Angemessenheitsgesetz (VorstAG)) in July 2009 is regarded as 'unprecedented interference with private compensation schemes' (Hitz and Müller-Bloch, 2015, p. 659). This act demands that remuneration contracts are reviewed and oriented towards the long term. Variable compensation components should be based on a multi-year assessment period (i.e., multi-year evaluation and deferment of awarded variable compensation), conditional payback provisions are recommended, and most specifically, the minimum vesting period

[38]For expositional convenience, I use the term *longer (incentive) duration* throughout the paper to refer to changes in compensation systems that relate to the adoption of a multi-year performance assessment, deferral of an awarded variable compensation, gearing of performance-related components to long-term determinants, increasing the vesting period of stocks or stock options, adopting a bonus-malus system, or requiring investment in the firm's own stocks or stock options. This use of the term extends Gopalan et al. (2014)'s understanding of duration to include the prolonged incentive effect inherent in adopting multi-year assessments, deferrals, and bonus-malus schemes.

[39]Recent survey responses (Edmans et al., 2021) indicate that shareholders and supervisory board members disagree in several regards on how to set executive compensation.

for stock options (if awarded) is increased from two to four years.

A mandated increase in the duration of executive compensation can affect a firm's investment efficiency either positively, negatively, or not at all. By increasing the duration, the board increases the length of the window during which the executive performance can be evaluated. As a consequence, the future pay-off to current-period investment will be realized to a larger extent in observable performance. This may lessen the incentive to empire build (Jensen, 1986), because if executives engage in unprofitable projects to increase their span of control, this unprofitability will materialize within the lengthened evaluation window. Likewise, the longer evaluation window mitigates myopic disincentives for current period investments with future payoffs (Graham et al., 2005), as more of the payoffs will be considered within the lengthened evaluation period.

However, compensation regulation may also interfere with efficient contracting and lead to a decrease in investment efficiency. Strong short-term objectives may be optimal when the executive's short-term stock price manipulation propensity is high (Schroth, 2018), when the executives have valuable alternatives, and when firms' operations are highly uncertain (Evans et al., 2017). Finally, if the regulation is too vague, firms might simply window dress without implementing any substantial changes to the compensation systems, leading to no effect. Which of these opposing forces dominates remains to be empirically explored.

To do this, I compare German firms with firms based in other European countries in a difference-in-differences design. In addition to comparing German firms with all other European firms, I obtain a matched subsample of the other European firms. I use different proxies for investment and operationalize firms' abnormal investment with the residuals of commonly used investment model regressions that follow the accelerator theory (e.g., Biddle et al., 2009; Chen et al., 2011, 2013) as well as Tobin's q theory (e.g., Li et al., 2018; McNichols and Stubben, 2008).[40] I find strong support for the notion that mandated longer duration increases investment efficiency across measures of investment and models of expected investment. My results reveal statistically and economically

[40]Smaller residuals imply lower abnormal investment, which implies higher investment efficiency.

significant reductions in abnormal investment following the adoption of the VorstAG act. Treated firms invest about 10% more efficiently, when viewed relatively to the sample mean of the respective investment variable.

Next, I analyze cross-sectionally whether the treatment effect is related to firms' pre-treatment corporate governance characteristics. In line with research finding positive effects of supervisory board independence on compensation incentives (e.g., Knyazeva et al., 2013; Kumar and Sivaramakrishnan, 2008), I investigate whether the reduced abnormal investment depends on the degree of independence that the firm's compensation committee had in the pre-period. The mandate for longer duration could impose a stronger change on the incentive system for executives in firms in which fewer independent supervisory board members were involved in the compensation contracting before the treatment, leading to stronger effects on investment efficiency. I further analyze a broader measure of corporate governance using a firm's overall corporate governance pillar score. I find evidence to support the notion of mandated longer duration having the largest effect on investment efficiency in firms that have low compensation committee independence or a weak overall corporate governance.

In further analyses, I mitigate concerns that different contemporaneous forces at a country level might drive the results, by comparing the treated German firms with a control group of unaffected private German firms. I again find that the treated German firms show greater improvements in investment efficiency than the untreated German firms following the VorstAG. I also attempt to differentiate between reductions in over- and under-investment following the mandated longer duration. Across specifications, it appears that the treatment effect is driven by reductions in over-investment. Longer duration therefore seems to hamper over-investment but may not be effective in providing incentives for under-investing executives to invest more. Finally, I employ a different approach where investment aggregated at the industry level proxies for the likelihood of over- or under-investing (e.g., Biddle et al., 2009; Lara et al., 2016) and find that the treatment also leads to lower abnormal investment in situations where whole industries diverge from optimal investment levels.

This study makes three contributions. First, I contribute to the debate on whether long-term incentive duration can improve firm outcomes. On the one hand, many theorists and empiricists argue for and show benefits of longer duration (e.g., Bebchuk and Fried, 2010; Flammer and Bansal, 2017; Jochem et al., 2018; Larcker, 1983). On the other hand, there are arguments for keeping incentives short-term under certain conditions (e.g., Evans et al., 2017; Schroth, 2018). Recent survey responses even indicate some disagreement between shareholders and supervisory board members, as most shareholders believe that more long-term incentives would lead to better investment decisions, while supervisory board members view incentives to be sufficiently long-term and raise concerns about further lengthening them (Edmans et al., 2021). My study allows an estimate of the effect of an exogenously triggered lengthening in incentive duration and shows that it reduces firms' abnormal investment. I generally interpret this as a positive effect on real investment behavior. However, there seems to be little evidence for reductions in under-investment, which demands caution in claiming universal benefits of longer duration.

Second, I contribute insights into the interplay of corporate governance characteristics and executive pay duration. Pay duration may generally be a substitute or a complement to alternate forms of corporate governance (Gopalan et al., 2014). More specifically, the independence of the supervisory board may play a major role in determining the right incentive pay (e.g., Knyazeva et al., 2013; Li and Wang, 2016) and providing the right monitoring (e.g., Kumar and Sivaramakrishnan, 2008) for optimal investment decisions. My cross-sectional tests indicate that longer duration matters in firms with low compensation committee independence as well as weak overall corporate governance. This rather suggests a complementary relationship between governance and incentive duration.

Third, I provide information for regulators and political decision makers by adding to the growing literature on the consequences of governance regulation in general and on the effects of compensation regulation in particular. There is a contemporaneous debate on the EU level regarding whether and how to reform executives' and supervisory boards' duties towards sustainable corporate governance.[41] My paper adds to this debate

[41] I refer to the European Commission's 'Study on directors' duties and sustainable corporate governance' published in July 2020. This study recommends far-reaching regulatory action and sparks heavy

the notion that mandated longer duration leads on average to more efficient investment decisions, which is in line with shareholders' long-term interests. Analyses suggest that generalization of this notion to other jurisdictions may depend on prevalent governance characteristics, and in particular, on whether there are any independent board members who are responsible for compensation contracting. Moreover, the expected effects in other jurisdictions may depend on whether over-investment is a major issue.

2.2 Research question

Investment efficiency is a key determinant of a firm's economic productivity (Biddle and Hilary, 2006).[42] A mandated increase in executive compensation duration can have either a positive, negative, or no effect on a firm's investment efficiency. Corporate finance theory provides that increasing duration can lead to more efficient investments by mitigating incentives for both empire building and myopic decision making. When the board increases the length of the window during which the executive is evaluated, the future pay-off to current period investment will to a larger extent be realized in observable performance. When an executive enjoys some discretion (Berle and Means, 1932; Jensen and Meckling, 1976), she may have an incentive to grow her empire beyond the optimum size (Jensen, 1986). If she takes on unprofitable projects to increase her span of control, this unprofitability can be evaluated to a greater extent within the lengthened window. At the same time, lower incentives for over-investment require less ex-ante capital rationing by outside suppliers of funds (Stiglitz and Weiss, 1981), which may have led to ex-post under-investment.

Longer duration also counters incentives for inefficient investment that can stem from myopic decision making (Keynes, 1936). Influential survey results show that the majority of US financial executives would pass up a positive NPV investment if initiating the project meant missing current earnings estimates (Graham et al., 2005). The longer evaluation

critique from accounting scholars that urges a more thorough review and analysis of existing governance research before implementing regulatory action (e.g., Roe et al., 2020).

[42]Investment efficiency means that firms undertake only projects and all projects with a positive net present value (NPV) (Modigliani and Miller, 1958). Inefficient investment may relate to over-investment, which corresponds to investing in negative NPV projects, and under-investment, which corresponds to passing up positive NPV opportunities (Biddle et al., 2009).

period reduces the disincentive for current period investments with future payoffs, as these payoffs can be realized within the extended evaluation period (Fudenberg et al., 1990). Likewise, in situations where myopia provides incentives for over-investment to signal a positive outlook for the firm, longer evaluation windows should discipline the executive, as she will have to bear the negative impact of later write-offs of poorly performing projects.[43]

However, mandating changes in compensation design, as in the VorstAG setting, may interfere with efficient contracting. Assuming that firms (i.e., supervisory board members) know best how to compensate their executives, exogenously imposed longer duration may lead to a decrease in efficient investment. It may be optimal for shareholders to implement strong short-term objectives when the executive's short-term stock price manipulation propensity is high (Schroth, 2018). Shareholders may prefer shorter performance periods when executives have valuable alternatives and when firms' operations are highly uncertain (Evans et al., 2017). Moreover, biases towards inefficient investment choices may not be solely eliminated through complete long-term contracting, if investors fail to actively acquire information about executives' activities (von Thadden, 1995). Such monitoring may also be costly, diminishing the returns to increasing duration. As a result, regulation may even distort incentives and reduce investment efficiency.

Finally, if the mandate for longer duration is vague or it is not (anticipated to be) rigorously enforced, there may be no effect on investment efficiency. Firms might simply window dress descriptions of their compensation systems without substantially changing the actual contract design and the resulting incentives. Therefore, whether and in which direction mandated longer incentive duration significantly alters the efficiency of firms' investments ultimately remains an empirical question.

2.3 Regulatory background

The Appropriate Director Compensation Act (VorstAG) came into force in July 2009 as an amendment to the German stock corporation law and prompted listed German

[43]Bebchuk and Stole (1993) show that when the *level* of investment in long-term projects is observable by investors, and only its *productivity* is private knowledge of the executive, she may have incentives to signal a positive long-term outlook of the firm and boost her reputation (Stein, 2003) by heavily investing in long-term projects, leading to over-investment.

firms to review and adapt their remuneration contracts. Prior to the VorstAG, no law demanding any particular incentive scheme was in place in Germany. The previous version of Paragraph 87 of the German stock corporation law merely mandated the supervisory board to implement 'appropriate' executive compensation.[44] The VorstAG is a hard-law regulation that directly affects the structure of executive compensation, while the disclosure of the compensation components had been mandatory since 2006.

The VorstAG act was a targeted response by the German government to the general perception that the global financial crisis had been fuelled by executive contracts that emphasized short-term performance (Deutscher-Bundestag, 2009). Apparently, the regulators presumed that executive compensation contracting had been inefficient in terms of the duration of incentives, demanding regulatory action to prolong them. To this end, the VorstAG act demands that remuneration contracts are oriented towards the long-term. Executives' variable compensation components should have a multi-year assessment period, which may imply that the evaluation spans a multi-year period and/or that a portion of the variable compensation already awarded is deferred for a certain period. Further, the act recommends conditional payback provisions. The most specific requirement demands that the minimum vesting period for stock options is increased from two to four years if firms have such compensation in place.[45] German accounting and law scholars consider the law to have a far-reaching impact on compensation systems (Hitz and Müller-Bloch, 2015).

I rely on two sources of data regarding firms' implementation of the act as stated in contemporaneous annual reports (AR). First, I check a published series on the act's implementation in the 80 firms listed in the two German blue chip indices.[46] Analyzing firms' AR 2009 and AR 2010 (or AR 2008/2009 and AR 2009/2010 if the fiscal year does

[44]Note that the German two-tier system imposes the duty of steering the firm and managing its operations on a public firm's executive board, while the supervisory board appoints, oversees and controls the former, which includes negotiating compensation contracts.

[45]As part of the general goal of orienting executive compensation more in the long term, the act also mandates some corporate governance policies. In particular, it mandates a personal deductible in directors' and officers' liability insurance contracts. Lin et al. (2019) show positive announcement returns regarding the introduction of this aspect. This paper focuses on the aspects of the law that are concerned with the structure of executive compensation.

[46]My final sample of treated German firms is largely a subset of those firms.

not end on December 31), the series finds that among other changes, 49 firms had geared performance-related components towards long-term measures, 38 had lifted the ratio of long-term components to short-term components, 14 had adopted a bonus-malus system, 12 had increased the vesting period for options to four years, and 11 had newly mandated their executives to invest in their own shares (Götz and Friese, 2010, 2011).

Second, I hand-collect AR descriptions of how the German firms included in the final sample implemented the act. My investigation begins with AR 2009 and extends to AR 2010 and AR 2011, until I find a substantial change regarding the remuneration structure. At times, I compare the descriptions of compensation systems with descriptions in earlier documents (i.e., AR 2008) to find apparent differences. Appendix A.2 lists the quoted descriptions exemplary for three firms, the full list is provided in the online Supplement.[47] To group the main changes and any potential concurrent or additional changes, I classify them into six broad categories. As set out in Appendix A.3, it appears that the adoption of a multi-year performance assessment and the deferral of bonus payouts were the major design choices, with a three-year observation window being used most frequently.[48] It further appears that both the later vesting of stocks and options and the requirement to invest in stocks and options were frequent, yet mostly secondary, changes.[49] As it can therefore be reasonably assumed that German firms were exogenously prompted to

[47]For instance, Deutsche Post AG states in AR 2009: '(...) the annual performance-related remuneration will in future no longer be paid in full for the year on the basis of having reached the agreed targets. Instead, 50 % of the annual performance-related remuneration will flow into a new medium-term component with a three-year calculation period (...)'. Similarly, the much smaller company Wincor Nixdorf AG states in AR 2009/2010: 'In accordance with the requirements of Germany's VorstAG Act, the vesting period for share options granted under the 2010 share option program has been extended from two to four years.'

[48]One third of the treatment group (15 firms) adopted multi-year performance assessment as the main change. One fourth (11 firms) deferred a portion of its awarded variable compensation. The most frequent term for both designs is three years. 7 firms switched to more long-term determinants in their bonus systems or increased the long-term to short-term ratio. While the adoption of a bonus-malus system is the explicit main change for only 4 firms, it is likely that other multi-year assessment or deferral schemes also entail such a system.

[49]While only 4 firms increased the vesting of stocks or options as a main change, many firms implemented this change along with a further main change (total frequency of 12). It further appears that a quarter of firms (11) newly required their executives to invest in their own stocks or options, in addition to other changes.

implement longer duration, I regard all German firms as treated.[50] [51] In a later section, I show that the exclusion of the few firms with apparently no - or likely very minor - changes leads to stronger results.

At the very latest, firms had to consider the VorstAG when extending existing executive compensation contracts or signing new ones. Still, many firms immediately switched to new contracts.[52] The published series covering the 80 blue chip firms shows that when analyzing firms' AR 2009, despite substantial changes made by many of them, 35 still stated that their assessment of the compensation system was ongoing or that it was postponed until the contract was due to be rewritten (Götz and Friese, 2010). However, an analysis of the firms' AR 2010 showed that the number of firms had dropped to 11 (Götz and Friese, 2011). I therefore exclude the transition year 2009 from the analysis and expect an effect to materialize for the year 2010, potentially becoming stronger in years 2011 and 2012. I make no distinction between arguably earlier or later effective dates of changes in firms' compensation systems, as the resulting incentives can have an impact on executives long before that date.

I sample several years of archival data for European firms around the act, making the research design potentially sensitive to concurrent and confounding regulatory action taken in any country that is part of the control sample. I check an EU document for such action in the major subset of firms in my control sample that are domiciled in an EU

[50]It is likely that my treatment group includes weakly treated firms, which may impair the detection of significant effects. Trying to differentiate between strongly and weakly treated firms among German firms may endanger the validity of the assigned treatment. Strongly treated firms have an incentive to downplay the impact of the act on compensation contracting so as not to convey that the supervisory board was unable to implement longer duration. Weakly treated firms may have an incentive to describe the changes in the compensation system as more substantial than they were, to meet the expectations of the general public in implementing major revisions.

[51]Flammer and Bansal (2017) exploit close call election outcomes around shareholder proposals on long-term executive compensation to obtain an exogenous variation in compensation contracts. Götz and Friese (2010) sample the degree of approval for a vote on the compensation system during the annual general meeting in the year after the VorstAG act took effect. 70% of the firms covered put their compensation system up for vote. The vote failed in just one firm. There was generally great approval for the systems with most firms' shareholders casting more than 90% of votes in favor. This indicates large support for the compensation systems (many of which already contained substantial changes regarding the incentive duration) in the German firms, making a similar design to Flammer and Bansal (2017) impossible.

[52]For instance, BASF SE states in AR 2009: 'Contracts with all Board members, regardless of existing contractual terms, were consensually and uniformly converted to this system effective as of January 1, 2010.'

country. The EU had released a (nonbinding) recommendation concerning the regulation of executive board remuneration (2009/385/EC in May 2009) only months before the VorstAG. A later European Commission staff working document ('Report on the application by Member States of the EU of the Commission 2009/385/EC Recommendation') samples the recommendation's implementation among EU member states. The document reports that by June 2010, only Germany had implemented legally binding rules regarding the structure of the remuneration policy. Belgium is the only country that also enacted regulations in this regard, yet these regulations are weaker.[53] UK firms make up the largest portion of the control sample. However, there were no reforms in the UK in the period before October 2013, which is outside my sample period.[54] The EU itself did not undertake any further binding compensation regulation before July 2015 (Edmans et al., 2017b), when the Shareholder Rights Directive 2007/36/EC was amended.[55]

2.4 Research design and variable measurement

2.4.1 Sample selection

I compare German and non-German firms with regard to the introduction of the VorstAG act in a difference-in-differences (DID) design. I use all firms that were listed in the STOXX Europe 600 index in July 2009, when the regulation came into force. The German firms in the index are the treated firms and all others are control firms. Many firms in my sample had switched to IFRS in 2005. There had also been changes in executive compensation disclosure requirements in that year. I therefore begin my sample selection with the 2006 financial year. I exclude fiscal year 2009 in the DID regressions, as the regulation became effective during 2009 and may have already affected some firms' investment decisions in that year but also allowed firms to only consider it for new con-

[53]Belgian firms may deviate from the requirement of basing a part of the variable compensation on multi-year performance assessment as well as minimum vesting periods with shareholders' approval (Royal Decree of June 6, 2010). Nevertheless, I exclude Belgian firms from the control group in untabulated robustness tests, which does not alter the results.

[54]In October 2013, the UK mandated the implementation of binding shareholder votes on forward-looking remuneration policies every three years (Petrin, 2015).

[55]The amended directive, which must be transposed into national law by EU member states, mandates that shareholders shall have binding votes on remuneration policies and prescribes the disclosure of several aspects of the remuneration policy.

tracting. Accordingly, the pre-treatment and post-treatment periods span the three years before and after 2009, respectively.

Table 1: Sample selection and distribution

Panel A: Sample selection procedure

Selection step	German	Other	Total
Thomson Reuters Datastream annual data from 2006 to 2012 for firms listed in the STOXX Europe 600 as of July 2009	392	3,808	4,200
− firms from finance and insurance industries	(49)	(868)	(917)
− firms with an 80% decline in equity market value during GFC	(21)	(98)	(119)
− firm-year observations with missing model or control data	(8)	(235)	(243)
Final sample of firm-years for estimating the investment models	314	2,607	2,921
− firm-years for the 2009 transition year	(46)	(385)	(431)
Final sample of firm-years for DID regressions	268	2,222	2,490
Final sample of unique firms	46	394	440
− other country firms that are not matched to a German firm	0	(1,961)	(1,961)
Final PSM sample of firm-years for DID regressions	268	255	523
Final PSM sample of unique firms	46	46	92

Panel B: Firm-year distribution among countries and FF12 industries

Country	Firm-years	FF12 industry	German	Other
Austria	40	(1) Consumer NonDurables	14	221
Belgium	61	(2) Consumer Durables	45	37
Denmark	88	(3) Manufacturing	49	441
Finland	111	(4) Oil, Gas, & Coal Extract. & Products	0	170
France	410	(5) Chemicals & Allied Products	46	127
Germany	314	(6) Business Equipment	27	190
Greece	30	(7) Telephone & Television Transmission	14	185
Ireland	35	(8) Utilities	14	177
Italy	121	(9) Wholesale, Retail, & Some Services	21	239
Luxembourg	26	(10) Healthcare, Medical Equipm., & Drugs	42	208
Norway	51	(11) Finance (excluded)	0	0
Portugal	48	(12) Other	42	612
Spain	155	Total	314	2,607
Sweden	195		2,921	
Switzerland	183			
The Netherlands	169			
United Kingdom	837			
Others	47			
Total	2,921			

I obtain accounting and stock-related data from Thomson Reuters (now Refinitiv) Datastream for the financial years 2006 to 2012. Panel A of Table 1 shows the sample selection procedure. First, I eliminate firms from finance and insurance industries as identified by the Fama and French 12 industry (FF12) classification (French, 2018). Financial firms have very different investment measures compared with other industries; moreover, in many jurisdictions, financial firms were subject to special remuneration regulations in the immediate aftermath of the global financial crisis (GFC).[56] Second, following earlier research based on a sample period covering the GFC (e.g., Ernstberger et al., 2017), I exclude firms with a decline in equity market value greater than 80% from October 2007 (the month with the highest STOXX Europe 600 closing price before the GFC) to March 2009 (the month with the lowest STOXX Europe 600 closing price during the GFC). My inferences remain unchanged when I do not apply this restriction. Third, I eliminate observations (firm-years) in which one or more model or control variables are missing. I arrive at a sample of 2,921 firm-years for the investment model regressions. The sample for the DID regressions contains 2,490 firm-years, of which 268 relate to German firms and 2,222 to others. Both samples correspond to 46 German firms and 394 non-German firms. Table 1 further shows the firm-years for a propensity score matched (PSM) control sample.

Panel B of Table 1 shows the distribution of firm-years among countries and FF12 industries for the full sample from 2006 to 2012 used in the investment model regressions. Firms in the United Kingdom account for the biggest portion of firm-years, followed by French and German firms. In terms of industry classification, the biggest portion of firm-years is attributable to 'Residual category 12 - Other', followed by 'Category 3 - Manufacturing'.

2.4.2 Measuring investment and abnormal investment

I rely on archival accounting and financial market data to assess a firm's investment in a given year. I use two common measures of a firm's investment level. The first investment measure (C/PPE) is the firm's capital expenditures (CAPEX), scaled by the average property plant and equipment balance (PPE) (e.g., Biddle et al., 2009, 2016; McNichols and Stubben, 2008). The second measure defines investment more broadly, as it also accounts for investments in R&D endeavors. I/TA adds R&D expenses to CAPEX, scaled by average total assets (e.g., Balakrishnan et al., 2016; Chen et al., 2013; Lara et al., 2016).

[56] The UK and Switzerland, for instance, implemented reforms applying to financial services firms (Bebchuk and Fried, 2010). Also, Germany implemented the Remuneration Ordinance for Institutions (Institutsvergütungsverordnung) in 2010. Moreover, financial firms receiving funds from bailout or relief programs were required to implement particular caps on compensation.

To do this, I assume missing values for R&D to be zero when CAPEX are available (e.g., Edmans et al., 2017a; Ernstberger et al., 2017). Throughout the paper, I use investment ($Invest$) as an umbrella term to cover the different measures. Abnormal investment ($AbnInvest$) refers to the residuals of different investment model regressions with different investment measures, $|AbnInvest|$ refers to the absolute residuals.

In frictionless markets, the expected benefits of an investment should be the sole factor determining the investment decisions of the firm. Expectations regarding future growth and product demand in turn determine the expected benefits. Growth expectations are based on executives' information sets. The core task of investment models is to proxy those growth expectations. I rely on two well-established models to estimate expected investment: the accelerator model and Tobin's q model.[57]

The accelerator model (e.g., Biddle et al., 2009; Chen et al., 2011; Hsu et al., 2015) builds on the assumption that investment is related to the level of output. When all available capital is fully utilized, changes in demand require adjustments in investment. Investment is modelled as a function of revenue growth, whereby the revenue growth proxies the (future) output growth, i.e., growth in demand. Expected investment is thereby commonly estimated either over the entire cross-section of the economy or on an industry level, where the regression residuals constitute abnormal investment. The implicit assumption is either that investment is efficient on average over the whole economy or in every industry over time. I report the results from estimating the regression models for each FF12 industry; the cross-sectional results are very similar. Equations (1) and (2) show the baseline specification along with a piecewise linear specification of that approach:

$$Invest_{i,t} = \alpha + \beta RevCh_{i,t-1} + \varepsilon_{i,t}, \tag{1}$$

$$Invest_{i,t} = \alpha + \beta_1 RevCh_{i,t-1} + \beta_2 RevChN_{i,t-1} + \beta_3 RevCh * RevChN_{i,t-1} + \varepsilon_{i,t}. \tag{2}$$

Equation (1) regresses firm i's investment in year t on the change in revenue ($RevCh$) from year *t-1* to year t. Chen et al. (2011) suggest a piecewise linear model that allows the

[57]Gao and Yu (2018) present a recent review of different models and their implicit assumptions offered by the investment literature.

relation between revenue growth and investment to differ between increasing and declining revenues. The authors introduce a variable indicating whether the revenue change is negative ($RevChN$ in Equation (2)). As both specifications produce qualitatively and quantitatively similar results, I only report results for the piecewise linear model.

The second approach is Tobin's q model (e.g., Lai et al., 2014; Li et al., 2018; McNichols and Stubben, 2008; Verdi, 2006), which also approximates expected investment by the fitted values of a linear model. This model builds on the assumption that marginal q summarizes a firm's growth opportunities.[58] Thus, I regress investment on the firm's ratio of the market valuation of asset stock to its replacement costs, as in the following model:

$$Invest_{i,t} = \alpha + \beta_1 Q_{i,t-1} + \varepsilon_{i,t}. \tag{3}$$

I measure Q in Equation (3)) as the beginning-of-year enterprise value divided by beginning-of-year total assets. My inferences are not affected when I combine the two proxies for growth opportunities of the two models as suggested by Chen et al. (2013).

Alongside expected growth opportunities, differences in internal financing capabilities may play a major role in explaining firms' investment decisions (e.g., Biddle et al., 2009; McNichols and Stubben, 2008). One way to address this is by adding the firm's operating cash flow to the investment model (e.g., McNichols and Stubben, 2008). I add firm i's cash flow from operating activities in year t scaled by average total assets (CFO) to both models. Equations (4) and (5) show the specifications for the adjusted accelerator and the Tobin's q model as I use them for presenting my empirical results:

$$\begin{gathered} Invest_{i,t} = \alpha + \beta_1 RevCh_{i,t-1} + \beta_2 RevChN_{i,t-1} + \\ \beta_3 RevCh * RevChN_{i,t-1} + \beta_4 CFO_{i,t} + \varepsilon_{i,t}, \end{gathered} \tag{4}$$

$$Invest_{i,t} = \alpha + \beta_1 Q_{i,t-1} + \beta_2 CFO_{i,t} + \varepsilon_{i,t}. \tag{5}$$

The core investment model variables to proxy the opportunity sets are clearly impacted by the GFC. For instance, the average change in revenue plunges in 2009 and recovers

[58] Average q is used to proxy for marginal q, which is unobservable.

in 2010.[59] One might suspect this to impact and potentially bias the empirical tests. The main tests regarding the evolution of investment efficiency rely on a comparison of German firms with other European firms in relative terms. Therefore, there would only be a bias if the extent to which the change in revenue proxies investment opportunities is systematically differently affected by the crisis for German firms relative to other firms. For example, it would be required that the extent to which firms had an incentive to decrease the change in revenue even further in 2009 and to show an even larger increase in the following years, accordingly, systematically differs between treatment and control firms. This seems very unlikely. To further mitigate this concern, I estimate the investment models not only on the basis of the full sample but also on a matched subsample, the full sample adding untreated German firms, and a sample of treated and untreated German firms only in later sections.

2.4.3 Empirical strategy

The absolute value of the residuals of the investment models in (4) and (5) proxy for abnormal investment. This approach is subject to the concerns raised by Chen et al. (2018) when residuals from one regression are the dependent variable of another. Following their advice, I add the first-step regressors to the second-step regression.[60] Equation (6) formalizes the second-step regression for both investment models:

$$\begin{gathered}|AbnInvest|_{i,t} = \alpha + \beta_1 Treat_i * Post_t + \beta_2 Treat_i + \beta_3 Post_t + \\ \beta_4 ResidN_{i,t} + 1^{st} StepVars + Controls + FEs + \varepsilon_{i,t}.\end{gathered} \tag{6}$$

$|AbnInvest|$ is the absolute residual of the respective first-step regression and measures firm i's yearly absolute deviation from the expected investment. $Treat$ indicates whether a firm is German. *Post* indicates whether an observation refers to a fiscal year after the regulation took effect. I code firm-years 2006 to 2008 as 0 and firm-years 2010 to 2012 as 1. The interaction of $Treat$ and *Post* is the variable of interest. A negative coefficient

[59]This pattern is, however, very similar for treated and control firms (not shown).

[60]The economic interpretation of the residuals requires that they be transformed into their absolute values, which may still bias the estimators. To mitigate this issue, I include a binary variable (*ResidN* in Equation (6)) that is coded 1 for positive and 0 for negative residuals of the first model.

indicates a reduction in abnormal investment for treated firms, which means an increase in investment efficiency. The coefficient captures reductions in both over-investment and under-investment. In later tests, I attempt to differentiate between the two.

In a later analysis, I employ a third approach that uses investment aggregated at the industry level as a proxy for the likelihood of over- and under-investing (Biddle et al., 2009; Lara et al., 2016). Untabulated results for a fourth investment model, which was used previously in the context of linking compensation to investment distortion (Eisdorfer et al., 2013), do not affect my inferences.[61]

2.4.4 Control variables

Control variables (*Controls*) fall into three categories: First, I add the model variables from the accelerator approach to the second step of the Tobin's q approach and vice versa, since these variables should explain abnormal investment behavior. Second and third, I add the control variables from Eisdorfer et al. (2013), who also investigate investment efficiency in a compensation context, and those from Chen et al. (2011), who also analyze investment efficiency in a multinational context. A firm's *Size* should be indicative of the level of attention and scrutiny it receives from financial market participants, which should be negatively associated with abnormal investment.[62] I employ Altman's Z score (Z) to proxy for financial distress, which I expect to be positively associated with incentives to make abnormal investment decisions. *Lever* is the firm's leverage. A higher leverage could financially constrain the firm. Thus, I would expect leverage to be negatively associated with the investment level. It is thus unclear whether the association between leverage and abnormal investment is positive or negative. *Regul* is a binary indicator of whether a firm operates in a regulated industry. Abnormal investment should be lower in such firms. *Homog* estimates homogeneity in different industries, whereby I also expect a negative

[61] These authors suggest analyzing the difference in a firm's investment from the same year's median industry investment after controlling for certain variables driving investment levels. They regard the median industry year investment as the expected investment. The absolute difference from the expected investment is the abnormal investment after controlling for leverage, firm size, the investment opportunity set, distress, regulation and industry homogeneity.

[62] I measure *Size* by the number of a firm's employees in a given year instead of other commonly used proxies for firm size, such as total sales or assets, because the firms in my sample report in different currencies.

association with abnormal investment decisions. *Tangb* measures the tangibility of the firm's assets. This control variable captures differences in the importance of tangible assets to the firm in general. Finally, *Slack* is added to control for a firm's cash abundance. As with low-leverage firms, cash-rich firms should be less constrained in financing their investments. Therefore this variable may be positively or negatively associated with abnormal investment.

I provide a description of the calculation of all variables in Appendix A.1. All continuous variables are winsorized yearly on the 1% level. My inferences also hold without winsorization. In line with prior literature in the field that uses a similar DID or pre-post design (e.g., Chen et al., 2013; Li et al., 2018), I control for time-invariant fixed effects between treated and untreated firms via firm fixed effects in my main specification. Thereby, the parameter of interest is akin to a weighted combination of all firms' two-period DID estimators.[63] I cluster standard errors on the country-year level as firms are assigned to the treatment group on the basis of their home country. As an alternative, I present results for industry fixed effects as a higher-level identification of units that requires fewer dummies (e.g., Chen et al., 2011). In this specification, I add year fixed effects to absorb potential effects of a time-varying shock on the industry level that may coincide with the regulatory treatment, and cluster standard errors on the level of the firm.[64]

2.5 Empirical results

2.5.1 Descriptive statistics

Table 2 presents descriptive statistics of the variables included in the regression analyses. Panel A shows investment model data for all firm-years between the financial years 2006 and 2012. The mean investment in terms of CAPEX is 22.80% of average PPE. The mean investment in terms of CAPEX and R&D is 7.12% of average total assets. Panel A shows that the investment models produce more and smaller negative residuals than

[63] This DID with firm dummies is less sensitive to few firms driving the effect, particularly in unbalanced designs where firms drop in and out of the sample.

[64] The inclusion of year dummies makes the DID less sensitive to few years (within the pre- or post-period) driving the effect. The results are unaffected when also adding the year fixed effects to the main specification.

positive ones, which is in line with previous research (Chen et al., 2011).[65] I distinguish between the treatment and control sample in Panel B of Table 2, where I focus on the sample of firm-years used in the DID regressions. Panel B shows that the distribution of the investment variables C/PPE and I/TA is similar across treatment and control firms. While the absolute residuals of the investment models for C/PPE (i.e., $|C/PPE_{Rev}|$ and $|C/PPE_Q|$) also do not differ, the absolute residuals for I/TA do. It appears that German firms have slightly lower $RevCh$ and OCF and substantially lower Q figures throughout the sample period. Panel B in Table 2 further shows that treatment firms are substantially larger ($Size$) and have a less tangible asset base ($Tangb$) than control firms.

Untabulated Bravais Pearson correlations show that the absolute residuals of the two investment models are highly correlated for each investment measure, for instance, the correlation between $|C/PPE_{Rev}|$ and $|C/PPE_Q|$ is 0.93. However, the investment model residuals are only weakly correlated across C/PPE and I/TA (between 0.30 and 0.33).

[65] For instance, 1,705 of C/PPE_{Rev} residuals are negative, with a mean of -7.57, while 1,216 residuals are positive, averaging 10.61.

Table 2: Descriptive statistics

Panel A: Final sample of firm-years for estimating the investment models

	N	Mean	STD	Min	25%	75%	Max
C/PPE	2,921	22.80	14.32	2.40	13.65	28.17	98.70
C/PPE_{Rev}	2,921	0	12.66	−32.20	−7.68	4.60	80.35
$C/PPE_{Rev} < 0$	1,705	−7.57					
$C/PPE_{Rev} > 0$	1,216	10.61					
C/PPE_Q	2,921	0	12.43	−32.92	−7.35	4.48	74.86
$C/PPE_Q < 0$	1,706	−7.41					
$C/PPE_Q > 0$	1,215	10.41					
I/TA	2,921	7.12	5.30	0.25	3.58	9.14	38.79
I/TA_{Rev}	2,921	0	4.55	−11.65	−2.71	1.86	31.93
$I/TA_{Rev} < 0$	1,711	−2.75					
$I/TA_{Rev} > 0$	1,210	3.89					
I/TA_Q	2,921	0	4.59	−18.23	−2.72	1.83	33.71
$I/TA_Q < 0$	1,720	−2.75					
$I/TA_Q > 0$	1,201	3.94					
$RevCh$	2,921	0.09	0.20	−0.51	−0.001	0.15	1.53
$RevChN$	2,921	[0.25]					
Q	2,921	1.37	1.05	0.18	0.73	1.65	8.25
CFO	2,921	0.11	0.07	−0.08	0.07	0.14	0.39

(continued on next page)

(continued from previous page)

Panel B: Final sample of German vs. other firm-years for DID regressions

	German firm-years (N = 268)						Other firm-years (N = 2,222)						Diff
	Mean	STD	Min	25%	75%	Max	Mean	STD	Min	25%	75%	Max	p
C/PPE	23.65	13.91	6.38	15.16	27.82	85.43	23.07	14.37	2.40	13.69	28.92	97.16	0.52
$\|C/PPE_{Rev}\|$	8.99	9.35	0.01	3.16	11.25	59.84	8.80	8.98	0.003	2.99	11.58	80.35	0.75
$\|C/PPE_{Q}\|$	8.50	9.63	0.01	2.25	10.38	60.54	8.67	8.74	0.01	2.95	11.54	74.86	0.75
I/TA	7.37	3.97	0.67	4.78	9.11	24.82	7.15	5.50	0.25	3.46	9.15	38.79	0.43
$\|I/TA_{Rev}\|$	2.48	2.51	0.001	0.76	3.21	16.42	3.30	3.33	0.003	1.10	4.36	31.93	0.00
$\|I/TA_{Q}\|$	2.50	2.43	0.01	0.77	3.30	15.61	3.31	3.38	0.001	1.15	4.26	33.71	0.00
$RevCh$	0.10	0.14	−0.51	0.03	0.14	0.90	0.11	0.20	−0.51	0.01	0.16	1.53	0.22
$RevChN$	[0.14]						[0.21]						0.00
Q	1.06	0.74	0.19	0.59	1.29	4.90	1.48	1.12	0.18	0.81	1.78	8.25	0.00
CFO	0.09	0.06	−0.08	0.06	0.11	0.37	0.11	0.07	−0.08	0.06	0.14	0.39	0.00
$Lever$	0.23	0.13	0.003	0.14	0.32	0.60	0.25	0.16	0.00	0.14	0.36	0.74	0.02
$Size$	82,667	111,418	759	11,994	95,193	469,680	41,347	64,154	175	6,134	50,817	469,680	0.00
Z	2.93	1.87	0.07	1.72	3.48	12.49	3.60	3.16	−0.02	1.85	4.10	22.28	0.00
$Regul$	[0.11]						[0.13]						0.42
$Homog$	27.14	7.83	18.94	20.23	27.98	46.01	29.88	8.78	15.22	24.46	34.40	52.67	0.00
$Tangb$	0.25	0.14	0.04	0.12	0.33	0.73	0.29	0.22	0.01	0.11	0.44	0.89	0.00
$Slack$	0.08	0.07	0.001	0.03	0.11	0.37	0.07	0.07	0.0000	0.02	0.09	0.41	0.01

This table shows descriptive statistics. Panel A shows descriptives for the investment measures and the investment model variables for the final sample of firm-years. The sample period is from 2006 to 2012. Panel B differentiates between German and other firms included in the final sample of firm-years for the DID regressions which exclude the transition year 2009. Panel B shows descriptives for the variables included in the multivariate regression analyses. All variables are defined in Appendix A.1. $|C/PPE_{Rev}|$, for instance, refers to the absolute value of the residuals of investment model regressions based on the accelerator model using CAPEX scaled by average PPE as the investment variable.

2.5.2 Main model results

Regression results

Table 3 shows the results of the accelerator model. $|C/PPE_{Rev}|$ is the dependent variable in the first three columns and $|I/TA_{Rev}|$ is the dependent variable in columns (4) to (6). Columns (1) and (4) ((2) and (5)) show results with firm (industry and year) fixed effects for the full control sample. All estimators for the interaction of $Treat$ and $Post$ are negative and significant. German firms have on average about 2.3 to 2.6% (about 0.5 to 0.9%) lower abnormal investment in terms of C/PPE (I/TA) after treatment. This corresponds to about 10% of the average level of the respective investment variable.[66] I consider this effect economically significant.

The coefficients for $Treat$ and $Post$ are mostly insignificant. The coefficients for $ResidN$ are negative and significant, as the negative residuals of the first-step models are smaller than the positive residuals. The first-step regressors are not tabulated, as they cannot be meaningfully interpreted. The control variable Q is positive but mostly insignificant. There is no clear pattern for $Lever$, as it has insignificant, significantly positive and significantly negative coefficients. $Size$ is mostly negatively correlated with abnormal investment. The distress variable Z is positive for $|C/PPE_{Rev}|$ and negative for $|I/TA_{Rev}|$. There are no clear patterns for industry regulation and homogeneity. $Tangb$ is negatively correlated with abnormal investment for $|C/PPE_{Rev}|$ in the full sample. The coefficient for $Slack$ is mostly insignificant. Overall, the different fixed effects structures produce quite different coefficients for many of the control variables.

In addition to comparing treated German firms with all other firms, I apply a propensity score matching. Panel B of Table 2 shows that German and other firms display significant differences for the majority of the control variables. PSM would address potentially non-linear effects of these controls. I therefore match on these variables' means over the years of the pre-period (2006 to 2008). More specifically, I match on $log(Size)$, $Tangb$, and $Slack$, and include dummies for FF12 industry factors into the matching.[67]

[66]For instance, the interaction coefficient of column (1) in Table 3 (-2.305) divided by the average C/PPE reported in Table 2 (22.80) gives 10.1%.

[67]A three-year-average of $Lever$ and Z makes little sense, so I do not include those variables in the matching. $Regul$ and $Homog$ vary on the level of the industry.

Table 3: Regression analyses for the accelerator approach

	Dependent variable: $\lvert C/PPE_{Rev} \rvert$			*Dependent variable:* $\lvert I/TA_{Rev} \rvert$		
	Main specification	Alternative FEs	PSM subsample	Main specification	Alternative FEs	PSM subsample
	(1)	(2)	(3)	(4)	(5)	(6)
Treat*Post	**-2.305*****	**-2.517****	**-2.620****	**-0.493****	**-0.733***	**-0.882*****
	(0.559)	**(1.045)**	**(0.865)**	**(0.216)**	**(0.403)**	**(0.248)**
Treat		2.665*			0.013	
		(1.361)			(0.381)	
Post	0.454		1.000	−0.146		0.279
	(0.316)		(0.648)	(0.113)		(0.209)
ResidN	−1.995***	−2.867***	−1.553**	−0.656***	−1.252***	−0.498**
	(0.448)	(0.540)	(0.704)	(0.177)	(0.230)	(0.244)
Q	0.512	−0.203	0.598	0.251	0.388**	0.649*
	(0.443)	(0.847)	(1.194)	(0.209)	(0.184)	(0.355)
Lever	−0.383	4.729	6.394*	−3.527***	−0.913	−0.989
	(2.382)	(3.040)	(3.669)	(1.126)	(0.865)	(1.129)
log(Size)	−0.005	−0.972***	−1.342	0.181	−0.482***	0.547
	(0.580)	(0.227)	(1.891)	(0.242)	(0.096)	(0.410)
Z	0.125	0.341	0.379	−0.025	−0.143**	−0.004
	(0.224)	(0.365)	(0.628)	(0.116)	(0.070)	(0.133)
Regul	−9.836	1.815	−44.934	2.561	−0.487	−19.458
	(10.575)	(1.340)	(116.121)	(5.146)	(0.453)	(23.755)
Homog	1.225*	0.131**	1.798	−0.086	−0.050**	0.982
	(0.665)	(0.067)	(5.861)	(0.319)	(0.024)	(1.221)
Tangb	−12.499**	−3.904*	−0.761	1.251	−0.044	2.708
	(5.366)	(2.106)	(5.754)	(1.695)	(0.734)	(2.301)
Slack	2.151	2.516	−0.584	−1.681	−0.273	−2.886*
	(5.111)	(4.064)	(6.176)	(1.278)	(1.475)	(1.699)
Intercept	−18.985	12.663***	−8.765	5.224	7.954***	−27.047
	(15.352)	(3.030)	(119.371)	(7.247)	(1.555)	(27.682)
Firm FE	Yes		Yes	Yes		Yes
FF12 + Year FE		Yes			Yes	
Observations	2,490	2,490	523	2,490	2,490	523
Adjusted R^2	0.455	0.153	0.609	0.566	0.233	0.476

This table presents the results of the treatment effect of longer duration on investment efficiency, in a two-step approach: The first step (not tabulated) estimates the accelerator model regression (4) per FF12 industry. The absolute values of the residuals of this regression (i.e., $\lvert C/PPE_{Rev} \rvert$ for the investment variable CAPEX sclaed by average PPE and $\lvert I/TA_{Rev} \rvert$ for the sum of CAPEX and R&D scaled by average total assets) are the dependent variables in the second-step regression (6). The table shows pooled panel OLS regression coefficients from the second step. The first-step regressors are included in the second step, but not tabulated. The dummies for the respective fixed effects are also included but not tabulated. All variables are defined in Appendix A.1. Standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

I match one of the control firms to each treated firm without replacement and exclude those control firms that do not have adequate counterparts among the treated firms. The mean propensity score distance between the 46 treated firms and the 46 matched control firms is 0.006.

Columns (3) and (6) of Table 3 show the results of the accelerator approach for the PSM sample. The interaction of $Treat$ and $Post$ remains significantly negative. The investment models implicitly assume that investment is efficient on average for all firms in the respective industry in the economy. Therefore, the reported figures in columns (3) and (6) show the results of the respective second step, in which the sample is reduced in line with the matching procedure after the first-step investment model estimations, such that the investment models employ the full sample of the STOXX Europe 600 universe. Nevertheless, in untabulated tests, the results also hold when I estimate both the first step and the second-step regressions after the matching based sample reduction.

Table 4 shows the results of the second regression approach, which follows Tobin's q theory. Again all estimators for the DID interaction are negative, with five of the six being statistically significant. The economic effects have a similar magnitude as before. The negative residuals of the first-step investment model are again smaller than the positive residuals ($ResidN < 0$). The other control variables also appear to have similar effects as in the accelerator model.

Tables 3 and 4 show that the results for C/PPE are generally stronger than those for I/TA. This is not surprising, given that the latter relies on assuming R&D to be zero when the variable is missing. I add a dummy variable for firms where missing R&D is set to zero (Koh and Reeb, 2015) in untabulated tests, which produces quantitatively similar results. In the investment efficiency literature, the I/TA measure is often used net of proceeds from disposals of PPE (e.g., Biddle et al., 2009; Chen et al., 2011; Cheng et al., 2013). In untabulated tests, I obtain similar but weaker results when I net both measures of investment by deducting PPE disposals, which are set to zero when the variable is missing. In all, the results for the two main approaches provide strong support for the notion that mandated longer duration increases investment efficiency.

Table 4: Regression analyses for the Tobin's q approach

	Dependent variable: $\lvert C/PPE_Q \rvert$			*Dependent variable:* $\lvert I/TA_Q \rvert$		
	Main specification	Alternative FEs	PSM sub-sample	Main specification	Alternative FEs	PSM sub-sample
	(1)	(2)	(3)	(4)	(5)	(6)
Treat*Post	**-2.563*****	**-2.839*****	**-3.080*****	**-0.465***	**-0.655**	**-0.925*****
	(0.508)	**(1.065)**	**(0.706)**	**(0.254)**	**(0.407)**	**(0.240)**
Treat		2.528*			0.003	
		(1.421)			(0.381)	
Post	0.349		0.661	−0.051		0.346*
	(0.284)		(0.643)	(0.113)		(0.194)
ResidN	−2.003***	−2.838***	−1.202**	−0.753***	−1.291***	−0.389*
	(0.387)	(0.540)	(0.596)	(0.175)	(0.239)	(0.220)
RevCh	−0.449	−1.296	0.459	−0.202	0.403	0.363
	(1.134)	(1.176)	(1.733)	(0.403)	(0.424)	(0.572)
Lever	2.325	5.252*	13.026**	−2.534**	−0.762	−0.747
	(2.553)	(2.848)	(5.629)	(1.000)	(0.885)	(1.152)
log(Size)	−0.530	−1.038***	−2.732	0.200	−0.512***	1.051***
	(0.624)	(0.228)	(1.941)	(0.224)	(0.101)	(0.295)
Z	0.349	0.242	1.516*	0.038	−0.118	−0.001
	(0.251)	(0.322)	(0.814)	(0.090)	(0.073)	(0.152)
Regul	3.503	1.514	−4.053	5.437	−0.482	−14.608
	(6.905)	(1.374)	(86.591)	(4.879)	(0.512)	(23.735)
Homog	0.376	0.129*	−0.353	−0.207	−0.056**	0.729
	(0.392)	(0.069)	(4.280)	(0.309)	(0.025)	(1.210)
Tangb	−11.722**	−3.687*	−0.731	−0.046	−0.186	1.682
	(5.339)	(2.124)	(5.786)	(1.739)	(0.739)	(2.651)
Slack	0.584	2.814	−0.759	−0.727	0.323	0.223
	(5.171)	(3.905)	(6.416)	(1.429)	(1.466)	(1.518)
Intercept	1.506	12.979***	52.665	6.927	8.300***	−27.200
	(11.450)	(3.118)	(90.759)	(6.872)	(1.632)	(26.535)
Firm FE	Yes		Yes	Yes		Yes
FF12 + Year FE		Yes			Yes	
Observations	2,490	2,490	523	2,490	2,490	523
Adjusted R^2	0.481	0.161	0.640	0.594	0.241	0.495

This table presents the results of the treatment effect of longer duration on investment efficiency, in a two-step approach: The first step (not tabulated) estimates the Tobin's q model regression (5) per FF12 industry. The absolute values of the residuals of this regression (i.e., $\lvert C/PPE_Q \rvert$ for the investment variable CAPEX sclaed by average PPE and $\lvert I/TA_Q \rvert$ for the sum of CAPEX and R&D scaled by average total assets) are the dependent variables in the second step-regression (6). The table shows pooled panel OLS regression coefficients from the second step. The first-step regressors are included in the second step, but not tabulated. The dummies for the respective fixed effects are also included but not tabulated. All variables are defined in Appendix A.1. Standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Parallel trends

A key assumption underlying my identification strategy is that German and other firms would have had parallel trends in abnormal investment had it not been for the VorstAG act. I conduct a parallel trends test and visualize the treatment effect by year over the sample period. I run Equation (6) by replacing the single *Treat*Post* interaction with separate interactions between $Treat$ and binary indicators for each year. I exclude the indicator for year 2008 immediately prior to the VorstAG, such that 2008 is the benchmark period (e.g., Christensen et al., 2017). I run this test with the full control sample and present results with the industry and year fixed effect specification that generally produces larger standard errors.

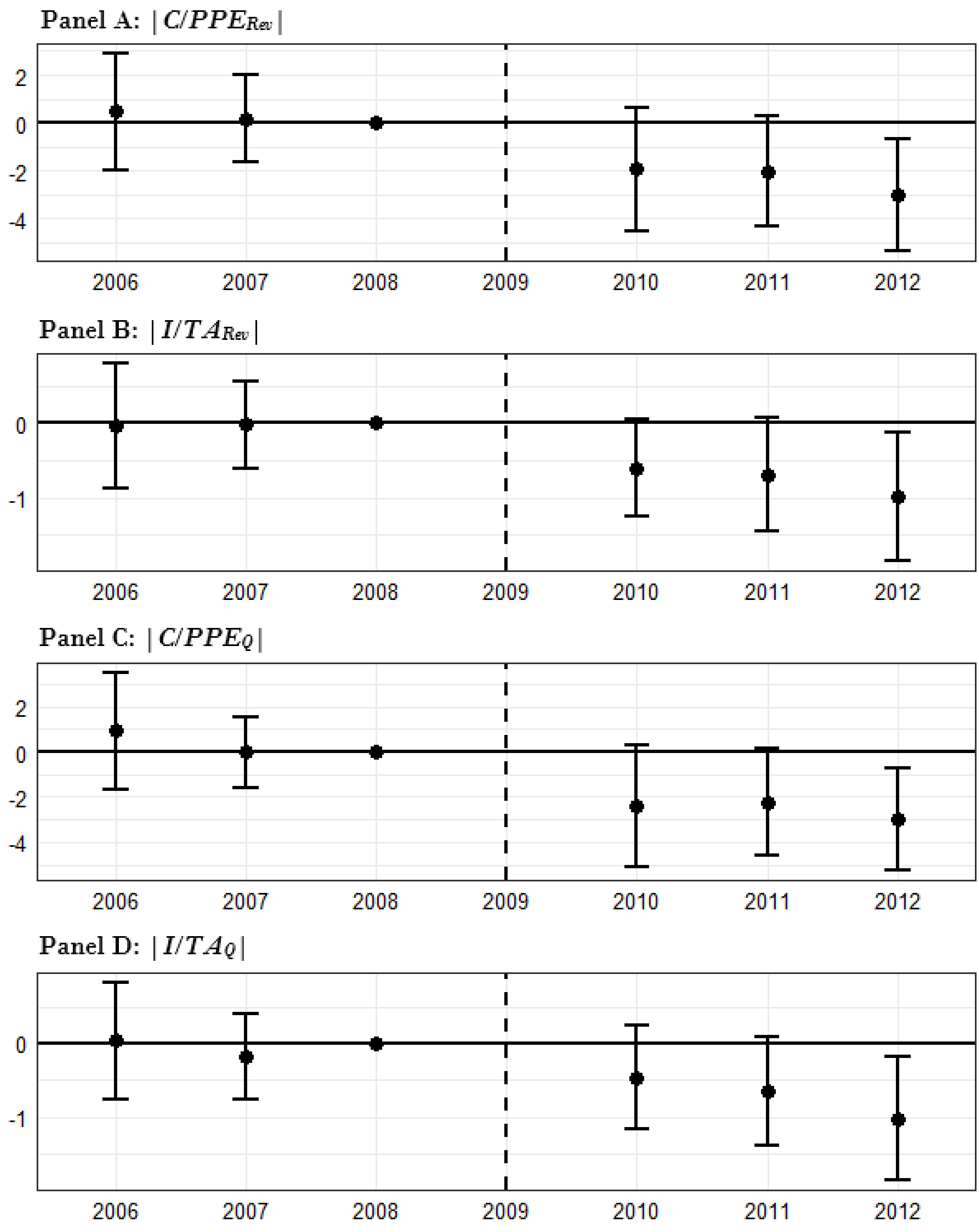
Figure 1 displays the regression coefficient estimates and their two-tailed 90% confidence intervals. Panels A and B map the coefficients for C/PPE and I/TA following the accelerator model. Panels C and D show results for both investment measures and the Tobin's q approach. The counter-factual treatment effects in the pre-treatment periods are small and statistically indistinguishable from the benchmark period in all four panels. The plots indicate that treated and control firms experienced common trends in abnormal investment levels in the period before the introduction of the VorstAG. This supports the assumption that, in the absence of the VorstAG, there would have been parallel trends in the post-period. Figure 1 also indicates that the treatment effects begin in year 2010 and increase until year 2012, lending additional support to the claim that longer duration reduces abnormal investment and increases investment efficiency.[68]

Removing firms with likely very minor changes

The hand-collected descriptions of changes to firms' compensation systems given in the annual reports (provided in the online Supplement) indicate that almost all firms implemented substantially longer duration following the VorstAG act. Five firms apparently

[68]Mapping the average dependent variable per treatment and control group for the years 2006 to 2012 (Appendix A.5) presents a similar picture of common trends in the pre-period, which would have continued in the post-period without the VorstAG intervention. It appears that the average treatment effect is already slightly apparent in 2009 and materializes fully in years 2010, 2011 and 2012. This is in line with the notion that some firms' investment decisions were already affected in 2009, when the VorstAG came into force, and that it took some time for all firms to implement the changes to their compensation systems and for the systems to affect executive decisions.

Figure 1: Pattern of the counter-factual treatment effects



This figure plots regression coefficient estimates and two-tailed 90% confidence intervals for a parallel trends test to visualize the treatment effect by year over the sample period. The DID interaction (*Treat*Post*) of the main tests is replaced with separate interactions between $Treat$ and binary indicators for each year. The indicator for year 2008 immediately before the VorstAG is excluded, such that 2008 is the benchmark period. Panels A and B map the coefficients for $|C/PPE_{Rev}|$ and $|I/TA_{Rev}|$, which are the absolute values of the regression residuals based on the accelerator model using the two alternative investment measures CAPEX scaled by average PPE and the sum of CAPEX and R&D scaled by average total assets. Panels C and D show results for both investment measures and the alternative Tobin's q approach (i.e., $|C/PPE_Q|$ and $|I/TA_Q|$).

made no[69] or likely very minor changes[70] to their compensation systems. In untabulated tests, I obtain larger effect sizes when I remove these five firms from the treatment sample. The effect sizes of the main analyses reported in Table 3 and Table 4 are about 0.5 larger for C/PPE and about 0.1 larger for I/TA across all models and for both the full and the PSM control sample.

2.5.3 Cross-sectional analyses

Next I analyze cross-sectionally whether the effect of increased incentive duration on investment efficiency is related to corporate governance characteristics. In particular, I test whether the reduced abnormal investment depends on how independent the firm's compensation committee had been in the pre-period. A substantial body of research documents positive effects of supervisory board independence on executive compensation contracting (e.g., Knyazeva et al., 2013; Kumar and Sivaramakrishnan, 2008; Li and Wang, 2016). One could expect that firms with fewer independent supervisory board members involved in the design of the executive compensation system likely leads to weaker incentives for efficient investment. In such firms, the VorstAG act should impose a stronger change to the incentive system for executives, which would then have a stronger effect on investment efficiency. There is tension, however, as monitoring by more independent boards may decrease even if they install better compensation incentives (Kumar and Sivaramakrishnan, 2008), which may allow more abnormal investment.

I obtain data on the degree of independence of the compensation committee from the executive board from the asset-4 data in Thomson Reuters Datastream and obtain it manually for German firms where it is missing.[71] I first test the cross-sectional effects

[69] Software AG states in AR 2009 that the system 'is geared in an exemplary fashion toward long-term company development' and shows no changes in later years. United Internet AG neither mentions the VorstAG in any of its reports, nor shows any changes throughout AR 2008 to AR 2012).

[70] Fresenius SE subjects the new provision 'that the share of long-term variable compensation components is at least equal in its amount to half of the total variable compensation components' to the discretion of the supervisory board (AR 2009). Symrise AG stated that 'against the background of the (...) [VorstAG,] (...) the remuneration is justified in both the internal vertical and external horizontal comparison' in AR 2009 and describes in AR 2011 that only the benchmark of the long-term incentive plan had been changed. Tognum AG states in AR 2009 that the 'remuneration already meets the objectives of the new legislation to a large degree' and presents a 'new market-based target remuneration structure' that contains the same components with likely similar weighting as before in AR 2010.

[71] The variable calculates the percentage of independent members in the compensation committee (i.e.,

within the sample of German treated firms. The effect of low compensation committee independence might be particularly strong when there are no or very few independent members and the effect among higher degrees of independence is likely not linear. I therefore define low compensation committee independence (CCI) to be 1 for firms that are in the lowest quartile for the score in 2007 of the pre-period and 0 otherwise. Panel A in Table 5 shows the results for the four different specifications regarding investment measures and models. All four interactions between CCI and the $Post$ dummy are negative, with three of them being significant, indicating that the reduction in abnormal investment is driven by those firms that had a low degree of compensation committee independence prior to the mandate to lengthen the duration.

'personnel committee', 'human resources committee', 'executive committee', 'presiding committee') as stipulated in firms' annual reports. I obtain members' independence (binary variable) on the basis of the Thomson Reuters Datastream methodology, pretests for German firms where the variable is available, and in line with bright-line criteria as they are used today and were used at the time in listing requirements of major stock indices (i.e., NYSE and Nasdaq). A member qualifies as independent if she was not employed by the company in the two years prior to joining the supervisory board, does not have a shareholding of more than 5%, does not depend on substantial executive compensation (i.e., the compensation beyond membership and committee fees paid as a variable compensation or in return for advisory contracts between the member and the firm should not exceed 100 thousand Euros (i.e., NYSE and Nasdaq use a threshold of 120 thousand U.S. dollars)), and has not been a member of the board for more than 10 years. I check firms' annual reports and compensation committee members' BoardEx profiles to assess independence.

Table 5: Cross-sectional analyses of low compensation committee independence

Panel A: Treated German firms

Dependent variable:	$\lvert C/PPE_{Rev}\rvert$	$\lvert C/PPE_{Q}\rvert$	$\lvert I/TA_{Rev}\rvert$	$\lvert I/TA_{Q}\rvert$
	(1)	(2)	(3)	(4)
CCI*Post	**-2.572*****	**-1.415**	**-0.857*****	**-0.922*****
	(0.701)	**(1.034)**	**(0.319)**	**(0.282)**
Post	−0.127	−1.003	−0.103	−0.092
	(0.778)	(0.744)	(0.177)	(0.143)
ResidN	−2.458**	−1.786**	−0.736	−0.466
	(1.035)	(0.702)	(0.517)	(0.397)
Firm FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	268	268	268	268
Adjusted R^2	0.667	0.678	0.470	0.463

Panel B: Full sample

Dependent variable:	$\lvert C/PPE_{Rev}\rvert$	$\lvert C/PPE_{Q}\rvert$	$\lvert I/TA_{Rev}\rvert$	$\lvert I/TA_{Q}\rvert$
	(1)	(2)	(3)	(4)
Treat*Post*CCI	**-4.791*****	**-4.082*****	**-0.623**	**-0.823****
	(1.071)	**(1.125)**	**(0.409)**	**(0.369)**
Treat*Post	**-1.475*****	**-1.998*****	**-0.239**	**-0.194**
	(0.540)	**(0.661)**	**(0.165)**	**(0.209)**
$Post * CCI$	2.343***	2.525***	−0.399	−0.219
	(0.869)	(0.865)	(0.288)	(0.247)
Post	0.352	0.256	−0.089	−0.027
	(0.325)	(0.277)	(0.119)	(0.114)
ResidN	−1.733***	−1.591***	−0.582***	−0.704***
	(0.424)	(0.345)	(0.175)	(0.169)
Firm FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	2,199	2,199	2,199	2,199
Adjusted R^2	0.450	0.482	0.598	0.624

This table presents the results of low compensation committee independence in the pre-period regarding the effect of longer duration on investment efficiency. The table shows pooled panel OLS regression coefficients from the second step of the two-step approach used before. Panel A focuses on the treated German firms. Panel B extends the analyses to the full sample and shows the incremental treatment effect for a low level of compensation committee independence in the pre-period. In both panels, the four columns correspond to the combinations of the two alternative investment measures with the two alternative investment models. First-step investment model regressors, other control variables, intercept, and fixed effect dummies are not tabulated. All variables are defined in Appendix A.1. Standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Second, I extend the analysis to the full sample, including a three-way interaction between $Treat$, $Post$, and CCI, and present the results in Panel B of Table 5.[72] The three-way interaction is negative in all four specifications and significant in three of them, indicating that German firms with a low degree of independence reduce their abnormal investment more than European firms with low independence after the treatment. The DID effect (*Treat*Post*) continues to be significantly negative in two specifications. This indicates that there is a treatment effect in firms with low independence, which is incremental to the general treatment effect.

I further analyze a broad measure of firms' corporate governance using the overall corporate governance pillar score from the same database.[73] In line with the discussion by Gopalan et al. (2014) that pay duration can be a substitute or a complement to alternate forms of corporate governance such as monitoring, there is tension as to whether or not the treatment effect depends on aggregated corporate governance characteristics. I also define *Gov* to be 1 for firms that are in the lowest quartile of German firms' 2007 score and 0 otherwise. Table 6, Panel A shows the results for German firms for which the variable is available and Panel B extends the analysis to the full sample. The significantly negative coefficient in three of the four specifications regarding the interaction of *Gov* and *Post* indicates that the reduction in abnormal investment is driven by firms with generally weak corporate governance in the pre-period. In Panel B, the three-way interaction is significant for the two C/PPE specifications, while the *Treat*Post* interaction is significant in all four. This again provides some support for an incremental treatment effect on investment efficiency for firms with particularly low overall corporate governance.

[72]The results in Panel B of Table 5 are based on a reduced sample of 2,199 firm-years for which data on compensation committee independence is available. As before, I cluster standard errors on the country-year level in Panel B, which reduces to year-level clusters in Panel A, which focuses on German firms.

[73]The corporate governance pillar score aggregates 35 metrics related to the management category (i.e., executive board structure and compensation characteristics), 12 shareholder category metrics (i.e., shareholder rights and the use of takeover defense mechanisms), and 9 corporate social responsibility metrics into a single score.

Table 6: Cross-sectional analyses of low corporate governance pillar score

Panel A: Treated German firms

Dependent variable:	$\lvert C/PPE_{Rev}\rvert$	$\lvert C/PPE_{Q}\rvert$	$\lvert I/TA_{Rev}\rvert$	$\lvert I/TA_{Q}\rvert$
	(1)	(2)	(3)	(4)
Gov*Post	**-2.857****	**-2.361**	**-0.810****	**-0.661*****
	(1.495)	**(1.512)**	**(0.362)**	**(0.233)**
Post	−0.333	−0.922***	−0.201	−0.232
	(0.627)	(0.280)	(0.191)	(0.222)
ResidN	−2.273**	−1.727***	−0.854	−0.640
	(1.078)	(0.659)	(0.545)	(0.465)
Firm FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	259	259	259	259
Adjusted R^2	0.671	0.678	0.481	0.466

Panel B: Full sample

Dependent variable:	$\lvert C/PPE_{Rev}\rvert$	$\lvert C/PPE_{Q}\rvert$	$\lvert I/TA_{Rev}\rvert$	$\lvert I/TA_{Q}\rvert$
	(1)	(2)	(3)	(4)
Treat*Post*Gov	**-3.846*****	**-3.777*****	**-0.281**	**-0.027**
	(1.308)	**(1.328)**	**(0.421)**	**(0.366)**
Treat*Post	**-1.370****	**-1.579*****	**-0.475****	**-0.477***
	(0.642)	**(0.591)**	**(0.217)**	**(0.276)**
$Post * Gov$	1.084*	1.200*	−0.484**	−0.594***
	(0.658)	(0.650)	(0.229)	(0.201)
Post	0.203	0.114	−0.001	0.104
	(0.393)	(0.348)	(0.133)	(0.125)
ResidN	−1.826***	−1.798***	−0.608***	−0.727***
	(0.450)	(0.374)	(0.183)	(0.188)
Firm FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	2,316	2,316	2,316	2,316
Adjusted R^2	0.443	0.476	0.585	0.611

This table presents the results of low overall corporate governance in the pre-period regarding the effect of longer duration on investment efficiency. The table shows pooled panel OLS regression coefficients from the second step of the two-step approach used before. Panel A focuses on the treated German firms. Panel B extends the analyses to the full sample and shows the incremental treatment effect for low corporate governance in the pre-period. In both panels, the four columns correspond to the combinations of the two alternative investment measures with the two alternative investment models. First-step investment model regressors, other control variables, intercept, and fixed effect dummies are not tabulated. All variables are defined in Appendix A.1. Standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

2.5.4 Further analyses

German control group of private firms

My main tests cannot fully control for contemporaneous country-level forces that may have affected firms' investment efficiency. It could be that German firms increased their investment efficiency (following the GFC) compared to their European peers, for reasons other than or in addition to the VorstAG treatment. For instance, in 2009, the German government offered several forms of financial aid to stabilize the economy. This may have forced firms to directly address inefficient investment to make them eligible for governmental support. Firms may have been forced to direct funds from former over-investment to support the current workforce and other stakeholders. Also, in the aftermath of the GFC, the public outcry and subsequent scrutiny regarding inefficient executive behavior may have been stronger in Germany than in other countries, which may have disciplined German executives regardless of the change in compensation incentives.

To rule out this concern, I obtain an alternative control group of untreated private German firms. I sample firms from Bureau van Dijk's Dafne database (e.g., Engel and Middendorf, 2009) with a minimum of 5,000 employees. I ascertain that the firms in this sample have a legal form other than a stock corporation, as the VorstAG was implemented as an amendment to the German stock corporation law and there exist some privately owned stock corporations that are not publicly listed. I retain only those firms for which there is at least some data in the pre- and post-periods. I further remove firms owned by a group, the government (e.g., major municipal utilities companies), or the church (e.g., church run healthcare companies). Appendix A.4 lists the names of the 32 firms included in this control group. The control group comprises 160 firm-years, after excluding the transition year, which results in 428 firm-years, after adding the 268 treated German firm-years.

I focus on the investment variable CAPEX scaled by average PPE, as R&D expenses are only available for two firms. I employ the accelerator model as before, because the Tobin's q model requires stock market data. Table 7 shows the results of the second step of comparing treated German firms with the private German control group across the same sample period as before. I use two different strategies to estimate the investment model for this test. First, I add the private German firms to the full sample of European firms (columns (1) and (2)). Second, I estimate the investment model for German firms only (columns (3) and (4)). I include all control variables as before with the exception of Altman's Z score, which requires market data.[74] The DID interaction is negative in all four specifications and significant in three. The effect sizes are comparable to the figures in the main sample. The distribution of the coefficients of the control variables shows some more diverse and extreme values, as this sample exhibits more variation in these firm characteristics. I also obtain significant results when I do not include the control variables.

It could still be, however, that the potential mechanisms in Germany described above simply affected the listed firms more strongly than the private firms, as the public outcry and scrutiny may have focused on the more high-profile listed firms. I would interpret private German firms exhibiting lower levels of abnormal investment when compared with the other European firms as indicative of this conjecture. Untabulated tests show that when comparing private German firms with other European firms, the DID coefficient is insignificantly positive (close to zero), therefore providing no support for this concern. I view the results for this alternative control group as providing strong additional support for the claim that longer duration of executive compensation fosters investment efficiency.

Over-investment vs. under-investment

I further attempt to assess which reductions are larger: those in over-investment or those in under-investment. On the one hand, it may be that mandated longer duration more easily prevents empire builders from wasting resources (Jensen, 1986) than it can encourage executives with a tendency to under-invest to invest at the optimal level. It may be very effective to force an over-investing executive to internalize later negative effects of his investment decisions via long-term compensation contracts. At the same time, long-term incentives may be rather ineffective in incentivizing a rather conservative

[74] I redefine *Lever* as total liabilities over total assets, because total debt is not available from Dafne. I winsorize the continuous variables for the private German firms on the 2.5% level. The specifications employing firm fixed effects (columns (1) and (3)) show yearly clustered standard errors, as the country-year clusters used before do not apply to this single-country sample.

Table 7: Regression analyses for the German control group

	Dependent variable: $\lvert C/PPE_{Rev}\rvert$			
	Investment model: Full sample		*Investment model: German sample*	
	Main specification	Alternative FEs	Main specification	Alternative FEs
	(1)	(2)	(3)	(4)
Treat*Post	**-2.110****	**-2.088**	**-3.731*****	**-3.321***
	(0.871)	**(2.009)**	**(1.352)**	**(1.725)**
Treat		−0.448		1.386
		(2.061)		(2.031)
Post	1.221		3.208**	
	(1.155)		(1.612)	
ResidN	−2.430***	−5.329***	−1.614**	−4.031***
	(0.761)	(1.533)	(0.679)	(1.270)
Lever	−9.591	6.003	4.552	4.427
	(6.232)	(5.688)	(5.956)	(4.833)
log(Size)	−3.503**	0.086	−1.226	0.188
	(1.411)	(0.429)	(2.545)	(0.450)
Regul	29.763	−5.128*	13.662	−5.428**
	(23.303)	(2.942)	(20.803)	(2.636)
Homog	−1.804	0.937**	−1.006	0.862**
	(1.321)	(0.426)	(1.212)	(0.350)
Tangb	−8.635	−14.583***	−2.443	−12.543***
	(7.507)	(5.077)	(5.144)	(4.653)
Slack	−1.207	19.241*	−7.032	10.528
	(10.087)	(9.858)	(9.042)	(7.464)
Intercept	100.460**	−11.702	41.018	−14.238
	(44.701)	(12.976)	(53.734)	(10.680)
Firm FE	Yes		Yes	
FF12 + Year FE		Yes		Yes
Observations	428	428	428	428
Adjusted R^2	0.542	0.354	0.520	0.325

This table presents the results of the effect of longer duration on investment efficiency for an alternative control sample of untreated German firms. The table shows pooled panel OLS regression coefficients from the second step of the two-step approach used before. The dependent variable in all columns is $\lvert C/PPE_{Rev}\rvert$, which is the absolute residual of regressions with C/PPE as investment measure based on the accelerator model. In columns (1) and (2), the first step investment regression is estimated on the main sample (including other European firms) supplemented with the untreated German firms. Alternatively, in columns (3) and (4), the first step investment regression is estimated on the German firms only. The respective first-step regressors are included in the second step, but not tabulated. The dummies for the respective fixed effects are also included but not tabulated. All variables are defined in Appendix A.1. Standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

executive to spend more capital on long-term projects, especially when career concerns (Holmstrom and Costa, 1986) dominate. Under this line of argument one could expect greater reductions in over-investment compared to reductions in under-investment.

On the other hand, the tendency to spend too many resources appears to be a stronger signal of some executive power (Bebchuk and Fried, 2003; Jensen and Meckling, 1976). Powerful executives may dominate the contract-writing process regardless of regulatory requirements for long-term duration, rendering the effect of the VorstAG weak. This would result in a smaller reduction in over-investment than in under-investment. It is possible that reductions in over- or under-investment are also a function of which behavior predominates, on average, among the sample firms.

The empirical literature on investment efficiency does not yet provide a framework for investigating the relative likelihood or magnitude of over- and under-investment. I run several tests concerning this question. First, I repeat the regressions of the main models and include the three-way interaction between the DID variables and the dummy that is coded 0 for positive and 1 for negative residuals of the respective investment model (*Treat*Post*ResidN*). This approach measures whether the treatment effect is different for over- or under-investing firm-years.

Table 8 shows the results for the four different combinations of investment measures and models, $|C/PPE_{Rev}|$, $|C/PPE_{Q}|$, $|I/TA_{Rev}|$, and $|I/TA_{Q}|$. Panel A reports the regressions for the full control sample and Panel B shows results for the PSM sample. I report results for firm fixed effects, the combination of industry and year fixed effects produces very similar results. The three-way interaction coefficients are positive and generally significant. All DID coefficients are negative and significant. The latter measure the reduction in average over-investment magnitude following treatment. For the reduction in average under-investment magnitude, it is necessary to total the two coefficients of each column. For instance, in the Tobin's q model with the full control sample (column (2) of Panel A), the coefficient for the reduction in the magnitude of over-investment is 4.469, whereas the coefficient for the reduction in the magnitude of under-investment is 0.657 (-4.469 + 3.812 = -0.657). Scaling these coefficients with the signed means of the first-step residuals results in 43% (4.469/10.41) for the reduction in over-investment and 9% (0.657/7.41) for the reduction in under-investment. For columns (1) and (4) of Panel A and columns (1) and (2) of Panel B, the coefficients even indicate an increase in the magnitude of under-investment.[75] This analysis shows that the VorstAG treatment has a larger effect on the average magnitude of over-investment.

The results in Table 8 would also be in line with slight over-investment becoming more frequent after treatment, depressing the average magnitude. Therefore, in an alternative approach, I redefine *ResidN-Pre* as 1 if the average investment residual of a firm in the pre-period is negative and 0 otherwise, to assess the relative treatment effect on former over-investors relative to former under-investors. The results tabulated in Appendix A.6 indicate that the treatment effect on the size of the residuals for former over-investors is larger than for former under-investors.[76] I draw the same inference when using signed residuals as the dependent variable. Finally, I draw the same inferences when looking at extreme values of former over- and under-investment (i.e., above the third quarter and below the first quarter of the average investment residual in the pre-period) and when looking at the last year before the treatment (i.e., 2008) only. I am therefore inclined to conclude that the mandated longer duration hampers over-investment more than it reduces under-investment.

[75] As Eisdorfer et al. (2013) find that executives under-invest when debt-like compensation is predominant, it may be inferred that the VorstAG act led to some emphasis on debt-like compensation components.

[76] For instance, column (2) in Panel A of Appendix A.6 shows that the coefficient for the reduction in abnormal investment of former over-investors is 4.310, while it is 0.432 for former under-investors (-4.310 + 3.878 = -0.432).

Table 8: Differentiating between over-investment and under-investment

Panel A: Full sample

Dependent variable:	$\lvert C/PPE_{Rev}\rvert$	$\lvert C/PPE_{Q}\rvert$	$\lvert I/TA_{Rev}\rvert$	$\lvert I/TA_{Q}\rvert$
	(1)	(2)	(3)	(4)
Treat*Post*ResidN	**6.308*****	**3.812****	**0.658**	**0.930****
	(1.341)	**(1.684)**	**(0.522)**	**(0.374)**
Treat*Post	**-5.940*****	**-4.469*****	**-0.720****	**-0.829****
	(1.155)	**(1.218)**	**(0.349)**	**(0.344)**
$Treat * ResidN$	-2.719^{**}	-1.781	-0.195	-0.067
	(0.895)	(1.097)	(0.348)	(0.346)
$Post * ResidN$	0.370	0.526	0.832^{***}	0.749^{***}
	(0.754)	(0.678)	(0.229)	(0.177)
$Post$	0.227	0.035	-0.631^{***}	-0.487^{***}
	(0.641)	(0.590)	(0.199)	(0.177)
$ResidN$	-2.222^{***}	-2.282^{***}	-1.068^{***}	-1.143^{***}
	(0.628)	(0.583)	(0.228)	(0.193)
Firm FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	2,490	2,490	2,490	2,490
Adjusted R^2	0.458	0.482	0.570	0.598

Panel B: PSM subsample

Dependent variable:	$\lvert C/PPE_{Rev}\rvert$	$\lvert C/PPE_{Q}\rvert$	$\lvert I/TA_{Rev}\rvert$	$\lvert I/TA_{Q}\rvert$
	(1)	(2)	(3)	(4)
Treat*Post*ResidN	**5.414*****	**3.431***	**0.883***	**0.963***
	(1.857)	**(2.052)**	**(0.528)**	**(0.502)**
Treat*Post	**-5.719*****	**-4.687*****	**-1.176*****	**-1.188*****
	(1.600)	**(1.460)**	**(0.349)**	**(0.422)**
$Treat * ResidN$	-3.286^{*}	-3.284^{**}	-0.777^{**}	-0.927^{**}
	(1.716)	(1.523)	(0.360)	(0.391)
$Post * ResidN$	1.657	0.979	0.516	0.649^{*}
	(1.574)	(1.527)	(0.316)	(0.382)
$Post$	-0.083	0.090	-0.080	-0.123
	(1.301)	(1.278)	(0.314)	(0.364)
$ResidN$	-2.128	-0.954	-0.602^{**}	-0.492^{*}
	(1.564)	(1.308)	(0.279)	(0.280)
Firm FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	523	523	523	523
Adjusted R^2	0.630	0.648	0.488	0.514

This table presents the results of the treatment effect of longer duration on over-investment and under-investment. *ResidN* is coded 0 for positive and 1 for negative residuals of the respective investment model. The coefficient of the DID interaction (i.e., the second line) measures the reduction in average over-investment magnitude following treatment. The total of the coefficient of the three-way interaction (i.e., the first line) and the DID interaction measures the reduction in average under-investment magnitude following treatment. In both panels, the four columns correspond to the combinations of the two alternative investment measures with the two alternative investment models. First-step investment model regressors, other control variables, intercept, and fixed effect dummies are not tabulated. All variables are defined in Appendix A.1. Standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Investment aggregated at industry level

My main tests assume that investment is efficient on the industry level over time. However, at times, whole industries may be over- or under-investing. In such times, it is more efficient to invest counter cyclically to one's industry peers. Therefore, in a further robustness analysis, I use investment aggregated at the industry level to proxy for the likelihood of over- and under-investing. I measure the industry-year average of the two investment measures and the independent variables of the investment model variables from before. I follow Lara et al. (2016) and Biddle et al. (2009) and use the accelerator model in this approach. I estimate Equation (7) cross-industrially for the financial years 2006 to 2012.

$$Invest_{Ind,t} = \alpha + \beta_1 RevCh_{Ind,t-1} + \beta_2 RevChN_{Ind,t-1} + \beta_3 RevCh * RevChN_{Ind,t-1} + \varepsilon_{Ind,t} \tag{7}$$

I rank the residuals from this regression into deciles, which I rescale from 0 to 1 to form estimates of aggregate over-investment for the industry-years ($OverInd$).[77] $OverInd$ indicates situations in which firms are likely to over-invest because of general over-investment throughout the industry. I then estimate Equation (8), including the interaction of $OverInd$ with the DID dummies, all investment model and control variables from before, and fixed effects on the firm or industry and year level. I again exclude the transition year 2009 from this regression. A negative coefficient on the three-way interaction between $Treat$, $Post$ and $OverInd$ would imply a lower level of investment in situations prone to over-investment for treated firms in the post-period.

$$\begin{gathered} Invest_{i,t} = \alpha + \beta_1 Treat_i * Post_t * OverInd_{Ind,t} + \beta_2 Treat_i * Post_t + \\ \beta_3 Treat_i * OverInd_{Ind,t} + \beta_4 Post_t * OverInd_{Ind,t} + \beta_5 Treat_i + \\ \beta_6 Post_t + \beta_7 OverInd_{Ind,t} + Controls + FEs + \varepsilon_{i,t} \end{gathered} \tag{8}$$

Table 9 presents the results of this approach. Columns (1) to (3) ((4) to (6)) correspond to investment measure C/PPE (I/TA). Columns (1), (2), (4) and (5) show the results for the full control sample with different combinations of fixed effects. Columns (3) and (6)

[77] I obtain similar results when ranking the residuals into quartiles or yearly quartiles.

Table 9: Regression analyses for the mean industry investment approach

	Dependent variable: C/PPE			*Dependent variable: I/TA*		
	Main specification	Alternative FEs	PSM subsample	Main specification	Alternative FEs	PSM subsample
	(1)	(2)	(3)	(4)	(5)	(6)
Treat*Post*OverInd	**-5.897*****	**-9.380***	**-3.307**	**-1.781*****	**-3.140***	**-2.396***
	(2.195)	**(5.157)**	**(3.727)**	**(0.634)**	**(1.611)**	**(0.791)**
$Treat * Post$	3.103	2.273	1.000	0.871**	1.233*	0.701
	(1.961)	(2.550)	(2.400)	(0.388)	(0.722)	(0.449)
$Treat * OverInd$	11.905***	4.561	12.459***	3.014***	−0.395	4.078***
	(3.011)	(6.781)	(3.961)	(0.770)	(1.821)	(0.805)
$Post * OverInd$	2.353*	0.358	0.486	0.366	0.243	1.505**
	(1.410)	(2.030)	(2.873)	(0.448)	(0.654)	(0.592)
$Treat$		0.711			0.984	
		(3.589)			(0.857)	
$Post$	−2.448**		−0.104	−0.940***		−0.942**
	(1.094)		(1.633)	(0.277)		(0.387)
$OverInd$	6.604**	10.780***	7.826**	1.470***	2.922***	0.658
	(2.069)	(2.565)	(3.328)	(0.523)	(0.804)	(0.622)
Firm FE	Yes		Yes	Yes		Yes
FF12 + Year FE		Yes			Yes	
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,490	2,490	523	2,490	2,490	523
Adjusted R^2	0.691	0.370	0.732	0.796	0.394	0.827

This table presents the results of the treatment effect of longer duration on investment efficiency for an alternative approach. In this approach, an industry average regression (Equation (7)) is used to obtain a proxy for the likelihood of over- and under-investing. $OverInd$ measures the likelihood of over-investing as the ranked deciles (scaled from 0 to 1) of the industry aggregate regression. The tabulated results are obtained from estimating Equation (8). Control variables, intercept, and fixed effect dummies are included but not tabulated. All variables are defined in Appendix A.1. Standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

show the results for the PSM control sample. The coefficients of the three-way interactions are negative and significant for five of the six specifications. The coefficients for the DID interaction are positive but mostly insignificant.[78] Firms with longer duration display lower over-investment when this is likely due to the preponderance of over-investment throughout the industry. However, there is less support for higher investment in treated firms in the absence of industry-level over-investment. The coefficients for the variable $OverInd$ are mostly significantly positive, which shows that firms generally invest more when there is over-investment in the industry. These results provide further support

[78]For C/PPE, the regression at the industry level has very low explanatory power with an adjusted R^2 of -2% (for I/TA the adjusted R^2 is 9%). I include the average CFO in this regression, as I also do in my main analyses (not tabulated). This increases the adjusted R^2 to 26% for the industry mean regression with C/PPE and leads to similar results in the second step.

for a reduction in abnormal investment following the VorstAG treatment. They also provide support for the notion that mandated longer duration more likely leads to less over-investment than under-investment.

2.5.5 Reconciliation with announcement returns

I find consistent results showing that mandated longer duration reduces abnormal investment. Early evidence on announcement returns for voluntary adopters of a particular longer duration compensation component are positive (Kumar and Sopariwala, 1992). However, previous research indicates that the mean stock market reaction to the VorstAG regulation was negative (Hitz and Müller-Bloch, 2015), suggesting that this regulation was perceived to involve costs to shareholders. Linked to increased incentive duration, a substantial body of literature states that having more debt-like compensation features better aligns executives with debtholders and, in particular, decreases the riskiness of firm investment (e.g., Cassell et al., 2012; He, 2015) as debt-like compensation features reduce risk-shifting over-investments (Edmans and Liu, 2011).[79] As different capital providers may have different investment preferences that follow from their different payoff structures (Eisdorfer et al., 2013; Roychowdhury et al., 2019), shareholders may have anticipated the mandate to inefficiently affect firms' risk-taking. The results that indicate a weak or potentially negative effect on firms' under-investment could be further in line with this notion.

Similarly, while the likely intention of the German (stakeholder-oriented) government was to force executive compensation incentive contracting closer to optimality, the act may have pushed at least some firms out of optimality. Shareholders may have anticipated that, by introducing contracting constraints, the VorstAG act would only be beneficial to a few firms and detrimental to others. The government may even have regulated all firms while actually targeting only a subset with particularly suboptimal incentive structures. My cross-sectional results indicate that much of the effect on investment efficiency is driven by

[79]This literature mainly focuses on pensions and deferred compensation, viewing such pay as an unsecured liability to the firm. While pension benefits were not subject to the VorstAG act, many firms increased the deferred portion of pay.

a small number of firms that had particularly weak compensation-related governance in the pre-period. This seems in line with further announcement return analyses, indicating that firms with higher abnormal compensation had lower returns, and firms with higher pay-performance sensitivity had higher returns (Hitz and Müller-Bloch, 2015).

2.6 Conclusion

I exploit an exogenous trigger of a regulation designed to prolong the duration of executive compensation to investigate whether this improves investment efficiency. Mandated longer duration increases the length of executives' evaluation window, which could force them to internalize the unprofitability of today's empire building that materializes in the future, reducing over-investment. At the same time, the longer evaluation window may mitigate myopic disincentives for current-period investments with future payoffs or incentives for over-signalling, reducing under- or over-investment. However, such a regulation may also interfere with efficient contracting or it may be immaterial if it is too vague.

Subjecting this to a series of tests, I find statistically and economically significant improvements in investment efficiency following the regulation. The results are robust to a wide range of (abnormal) investment specifications and different plausible control groups. Cross-sectional analyses suggest that the effects are driven by firms with weak corporate governance characteristics, in particular, firms whose compensation committee had a low degree of independence before the regulation. It further appears that these improvements are mainly driven by reductions in over-investment.

Overall, my findings support the notion that long-term orientation in executive compensation serves a role in mitigating frictions that hamper efficient investment decisions. Governmental and regulatory intervention into corporate governance in general and executive compensation in particular is frequently discussed and demanded. This paper underscores how regulation potentially affects real firm behavior and emphasizes the need to report and investigate the economic impacts of such intervention.

3 The Stock Market Valuation of Human Capital Creation

Abstract

We develop a measure of firm-year-specific human capital investment from publicly disclosed personnel expenses (PE) and examine the stock market valuation of this investment. Measuring the future value of PE ($PEFV$) based on the relation between lagged PE and current operating income, we first show that $PEFV$ is positively associated with characteristics of human-capital-intensive firms. Next, we find that $PEFV$ has a positive pricing coefficient, implying that the market recognizes some of its variation. In our main analysis, we find that market participants fail to fully impound the investment in human capital. The absolute value of analyst forecast errors is increasing in firm $PEFV$, and the signed value of these errors reveals that analysts are pessimistic for earnings of firms with high human capital investments. A long-short portfolio based on $PEFV$ produces annualized value-weighted (equal-weighted) abnormal returns of 6.5% (3.5%). Portfolios formed by interacting $PEFV$ with total PE, which combines the current potential investment in human capital with the historic portion of PE that created human capital, increase these returns to between 4.8% and 7.8%. These results are insensitive to numerous empirical choices.

Keywords: Intangibles, Market valuation, Human capital
JEL classification: M41, E22

Authors: Matthias Regier, Ethan Rouen (Harvard Business School)
Lead author: Matthias Regier

Status: Working Paper under Revise & Resubmit at Journal of Corporate Finance

Acknowledgements We thank Ramji Balakrishnan, Luminita Enache, Cristi Gleason, Rong Huang (discussant), George Serafeim, Anup Srivastava, Charles Wang, and workshop participants at Harvard Business School, the Human Capital Management Coalition, Technical University of Munich, the University of Calgary, the 2021 FARS Midyear Meeting, and the 2021 AAA Annual Meeting for helpful feedback.

3.1 Introduction

Accounting rules require that most expenditures related to employees be treated as costs and expensed as incurred.[80] The reason for this treatment is that unlike with assets, firms do not really have control over their employees. Employees are not forced to remain employed by the firm but can deprive the firm of their capacity and skills (e.g., Blair, 2003). Still, costs related to employees likely consist of two components, the immediate expense that ensures that employees contribute to maintaining current business operations, and the investment that encourages employees to improve in their roles and grow the firm. This latter component, which can take various forms ranging from incentive-based compensation to on-the-job training, gives rise to the trope illustrated by Xerox CEO Anne Mulcahy in 2003 that "Employees are a company's greatest asset."[81]

This paper seeks to better understand the information contained in employee expense disclosures, specifically the "Personnel Expense" (PE) line item that firms are required to disclose under IFRS.[82] To do so, we develop a methodology to identify at the firm-year level how successful a firm is at investing in personnel (i.e., the human capital investment). Our main analysis examines whether market participants recognize and appropriately value this component of PE, what we refer to as the future value of PE, or $PEFV$.

While measuring human capital creation is complex and imperfect, it is growing increasingly important. In a 2000 paper, Luigi Zingales wrote, "The wave of initial public offerings of purely human capital firms... is changing the very nature of the firm" (Zingales, 2000). If anything, the change has accelerated since the time of that writing. On

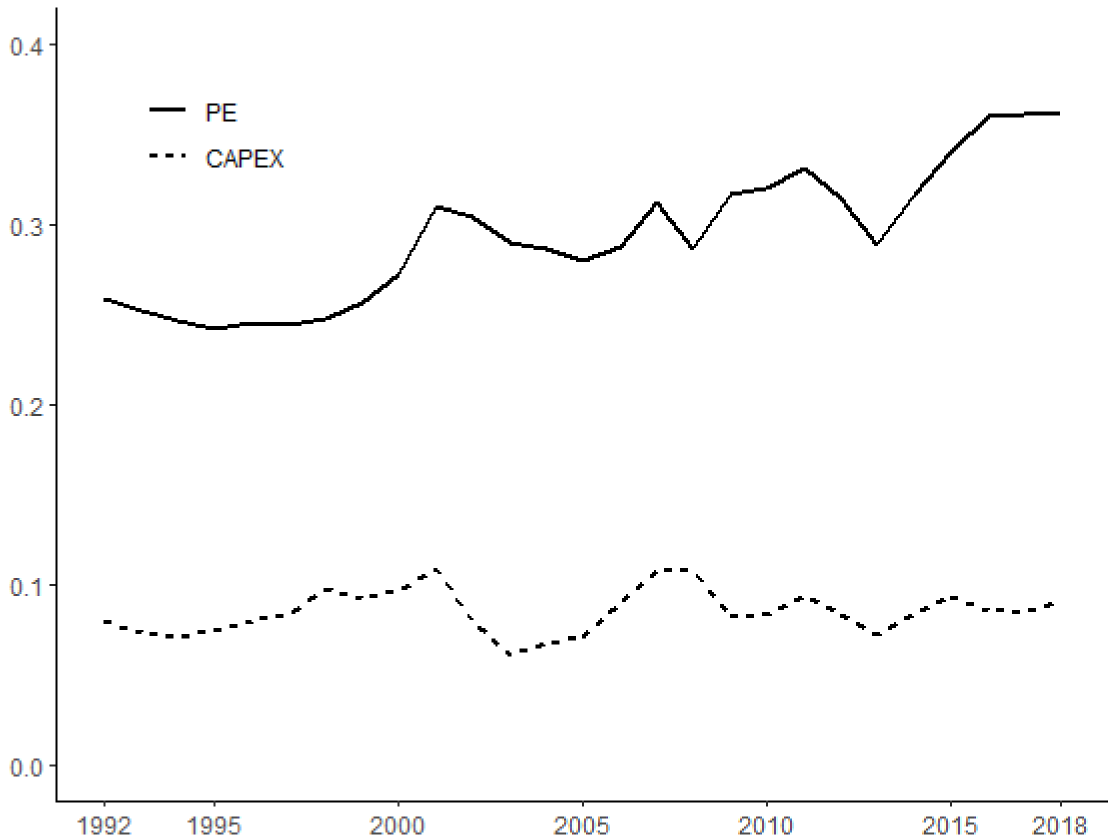
[80] This is not to say that employee expenditures cannot be capitalized at all. For instance, both U.S. Generally Accepted Accounting Principles (GAAP) and International Financial Reporting Standards (IFRS) allow the capitalization of manufacturing work and IFRS even allows to capitalize development costs (i.e., as opposed to research costs) under certain criteria.

[81] This quotation is attributed to a speech Mulcahy made in May 2003 at the Doral Arrowwood Resort in Rye Brook, N.Y. (e.g., Viswanathan and Chopra, 2015).

[82] Throughout the paper we use the terms "personnel expense" or "personnel expenditure" to refer to the Thomson Reuters Datastream item *Personnel expense for all employees and officers* (mnemonic WC01084). This item mostly relates to personnel expenditures that are expensed as incurred. As defined in IAS 19, the item includes, among others, the costs for hiring, wages, salaries and bonuses, social security and insurance costs, perquisites like catering and work wear, and post-termination benefits. While some firms disaggregate expenses by nature so that PE is visible on the income statement, most firms disaggregate by function and provide total PE somewhere in the notes. In the latter case, PE is typically divided among firms' cost of goods sold, selling, general and administrative expense, and research and development expense.

the one hand, as shown in Figure 2, from 1992 to 2018, capital expenditures as a percentage of total sales remained relatively flat at about 10%. On the other hand, *PE* almost continuously increased during that time. By 2018, *PE* consumed more than one third of all of the average firm's revenues in our large sample of publicly traded European firms reporting under IFRS. This comparison illustrates the growing importance of *PE* as an input resource for firms over time.[83]

Figure 2: Development of personnel expenditure and capital expenditure over time



This figure plots the annual average personnel expenditure (solid line) and capital expenditure (dashed line) scaled by total sales over the sample period from 1992 to 2018. The figure includes all firms listed in 30 European countries with available data as described in Panel A of Table 10 where 1992 is the first year with wide availability of data for firms' *PE*.

The growing importance of human capital to firms' profit generating abilities, com-

[83]The growth in *PE* relative to capital expenditures may seem at odds with a decline in national labor income shares documented globally over the same sample period (see, for instance, the report prepared by the International Labour Organization for the G20 Employment Working Group in Turkey, 2015). Be aware that the plotted figures are yearly averages across firms that may be influenced by many factors such as an increase in firms with high personnel costs but little tangible capital investments as suggested by Zingales (2000) or firms shifting their operating systems to have more operating expenses and lower capital expenditures. The figures may also be influenced by the development of the denominator, for instance, when firms increasingly go public earlier with lower revenues compared to the personnel costs already incurred.

bined with the paucity of disclosures related to employees and investment in the workforce, potentially creates an information gap that distorts valuation of the firm (Zingales, 2000). While IFRS requires firms to disclose *PE*, under U.S. GAAP, firms are required to disclose only the total number of employees and, since 2018, the salary of the median employee, a measure that lacks relation to future performance (Rouen, 2020).

Given these limited disclosures, investors face informational challenges when attempting to recognize variation in firms' abilities to effectively invest in intangible assets broadly and generate human capital specifically. Prior research investigates whether markets realize the future value generated by firms' expenditure on input resources such as research and development (R&D) (e.g., Eberhart et al., 2004; Lev and Sougiannis, 1996), advertising (Chan et al., 2001), and selling, general, and administrative (SG&A) (Banker et al., 2019). Moreover, accounting and finance scholars have shown the need for markets' recognition of firms' human capital quality (e.g., Ballester et al., 2002; Edmans, 2011; Lee et al., 2018; Pantzalis and Park, 2009). To provide a better understanding of human capital investments, we analyze the stock market valuation of *PE*, the expenditure of the input resource that is most intuitively related to the ability to create human capital.

It is unclear whether and how information regarding firm's human capital creation implicit in *PE* is recognized in market valuations. To a large extent, *PE* consists of the wages paid to workers in the period in which that work is done. If intangible human capital investments are absent from (or an insignificant component of) *PE*, then there should be little relation between *PE* and (current or future) returns given that the resource is consumed in the period in which it is reported. Alternatively, Bertrand and Mullainathan (2003) suggest that abnormally high *PE* may be due to a failure of governance, with managers paying more than is required to reduce their obligations at a cost to shareholders, meaning that higher *PE* may be associated with lower returns. Lastly, it may be that a portion of a firm's *PE* supports current operations as a cost, while another significant portion constitutes a personnel investment to develop human capital for future income (Flamholtz, 1971). The latter should be recognized in current or future returns.

Prior literature has provided suggestive evidence of the usefulness of *PE* for valuation purposes. While expenditures associated with human capital investments are not recognized on firms' balance sheets, total *PE*, as reported on the income statement, has been shown to increase earnings predictability and have value relevance (Schiemann and Guenther, 2013; Rouen, 2019). If a meaningful portion of *PE* represents investment in human capital, and human capital accounts for a relevant portion of firms' market values, then these investments, when properly measured, should be predictive of returns (Ballester et al., 2002). Moreover, *PE* clearly supports employee satisfaction, which correlates with abnormal returns (Edmans, 2011).

Another potential reason why the human capital creation implicit in *PE* could be value relevant relates to risk. Human capital creation in general and high *PE* in particular increase firm risk, given that these investments, much like R&D, have uncertain outcomes, and, in a way, may have even greater uncertainty than research investments: Similar to research, investing in employees comes with the risk that the investment might fail due to a misunderstanding of the employees or skill in which the firm invests. In addition, because firms do not own their employees, human capital is reduced when employees leave the firm (Lev and Schwartz, 1971), while successful R&D investments can be protected via patents. Additional risks may therefore stem from greater uncertainty regarding whether firms are able to control and monetize the created human capital. High *PE* may also be difficult or costly to adjust in the short run, leading to high labor leverage and increasing firms' equity risk (Donangelo et al., 2019; Rosett, 2003), which could lead markets to demand a risk premium. For example, Donangelo et al. (2019) finds that firms with high labor bills have higher expected returns, in part because these firms' operating profits are more sensitive to economic shocks given the stickiness of employee costs.

Our approach differs from prior studies in that we acknowledge that *PE* can impact future earnings (Schiemann and Guenther, 2013), and that there are firms where *PE* constitutes a substantial human capital investment (Ballester et al., 2002), but we capture cross-industry and cross-firm *variation* in the ability to create future value from *PE*. This strategy explained below allows us to explore whether and when the stock market realizes the future value created by *PE*.

This paper takes several steps to further the nascent literature on the relations among employee expenditures, human capital creation, and firm performance. Adapting methodologies to extract from an expenditure the intangible assets created by that expenditure, we create a proxy for the component of PE consisting of investments in human capital by identifying the relation between prior period PE and current firm performance (e.g., Banker et al., 2019; Chen et al., 2012; Huson et al., 2012; Lev and Sougiannis, 1996). For a large sample of firms across 30 European countries, in our main analyses we begin by regressing at the industry level current operating income on several years of lagged PE to identify the optimal lag structure for each industry in terms of the number of years in which PE influences income after that PE is initially incurred. The optimal lag structure is thereby determined by identifying the number of prior years in which PE has a statistically significant relation with current operating income and choosing the structure with the most explanatory power. In some industries, as many as three years of lagged PE are significantly positively associated with current operating income (e.g., manufacturing) while in other industries, prior PE has no relation to current performance (e.g., chemicals). Next, we rerun these regressions at the firm-year level using the industry-determined lag structure. Summing the coefficients on prior PE from these regressions provides a firm-year estimate of the PE future value ($PEFV$), our main variable of interest.

We begin our empirical analysis by validating our proxy for human capital investment ($PEFV$), examining whether $PEFV$ is associated with firm characteristics that are likely to be related to the importance of human capital creation. We find that firms with higher $PEFV$ are smaller, have higher market-to-book ratios, have fewer tangible assets, and provide more training days to employees. Documenting that growth firms and less capital-intensive firms are associated with higher $PEFV$ provides us with confidence in this measure as an effective proxy for investments in human capital.

Next, we examine the association between $PEFV$ and contemporaneous stock price. While the relation between total PE and stock price is negative and significant, the relation between $PEFV$ and contemporaneous price is positive and significant. This result suggests that the stock market, to some extent, differentiates between the current operating expense component of *PE* and the future value of *PE*, which is treated as an intangible asset. In other words, the stock market recognizes at least some of firms' human capital creation at the time when the investment in that human capital materializes (i.e., when the prior period investment is consumed). This result is robust to a battery of different controls and specifications. While *PEFV* is measured with error, these results provide further evidence that *PEFV* captures, in part, the investment we are attempting to measure.

Our main analyses examine whether market participants fully recognize this future value of the intangible asset included in *PE*. To do so, we analyze the predictive power of *PEFV* for sell-side analysts' earnings forecast errors and firms' future stock returns. First, we find a significant positive relation between the magnitude of *PEFV* and forecast errors, as well as absolute forecast errors. These results suggest that, not only do analysts fail to incorporate into their forecasts the full value of the investment component of *PE*, but that they overweight the expense component, resulting in pessimistic forecasts.

Next, we build two types of portfolios based on *PEFV*. First, we sort firms into portfolios based solely on *PEFV*. Second, we sort firms into portfolios based on the combination of *PEFV* and *PE* scaled by total assets, *PEFV*PE/TA*. This second set of analyses provides insights into both the investment in human capital and the opportunity to make that investment, based on the total amount spent on employees in the current period. A value-weighted (equal-weighted) long-short investment strategy based on the level of *PEFV* returns annualized abnormal returns of 6.5% (3.5%), while a strategy that divides firms into portfolios based on *PEFV*PE/TA* results in abnormal annualized returns of 7.8% (4.8%) in the subsequent year. These results, which are statistically and economically significant, suggest that the market fails to fully impound the human capital development embedded in *PE*, as well as the opportunity to make that investment. The results are also robust to numerous alternative specifications, including assigning portfolios based on industry, excluding firms from countries with illiquid currencies, using different factor models, and requiring identical lag structures across all firms when calculating *PEFV*.

The results also remain unchanged when conducting Fama-MacBeth cross-sectional regressions (Fama and MacBeth, 1973). Lastly, we find that the abnormal portfolio returns decrease monotonically over time, with statistically significant value-weighted returns of 5.1% in the second year after portfolio formation, and insignificant returns of 3.1% in the third year.

Given that *PEFV* likely fails to include some investment in human capital that did not materialize, we also explore whether the results are robust to an alternative measure of human capital investment that is likely to capture these investments. We adapt Enache and Srivastava (2017) in creating an alternative measure of human capital investment and find that portfolios formed using this measure continue to produce abnormal returns. Still, we acknowledge that our proxies for human capital investment are measured with error. Total personnel expense is an admittedly crude starting point to approximate measures of human capital. Included in *PE* is not only wages, social security expenses, and training costs, but also costs like uniforms, firm-hosted daycare centers, and meals. Exacerbating the challenge is that firms in our sample do not disaggregate this significant operating expense in any meaningful way. Therefore, the findings in this paper should serve as evidence that even disclosures that are only vaguely related to human capital have an information content that is value relevant. This can provide a basis for how firms and regulators can improve employee-related disclosures as they become increasingly relevant in the knowledge economy.

This paper makes several contributions to the literature. First, we provide evidence of the value of human-capital-related disclosures to market participants. There is little evidence of the relation between employee expense and future firm performance, and we are the first to develop an effective way to extract the future value of the expenditure from the total expense.[84] We show not only that there is significant variation in the ability of firms to generate future value from their investment in employees through *PE*, but also that employee expenses are relevant for future performance and mispriced by the market.

Second, we contribute to ongoing regulatory debates. Broadly, our result that the

[84]Papers that examine labor expenses' relation to firm market value are Schiemann and Guenther (2013) and Ballester et al. (2002).

market does not fully recognize the human capital creation implicit in *PE* supports the need to consider changes in the accounting for input resource expenditures (e.g., Enache and Srivastava, 2017; Lev, 2019). More specifically our results are informative to U.S. investors and the U.S. SEC, which recently passed an amendment to its Regulation S-K requiring firms to provide a description of the importance of their human capital resources to the underlying business. The current requirement gives firms wide latitude in terms of what they define as material human capital information, and large investors continue to engage regulators on which human capital disclosures are value relevant (Human-Capital-Management-Coalition, 2019; Maurer, 2021). Our results provide guidance on the disclosures that are relevant to investors. Crude total *PE* could already contain important information. Relatedly, the convergence project between the Financial Accounting Standards Board and the International Accounting Standards Board has discussed whether it is more informative to disaggregate costs by their function or by their nature, including a debate as to whether disclosure of nature of expense items like *PE* should also be mandatory under by-function systems.[85] Moreover, separating total *PE* into its expense and investment components appears to be meaningful. Our paper uses two methodologies to estimate the latter, simplified versions might also be suitable for financial reporting.[86]

Finally, we add the nature of expense perspective to the stream of research on the stock market valuation of intangible assets. Prior literature shows that the market misvalues functional expenses like R&D (Chan et al., 1990; Eberhart et al., 2004; Lev and Sougiannis, 1996), advertising (Chan et al., 2001), and SG&A (Banker et al., 2019). Until now, this research has paid little attention to the nature of the expense, broadly, and *PE* specifically. Relatedly, we expand the emerging literature on the impact of firms' ability to generate future value from input resource expenditures. For instance, scholars have analyzed the effects of executive compensation and cost decisions on market valuations (Banker et al., 2011; Chen et al., 2012; Huson et al., 2012; Banker et al., 2019). These studies limit their

[85] We refer to the Financial Accounting Standards Board and International Accounting Standards Board Joint Meeting on Primary Financial Statements in June 2018.

[86] Our paper can neither quantify the benefits of human capital creation disclosures for other stakeholders, nor identify which kinds of disclosures are most relevant for them. It seems straightforward, though, that employees or their interest groups could benefit heavily from more such disclosures.

evidence to a subset of employees and rely only on evidence for U.S. firms. We examine an intuitive, widely reported input resource that can be analyzed in an IFRS cross-country setting.

The remainder of this paper is organized as follows. Section 2 of the chapter (i.e., 3.2) describes the data used and the research design. Section 3 reports descriptive statistics, and Section 4 describes the main empirical results and robustness analyses. Section 5 concludes the paper.

3.2 Research design, data, and variable measurement

3.2.1 Research setting and sample selection

To test whether the market realizes firms' human capital creation from *PE*, we exploit the mandate to disclose *PE* for firms listed in European Union (EU) countries. Firms listed on an EU regulated market must report according to IFRS, which requires disclosure of *PE*. We include in our sample the current 27 members of the EU as well as the United Kingdom, which left the EU in early 2020. We further add Norway and Switzerland (e.g., Armstrong et al., 2010; Byard et al., 2011). We therefore begin the sample selection with all firms listed in any of these 30 countries.

Panel A of Table 10 shows the sample selection procedure. We consider 11,569 non-financial and non-utilities firms (e.g., He and Narayanamoorthy, 2020) that were active at some point in time during the period 1991 to 2018. For those firms, we obtain Thomson Reuters Datastream data for 124,507 firm-years from 1992 to 2018, which begins one year later since we use average total assets (*TA*) to deflate the financial statement variables.[87] We remove firm-years with missing financial statement items (*TA*, *PE*, operating income (*OI*), and depreciation & amortization), stock-related data (share price and market capitalization), and number of employees. We consider only firm-years with at least \$20 (\$0.5) million in total assets (personnel expenses). We require at least five firms in every SIC-

[87]Many countries required the disclosure of *PE* prior to the EU's IFRS adoption in 2005. We neither observe a kink in data availability around 2005, nor in any other year. We therefore begin our sample in 1992, when these data become widely available.

Table 10: Sample selection and distribution

Panel A: Sample selection procedure

Selection step	Firms	Firm-years
Thomson Reuters Datastream annual data from 1992 to 2018 for non-financial/non-utilities firms from 30 countries	11,569	124,507
− firm-years with missing financial statement items, stock-related data and number of employees	(2,176)	(39,351)
− firm-years with total assets below 20 million US$ and personnel expenses below 0.5 million US$	(1,202)	(10,740)
− SIC-4-industry-years with less then five firms	(649)	(9,837)
Sample used to winsorize yearly and to estimate $(PE/TA_{predicted})_{i,t}$ with the instrumental variable approach	7,542	64,579
− firm-years with missing data in any of the preceding four years	(3,034)	(30,561)
− FF12-industry-years with less than 15 firms	(3)	(29)
Sample for estimation of optimal lag structure for each FF12-industry from 1996 to 2018 (N in Panel B below)	4,505	33,989
− firm-years with not sufficient data for firm-year-specific regressions	(1,549)	(12,281)
Sample for estimation of PEFV per firm-year	2,956	21,708
− firm-years with negative PEFV	(472)	(9,638)
− firms from industries with zero lags (zero PEFV)	(130)	(1,061)
Final sample with positive PEFV per firm-year from 1998 to 2018 (N_{final})	2,354	11,009
− firm-years with missing earnings per share forecast	(431)	(2,183)
Subsample with forecast data availability from 1998 to 2018	1,923	8,826
− firm-years with missing forecast data to calculate forecast errors	(28)	(154)
Subsample with forecast error data availability from 1998 to 2018	1,895	8,672

Panel B: Firm-year distribution among countries and FF12-industries

Country	N	N_{final}	FF12-industry	N	N_{final}
Austria	595	208	(1) Consumer NonDurables	3,757	1,328
Belgium	805	234	(2) Consumer Durables	1,412	557
Denmark	1,072	333	(3) Manufacturing	6,659	1,944
Finland	1,186	441	(4) Oil, Gas, & Coal Extract. & Products	1,044	411
France	5,226	1,780	(5) Chemicals & Allied Products	1,221	-
Germany	4,723	1,492	(6) Business Equipment	5,371	1,991
Greece	401	70	(7) Telephone & Television Transmission	1,017	434
Hungary	121	57	(8) Utilities (excluded)	-	-
Ireland	474	171	(9) Wholesale, Retail, & Some Services	3,949	1,102
Italy	1,781	595	(10) Healthc., Medical Equipm., & Drugs	1,613	517
Luxembourg	121	44	(11) Finance (excluded)	-	-
The Netherlands	1,373	455	(12) Other	7,946	2,725
Poland	756	164	Total	33,989	11,009
Portugal	501	194			
Spain	1,152	420			
Sweden	1,679	567			
United Kingdom	8,692	2,851			
Switzerland	1,820	562			
Norway	962	281			
Others (BG, CY, CZ, EE, HR, LT, LV, MT, RO, SI, SK)	549	90			
Total	33,989	11,009			

4-industry-year (e.g., Banker et al., 2011; Lev and Sougiannis, 1996).[88] This procedure results in an initial sample of 64,579 firm-years. Based on this sample, we winsorize the financial statement ratio variables yearly at the 1% and 99% level (Banker et al., 2019). We then remove firm-years where less than four years of lagged data are available, which leaves a sample period from 1996 to 2018. Removing FF12-industry-years with less than 15 firms gives the sample of 33,989 firm-years used to obtain the optimal lag structure per FF12-industry.[89] Of that sample, 21,708 firm-years have sufficient lagged data to allow the firm-year-specific calculation of the human capital investment (i.e., the personnel expenditure future value, $PEFV$). In our main analyses, we focus on the firms from industries with at least one lag and positive $PEFV$ estimates. The earliest year where a calculation with one lag is possible is 1998. Our final sample contains 11,009 positive $PEFV$ firm-years from 1998 to 2018.

Panel B of Table 10 shows the distribution of firm-years among countries and FF12-industries for the 33,989 firm-years used to obtain the optimal lag structure per industry and for the final sample of 11,009 positive $PEFV$ firm-years. United Kingdom firms account for the largest portion of firm-years, followed by French and German firms. The relative weight of the sampled countries is comparable with other studies on EU firms (e.g., Armstrong et al., 2010; Byard et al., 2011; Christensen et al., 2013), implying that the required data availability does not distort the sample such that generalization of the results to the universe of EU firms is not warranted. Firms in the Manufacturing, Business Equipment and residual category Other industries account for the largest portion of firm-years. The sample reduction induced by focusing on the firm-years with positive $PEFV$ (i.e., from N to N_{final}) distorts neither the country nor the industry distribution.

[88] This requirement is needed for the instrumental variable approach that we explain in the next section. If there are less than five firms available in the SIC-4-industry-year, we pool the firms on the SIC-3 level, where we again require at least five firms in the industry-year.

[89] We use the Fama and French industry classification as it provides intuitive and consistent categories to assess the industry-specific lag structure. At the same time, we rely on the SIC categorization for those parts of our methodology that require a numerical disaggregation.

3.2.2 Measurement of personnel expenditure future value

We begin our analysis of the human capital investment implicit in *PE* by estimating the long-term effect of lagged *PE* on current operating income following a two-step procedure. First, we obtain the optimal *PE* lag structure for the relation between operating income and *PE* for each FF12-industry using the following equation:

$$OI/TA_{i,t} = \alpha + \sum_{k=0}^{n} \beta_k (PE/TA_{predicted})_{i,t-k} + \gamma log(\#E)_{i,t} + \eta_t + \varepsilon_{i,t}. \qquad (9)$$

Equation (9) is adapted from earlier methodological approaches to be currency neutral (e.g., Banker et al., 2011; Lev and Sougiannis, 1996). We estimate equation (9) for each FF12-industry with different numbers of lags (different n).[90] $OI/TA_{i,t}$ is operating income before depreciation & amortization and *PE* (e.g., Banker et al., 2019) deflated by average *TA*. $(PE/TA_{predicted})_{i,t-k}$ is the predicted value using an instrumental variables approach for the deflated *PE* of year *t-k* as follows:

Following Lev and Sougiannis (1996) and Banker et al. (2019), we use industry-year *PE* as an instrument in equation (9) to address a potential simultaneity problem when a shock to the residual affects both the dependent (*OI*) and the independent variable (*PE*).[91] For each firm-year observation, we calculate the average *PE* of all other firms in the SIC-4-industry ($(PE/TA_{SIC4-i})_{i,t}$). We assume that firm idiosyncratic shocks do not affect industry-year *PE*.[92] At the same time, industry-year *PE* should be highly correlated with firm-year *PE*. For each year and SIC-2-industry, we regress $PE/TA_{i,t}$ on the industry-year *PE*:

$$PE/TA_{i,t} = \alpha + \beta (PE/TA_{SIC4-i})_{i,t} + \varepsilon_{i,t} \qquad (10)$$

[90] Banker et al. (2019) consider models ranging from zero to seven years, Huson et al. (2012) consider up to five lagged years in their industry-specific analyses of the future value of SG&A. It appears unlikely that rather old human capital is still systematically relevant for operating income. Moreover, Ballester et al. (2002) find that human capital assets depreciate, on average, over three years. Thus, we consider models ranging from zero to four lags of *PE* in the industry-specific analysis.

[91] For example, demand for a firm's products may increase due to some exogenous shock. This could lead to both an increase in *OI* and an increase in the returns to input resource expenditure like *PE*, which would in turn lead to an increase in *PE*. *PE* could therefore no longer be treated as an exogenous variable.

[92] The firms in a SIC-4-industry may still be subject to a SIC-4-industry idiosyncratic shock.

We obtain the predicted value $(PE/TA_{predicted})_{i,t}$ from equation (10) and use it in the industry-level and firm-year-level estimations of equation (9).

In equation (9), we include the natural logarithm of the number of employees to account for firm size as there may be scale effects when analyzing how intangible investments are reflected in future income (Ciftci and Cready, 2011) and also include year indicators (η_t).[93] For each FF12-industry, we determine the lag structure with all positive and statistically significant (at the one-sided 10% level) coefficients and the most explanatory power.[94]

Second, we fix the optimal lag structure from the first step for all firms of a given industry. We next rerun equation (9) at the firm-year level.[95] For each firm-year, we use current and historical data of that firm, compatible with an investor's information set at a given point in time. We only run the regression in firm-years where there is sufficient historical data to obtain all coefficients of the respective model.[96] We use rolling windows of historical years in the firm-specific regressions using the number of lags determined at the industry level in the first step. $PEFV_{i,t}$ is calculated as the (discounted) sum of the firm-year-specific coefficients on past *PE* ($PEFV_{i,t} = \sum_{k=1}^{n} \beta_k/(1.1)^k$) and serves as the proxy for human capital investment.[97] The intuition is that it reflects the total effect of a currency unit of spending of current *PE* on future *OI*. To allay concerns about measurement error, we show in Section 3.4.4 that our main results are robust to an alternative strategy for measuring human capital investment.

[93] Banker et al. (2019) add current R&D and advertising expenditures to the model when estimating the future value of SG&A. *PE* already contains the personnel expenses included in SG&A, R&D and advertising, so we do not add any of the functional expenditure items to the model.

[94] We assess the explanatory power according to the Akaike Information Criterion (AIC), the Schwartz Bayesian Criterion (SBC), and adjusted R^2. We thereby regard a model to have the highest explanatory power when both AIC and SBC are lowest for this model. If the AIC and SBC criterion leave two different models, the model with the higher adjusted R^2 is chosen.

[95] We do not include the proxy for firm size when running the regressions on the firm-year level. Those regressions also do not provide the degrees of freedom to include year indicators.

[96] For instance, for a firm with full data coverage from 1992 to 2018, 1996 is the first year where data of the four preceding years is available. If the firm operates in a FF12-industry where we identify three lags to have the highest explanatory power, then the firm-year-specific regression for this firm has five coefficients (α and β_0 to β_3). This regression is possible from year 2000 onward.

[97] We use the same interest rate of ten percent to discount the coefficients as earlier papers (e.g., Banker et al., 2011, 2019). The results are not sensitive to the choice of the interest rate.

3.2.3 Optimal lag structure

The first step of the two-step-procedure to estimate $PEFV$ is to define the optimal lag structure for each industry by estimating equation (9) at the industry level. To gain initial insight on the impact of past PE on current OI, we show results for estimating equation (9) *across* industries including FF12-industry indicators in Panel A of Table 11. We show results for structures of one to four lags. The table shows that past streams of PE with different lag structures have significantly positive effects on current OI. In each of the four models, the discounted coefficients on past PE add up to between 0.355 and 0.417. It thus appears that a substantial portion of PE is a value-creating investment on average.

Next, we obtain the optimal lag structure per industry. We run equation (9) industry-by-industry. Panel B of Table 11 provides the coefficient estimates for the lag structure with all positive and significant coefficients and the highest explanatory power for each industry. The optimal number of lags varies substantially from zero to three. Past PE has no impact on current OI in the Chemicals & Allied Products industry. It appears meaningful that the lag structure persists into two or three earlier years in industries like Manufacturing and Healthcare, Medical Equipment, and Drugs, where firms can add relatively high value to the products and services they offer through human capital investments. Consumer-oriented industries like Consumer NonDurables and Wholesale, Retail, and Some Services also seem to have longer lag structures. Overall, the results support the notion that the magnitude of the future values generated by PE varies considerably across industries. In a later section, we apply two or three lags across all industries to allow firms to "compete on equal grounds" and find that our main results remain unchanged.

3.3 Descriptive statistics

3.3.1 Descriptive statistics for important variables

Table 12 shows descriptive statistics for the lag structure variables of equation (9) and for the variables used in the contemporaneous stock price and forecast error analyses.

Table 11: Lag structure regressions

Panel A: Cross-sectional regressions with different lag structures

	Dependent variable:			
	$OI/TA_{i,t}$			
	(1)	(2)	(3)	(4)
$(PE/TA_{predicted})_{i,t}$	0.664***	0.609***	0.566***	0.552***
	(0.038)	(0.039)	(0.039)	(0.038)
$(PE/TA_{predicted})_{i,t-1}$	0.391***	0.164***	0.148***	0.113**
	(0.037)	(0.046)	(0.045)	(0.046)
$(PE/TA_{predicted})_{i,t-2}$		0.296***	0.096**	0.079*
		(0.039)	(0.046)	(0.045)
$(PE/TA_{predicted})_{i,t-3}$			0.271***	0.079*
			(0.039)	(0.044)
$(PE/TA_{predicted})_{i,t-4}$				0.270***
				(0.037)
$log(\#E)_{i,t}$	0.016***	0.016***	0.016***	0.016***
	(0.001)	(0.001)	(0.001)	(0.001)
Intercept	−0.034***	−0.037***	−0.038***	−0.043***
	(0.009)	(0.009)	(0.009)	(0.009)
$\sum_{k=1}^{n} \beta_k/(1.1)^k$	0.355	0.394	0.417	0.412
Year dummies	Yes	Yes	Yes	Yes
FF12 dummies	Yes	Yes	Yes	Yes
Robust SEs	Yes	Yes	Yes	Yes
AIC	-14286	-14366	-14435	-14508
BIC	-13983	-14054	-14114	-14179
Observations	33,989	33,989	33,989	33,989
Adjusted R^2	0.355	0.357	0.358	0.360

Panel B: Optimal lag structure per FF12-industry

FF12-industry	β_0	β_1	β_2	β_3	$\sum_{k=1}^{n} \beta_k/(1.1)^k$	$R^2_{adj.}$
(1) Consumer NonDurables	.586	.179	.274		.390	.176
(2) Consumer Durables	.957	.578			.526	.114
(3) Manufacturing	.455	.259	.122	.289	.554	.214
(4) Oil, Gas, & Coal Extract. & Products	.310	.451			.410	.260
(5) Chemicals & Allied Products	.833				-	.120
(6) Business Equipment	.517	.524			.477	.209
(7) Telephone & Television Transmission	.623	.612			.556	.015
(8) Utilities (excluded)	-	-	-	-	-	-
(9) Wholesale, Retail, & Some Services	.614	.139	.186	.292	.499	.330
(10) Healthc., Medical Equipm., & Drugs	.554	.499	.295		.697	.299
(11) Finance (excluded)	-	-	-	-	-	-
(12) Other	.643	.136	.248		.328	.415

This table shows the derivation of the optimal lag structure per industry. Panel A reports results of cross-sectional regressions for different lag structures following equation (9). All variables are defined in Appendix B.1. Columns (1) to (4) present results for one to four lags. Coefficients on industry dummies are not reported. Robust standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. Panel B reports the optimal lag structure per industry where equation (9) is estimated industry-by-industry including year dummies for lag structures of zero to four lags. Only lag structures where all coefficients are positive and significant on the one-sided ten percent level are considered for the choice of the optimal model. The table reports coefficient estimates for the lag structure with the highest explanatory power for each FF12-industry. The last two columns report the discounted sum of the coefficients for the respective optimal lags and the adjusted R^2.

Panel A gives descriptive statistics for the initial sample before requiring four years of lagged data. Measured in U.S. dollars, the mean (median) *TA* value is $2,861 million ($239 million) and the mean (median) *PE* value is $417 million ($52 million). The mean (median) *PE* scaled by average *TA* (*PE/TA*) amounts to 0.27 (0.23). Panel B shows descriptive statistics for the final sample of positive *PEFV* firm-years. The observations included in this sample are larger in terms of *TA* and *PE* compared with the initial sample. We calculate $PEFV_{i,t}$ as the sum of the present values of the coefficients on lagged *PE* for each firm-year. Focusing on positive $PEFV_{i,t}$ estimates gives a highly right-skewed variable. We therefore winsorize it at the 95% level. The resulting mean value is 2.08, and the median is 1.28, which implies that the total effect of spending of *PE* on future operating income is larger than its nominal value. Panel B further describes the variables used in the contemporaneous price analyses and the contemporaneous forecast error analyses. All variables are defined in Appendix B.1.[98]

3.3.2 *PEFV* and firm characteristics

To assess the plausibility of *PEFV* as a proxy for human capital creation, Table 13 presents evidence of the association between firm characteristics and *PEFV*. We use deciles of *PEFV*, rescaled to range from zero to one for firm-years with a positive *PEFV*. Firms' logged number of employees as a proxy for size or life-cycle is significantly negatively associated with *PEFV*, implying that smaller firms are more likely to generate high future values from their *PE* investments. The significantly positive coefficient on the market-to-book ratio suggests that growth firms have higher *PEFV*. The coefficient for asset tangibility is significantly negative, and the coefficient for current *PE/TA* is significantly positive, which means that firms that are less capital-intensive and more reliant on employees are more effective at investing in human capital. Average pay per employee is significantly positively associated with *PEFV* on a stand-alone basis. When examining all variables in a single model in column (6), our inferences remain unchanged, with the exception of the coefficient on *MeanPay*, which becomes insignificant. Column (7) re-

[98]All variables are scaled by $P_{i,t-1}$ and winsorized at the 5% and 95% level.

Table 12: Descriptive statistics

Panel A: Characteristics of sample firms from 1992 to 2018

	N	Mean	STD	Min	25%	Median	75%	Max
$TA_{i,t}(\$m)$	64,579	2,861	13,615	20	80	239	998	397,812
$PE_{i,t}(\$m)$	64,579	417	1,524	0.5	17	52	197	40,950
$PE/TA_{i,t}$	64,579	0.27	0.21	0.01	0.13	0.23	0.36	1.33
$(PE/TA_{predicted})_{i,t}$	64,579	0.27	0.13	−0.30	0.18	0.26	0.34	1.28
$OI/TA_{i,t}$	64,579	0.37	0.25	−0.25	0.21	0.33	0.49	1.57
$PE/SALES_{i,t}$	64,579	0.30	0.36	0.02	0.15	0.24	0.34	6.08
$CAPEX/SALES_{i,t}$	63,794	0.09	0.19	0.00	0.02	0.04	0.07	1.98

Panel B: Descriptive statistics of final sample with positive *PEFV*

	N	Mean	STD	Min	25%	Median	75%	Max
$TA_{i,t}(\$m)$	11,009	5,540	21,649	20	177	517	2,459	396,812
$PE_{i,t}(\$m)$	11,009	729	2,089	0.5	43	120	449	38,762
$PEFV_{i,t}$	11,009	2.08	2.04	0.00	0.50	1.28	3.03	6.51
$PE/TA_{i,t}$	11,009	0.28	0.22	0.01	0.13	0.23	0.36	1.33
$PEFV * PE/TA_{i,t}$	11,009	0.49	0.53	0.00	0.09	0.26	0.75	1.64
$P_{i,t}/P_{i,t-1}$	11,009	1.09	0.47	0.07	0.81	1.04	1.30	5.32
$OIPS_{i,t}/P_{i,t-1}$	11,009	0.63	0.57	0.07	0.22	0.43	0.84	2.19
$PEPS_{i,t}/P_{i,t-1}$	11,009	0.56	0.56	0.05	0.16	0.35	0.76	2.09
$PEFV_{i,t}/P_{i,t-1}$	11,009	0.48	0.70	0.00	0.03	0.12	0.56	2.21
$EPS_{i,t}/P_{i,t-1}$	8,826	0.07	0.05	−0.06	0.04	0.07	0.10	0.18
$SGAPS_{i,t}/P_{i,t-1}$	7,164	0.44	0.45	0.03	0.12	0.27	0.57	1.70
$RNDPS_{i,t}/P_{i,t-1}$	11,009	0.02	0.03	0.00	0.00	0.00	0.01	0.13
$\|FE_{i,t}\|/P_{i,t-1}$	8,672	0.04	0.10	0.00	0.01	0.01	0.04	1.46
$FE_{i,t}/P_{i,t-1}$	8,672	0.01	0.10	−0.40	−0.02	−0.004	0.01	1.30
$\#Analysts_{i,t}$	8,672	8.22	7.81	1	2	5	12	44

This table reports descriptive statistics for different samples. Panel A reports characteristics for the initial sample of firms from 1992 to 2018 and Panel B reports descriptive statistics for the final sample with positive *PEFV*.

ports the relation between the average training days per employee and $PEFV$ for the small subsample of firms that report this information. Consistent with $PEFV$ being a proxy for human capital investment, the coefficient on *TrainingDays* is positive and significant. Overall, these results suggest that firms that are more reliant on labor, faster growing, and less reliant on capital, as well as those that invest more in training, are, on average, more effective at creating human capital, lending credence to the claim that $PEFV$ is an intuitive proxy for human capital investment.

Table 13: Firm characteristics and $PEFV$

	Dependent variable:						
	$PEFV - Decile_{i,t}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$log(\#E)_{i,t}$	-0.029^{***}					-0.030^{***}	-0.035^{***}
	(0.002)					(0.002)	(0.010)
$MTB_{i,t}$		0.006^{***}				0.007^{***}	
		(0.002)				(0.002)	
$Tangibility_{i,t}$			-0.059^{***}			-0.037	
			(0.022)			(0.023)	
$PE/TA_{i,t}$				0.076^{***}		0.080^{***}	
				(0.023)		(0.024)	
$MeanPay_{i,t}$					0.001^{***}	0.00004	
					(0.0001)	(0.0001)	
$TrainingDays_{i,t}(\%)$							0.020^{*}
							(0.012)
$Intercept$	0.891^{***}	0.629^{***}	0.670^{***}	0.633^{***}	0.630^{***}	0.872^{***}	0.838^{***}
	(0.031)	(0.019)	(0.021)	(0.018)	(0.016)	(0.033)	(0.094)
FF12 dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,009	10,627	11,008	11,009	11,009	10,626	1,013
Adjusted R^2	0.089	0.060	0.059	0.060	0.062	0.097	0.133

This table reports the results of OLS regression of $PEFV - Decile_{i,t}$ on human-capital-related firm characteristics. All variables are defined in Appendix B.1. Industry and year dummies are not shown. Two-way-cluster robust standard errors, clustering at the firm and year levels, are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

3.4 Market participants' recognition of human capital creation

3.4.1 Contemporaneous stock prices and *PEFV*

Having shown that *PEFV* is a plausible proxy for firms' human capital investment, we next turn to our main analysis, examining whether stock market participants recognize this investment in a timely manner. In our first market realization analysis, we estimate the association between contemporaneous stock prices and *PEFV*. To do so, we estimate the model from Kothari and Zimmerman (1995) as follows:

$$P_{i,t}/P_{i,t-1} = \alpha + \beta OIPS_{i,t}/P_{i,t-1} + \gamma PEPS_{i,t}/P_{i,t-1} + \delta PEFV_{i,t}/P_{i,t-1} + Controls + \varepsilon_{i,t}, \tag{11}$$

where $P_{i,t}$ is the end of year stock price, $OIPS_{i,t}$ is a per-share measure of OI excluding *PE*, $PEPS_{i,t}$ is *PE* per share, and $PEFV_{i,t}$ is the firm-year-specific future value of *PE*. All variables are converted to U.S. dollars and deflated by the beginning of year stock price to address scale differences. If the future value of human capital investment has a positive impact on contemporaneous price, then we expect a positive coefficient on δ. We expect β to have a positive pricing coefficient. If the contemporaneous stock market values *PE*'s current portion negatively (given that the expense mechanically reduces earnings), γ will be negative.

Table 14 shows the regression results of implementing equation (11). The coefficient on $OIPS_{i,t}/P_{i,t-1}$ is positive and significant in all specifications, indicating a positive relation between *OI* and contemporaneous stock prices. The coefficient on $PEPS_{i,t}/P_{i,t-1}$ is significantly negative in most specifications, and the coefficient on $PEFV_{i,t}/P_{i,t-1}$ is significantly positive in all specifications. This result indicates that the contemporaneous stock market values *PE*'s current portion negatively and its future value portion positively. The results support the conjecture that high *PEFV* (i.e., high human capital investment) is, at least partially, reflected in contemporaneous prices.

Table 14: *PEFV* and contemporaneous stock prices

	Dependent variable:							
	$P_{i,t}/P_{i,t-1}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$OIPS_{i,t}/P_{i,t-1}$	0.735***	0.712***	0.717***	0.631***	0.332***	0.397***	0.461***	0.336***
	(0.058)	(0.058)	(0.059)	(0.051)	(0.073)	(0.095)	(0.096)	(0.074)
$PEPS_{i,t}/P_{i,t-1}$	−0.532***	−0.506***	−0.521***	−0.518***	−0.236***	−0.330***	−0.049	−0.230***
	(0.042)	(0.057)	(0.056)	(0.045)	(0.073)	(0.099)	(0.118)	(0.073)
$\boldsymbol{PEFV_{i,t}/P_{i,t-1}}$			**0.055****	**0.043****	**0.046****	**0.041***	**0.073*****	
			(0.022)	**(0.021)**	**(0.019)**	**(0.022)**	**(0.027)**	
PEFV-Decile $\boldsymbol{_{i,t}}$								**0.053****
								(0.025)
$EPS_{i,t}/P_{i,t-1}$					2.136***	2.090***	2.767***	2.132***
					(0.242)	(0.259)	(0.376)	(0.242)
$SGAPS_{i,t}/P_{i,t-1}$						0.071***		
						(0.023)		
$RNDPS_{i,t}/P_{i,t-1}$						1.376***		
						(0.322)		
$log(\#E)_{i,t}$	−0.012***	−0.013***	−0.006	−0.005	−0.010**	−0.009*	−0.105***	−0.015***
	(0.003)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.015)	(0.004)
Intercept	1.033***	1.037***	0.962***	1.185***	1.073***	1.024***	2.020***	1.089***
	(0.058)	(0.055)	(0.056)	(0.069)	(0.055)	(0.059)	(0.149)	(0.050)
FF12 dummies				Yes	Yes	Yes		Yes
Year dummies				Yes	Yes	Yes	Yes	Yes
Firm dummies							Yes	
Observations	21,708	11,009	11,009	11,009	8,826	5,960	8,826	8,826
Adjusted R^2	0.097	0.095	0.100	0.341	0.411	0.420	0.491	0.409

This table reports the results of OLS regression of contemporaneous stock price on *PEFV* following equation (11). All variables are defined in Appendix B.1. Columns (1) and (2) show the sample reduction from all firm-years where *PEFV* is calculated to all firm-years with *PEFV* larger than zero. Column (3) presents the inclusion of *PEFV* and column (4) shows the effect of including industry and year dummies. In columns (5) to (8), the sample is further reduced regarding analyst earnings per share forecast availability. Column (6) presents the inclusion of controls for SG&A per share and R&D per share as in Banker et al. (2019). Column (7) shows results for firm and year dummies. Column (8) shows deciles of *PEFV* rescaled to range from zero to one as an alternative measure. Two-way-cluster robust standard errors, clustering at the firm and year levels, are shown in parentheses. Industry, year and firm dummies are not shown. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

In this table and those that follow, we follow Banker et al. (2019) and exclude negative *PEFV* firm-years from the analysis to mitigate the effect of measurement errors in *PEFV*. We further exclude firm-years from industries with zero lags (i.e., zero *PEFV*) to capture the contemporaneous pricing effect of relative differences in *PEFV*. Comparing columns (1) (which includes all firm-year observations) and (2) (which makes the above exclusions), we find that reducing the sample to include only *PEFV* values larger than zero does not substantially affect the coefficients on $OIPS_{i,t}/P_{i,t-1}$ and $PEPS_{i,t}/P_{i,t-1}$. Columns (3) and (4) show results for the effect of *PEFV* without and with the inclusion of industry and year fixed effects. Column (5) shows that the pricing coefficient on *PEFV* remains significantly positive after we include the contemporaneous analyst forecast for earnings per share. This result suggests that investors make *PE*-related adjustments to analyst forecasts and do not necessarily take them at face value.

Column (6) shows that the results are robust to the inclusion of SG&A per share as well as R&D per share as in Banker et al. (2019). Column (7) presents results for the inclusion of year and firm indicators as an alternative fixed effects specification. This specification increases the magnitude of the positive pricing coefficient of *PEFV* (compared with column (5)) and turns the pricing coefficient of current *PE* insignificant. Finally, in column (8), we consider the deciles measure of *PEFV* that we use in Table 13 as an alternative, which also has a significantly positive pricing coefficient. The results presented in Table 14 provide strong evidence that investors contemporaneously recognize some of the human capital investment made by firms.[99]

[99]While Banker et al. (2019) find that the future value of SG&A (*SGAFV*) is positively associated with contemporaneous and future returns, due to its required disclosure, *PE* is more broadly available for IFRS firms than is SG&A. Given that personnel expense is likely to make up a significant portion of SG&A, we examine whether our results are robust to including *SGAFV* in our analysis for the subset of firms that disclose SG&A, using the methodology described in Section 2 of the chapter. Appendix B.2 shows the results of regressing the contemporaneous price on *SGAFV* and control variables. Column (1) shows that the calculation of *SGAFV* is meaningful in the sense that there also is a positive pricing coefficient as in Banker et al. (2019). In column (2), *PEFV* is included and shows a positive pricing coefficient while the coefficient for *SGAFV* turns insignificant. This analysis suggests that *PEFV* is incrementally informative to *SGAFV* in relation to contemporaneous price changes and further stresses the importance of understanding human capital investment for valuation purposes.

3.4.2 Analysts' forecast errors and *PEFV*

Next we examine the relation between analysts' earnings forecast errors and *PEFV*. Given that the information contained in *PEFV* is not directly observable and contains uncertainty, as well as the negative mechanical relation between *PE* and earnings, it is plausible that analysts do not correctly forecast earnings when firms invest heavily in human capital. We first look at the relation between *PEFV* and the absolute value of the contemporaneous mean forecast error in a specification similar to the contemporaneous return analysis.

In column (1) of Table 15, we regress the absolute difference between reported earnings per share and the mean analyst consensus forecast on *PEFV*, operating income per share, and *PE* per share, scaled by the beginning of year share price. We add the number of analysts following the firm to control for analysts' attention to the firm, as well as the controls included in the prior analysis. The coefficient on *PEFV* is positive and significant in columns (1) through (3), indicating that analysts are less able to anticipate earnings of firms that invest more in human capital. In column (2), we add controls for the change in operating income and *PE* to capture year-over-year surprises in these measures. We add the change in SG&A and R&D in column (3) and continue to find the positive effect for *PEFV*. We repeat those analyses using signed forecast errors in columns (4) through (6). Finally, in column (7), we show that we get similar results when using the alternative *PEFV* deciles measure. Taken together, these results indicate that analysts do not fully incorporate investment in human capital into their forecasts, and that they are, on average, pessimistic in their forecasts, possibly because human capital investments are not directly disclosed but mechanically reduce earnings in the current period due to the expensing of *PE*.

3.4.3 Future portfolio returns

Having established that markets put a positive contemporaneous pricing coefficient on high human capital creation and that analysts seem to underestimate its influence on earnings, we investigate the effect of human capital investment on firms' future returns.

Table 15: *PEFV* and contemporaneous forecast errors

	Dependent variable:						
	$\lvert FE_{i,t}\rvert/P_{i,t-1}$			$FE_{i,t}/P_{i,t-1}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PEFV $_{i,t}$**/*P*** $_{i,t-1}$	**0.013*****	**0.017*****	**0.015*****	**0.005****	**0.006****	**0.006****	
	(0.002)	**(0.003)**	**(0.004)**	**(0.002)**	**(0.002)**	**(0.003)**	
PEFV-Decile $_{i,t}$							**0.010*****
							(0.003)
$OIPS_{i,t}/P_{i,t-1}$	−0.175***			−0.205***			
	(0.033)			(0.031)			
$PEPS_{i,t}/P_{i,t-1}$	0.211***			0.218***			
	(0.038)			(0.036)			
$\Delta OIPS_{i,t}/P_{i,t-1}$		−0.041	−0.047		−0.081**	−0.078**	−0.080**
		(0.036)	(0.033)		(0.035)	(0.031)	(0.035)
$\Delta PEPS_{i,t}/P_{i,t-1}$		−0.052	−0.050		0.017	−0.013	0.015
		(0.053)	(0.034)		(0.049)	(0.033)	(0.048)
$\Delta SGAPS_{i,t}/P_{i,t-1}$			−0.018			0.005	
			(0.022)			(0.023)	
$\Delta RNDPS_{i,t}/P_{i,t-1}$			−0.415***			−0.319**	
			(0.116)			(0.142)	
log(#Analysts) $_{i,t}$	−0.001	−0.012***	−0.011***	0.001	−0.003	−0.002	−0.004*
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
log(#E) $_{i,t}$	−0.004***	0.0004	−0.0001	−0.005***	−0.003***	−0.004***	−0.004***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Intercept	0.060***	0.048***	0.059***	0.036***	0.027***	0.029***	0.026***
	(0.010)	(0.008)	(0.010)	(0.007)	(0.006)	(0.010)	(0.006)
Observations	8,672	8,672	5,739	8,672	8,672	5,739	8,672
Adjusted R^2	0.158	0.100	0.118	0.113	0.055	0.066	0.055

This table reports the results of OLS regression of contemporaneous forecast errors on *PEFV*. All variables are defined in Appendix B.1. Absolute forecast errors are the dependent variable in columns (1) to (3). Signed forecast errors are the dependent variable in columns (4) to (7). Column (7) shows deciles of *PEFV* rescaled to range from zero to one as an alternative measure. Two-way-cluster robust standard errors, clustering at the firm and year levels, are shown in parentheses. Industry and year dummies are not shown. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

To do so, we conduct future portfolio returns analyses where the portfolios are formed based on human capital investment (*PEFV*). Our main results are based on the five-factor model (Fama and French, 2015) as follows:

$$\begin{aligned} R_{p,\tau} - R_{f,\tau} = \alpha + \beta_{market}(R_{m,\tau} - R_{f,\tau}) + \beta_{size}SMB_\tau + \\ \beta_{value}HML_\tau + \beta_{profit}RMW_\tau + \beta_{invest}CMA_\tau + \varepsilon_{p,\tau}. \end{aligned} \tag{12}$$

This model consists of the three factors for general market risk, firm size, and value-growth plus two additional factors for operating profitability robustness and investment aggressiveness (Fama and French, 1993, 2015).[100] We form the portfolios at the end of June of year *t+1*, assuming that year *t*'s fiscal results are disseminated by then. We calculate equal- and value-weighted monthly returns on the portfolios ($R_{p,\tau} - R_{f,\tau}$) for

[100] We obtain all data for the factor returns from the monthly European five-factor files on Kenneth French's data library (French, 2019).

the subsequent twelve months, i.e., from July of year *t+1* to June of year *t+2* (e.g., Fama and French, 1992).[101]

The variable of interest in equation (12) is the intercept α which measures the abnormal return. If stock market participants fail to fully incorporate the impact of human capital investment on future performance, then α will increase with portfolios built from higher quintiles (i.e., those with greater human capital investment). $R_{p,\tau}$ is the return on portfolio p in month τ. The coefficient on $R_{m,\tau} - R_{f,\tau}$ captures the portfolio's exposure to the general market risk premium over the risk-free interest rate with $R_{m,\tau}$ being the value-weighted market return and $R_{f,\tau}$ being the rate of the one-month Treasury bill. The coefficient on SMB_τ measures exposure to the size premium. The coefficient on HML_τ measures association with the value-growth factor where portfolios are built with book-to-market quantiles. The coefficient on RMW_τ captures exposure to a factor that measures robustness of firms' operating profitability. Finally, the coefficient on CMA_τ estimates the association with the investment aggressiveness factor.

[101] The factor returns take the perspective of a U.S. investor, thus we measure all returns in U.S. dollar (e.g., Fama and French, 2017). We obtain monthly stock-related Thomson Reuters Datastream items for these analyses.

Table 16: Future monthly returns in the year from July to June after portfolio formation

	PEFV						*PEFV ∗ PE/TA*					
	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	*LS*	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	*LS*
Panel A: Value-weighted:												
Intercept	−.05	.18	.20	.11	.49***	.54***	−.06	.06	.03	.55***	.59***	.65***
	(.10)	(.12)	(.16)	(.19)	(.14)	(.18)	(.09)	(.12)	(.12)	(.16)	(.18)	(.22)
$R_{m,\tau} - R_{f,\tau}$	1.06***	1.01***	1.06***	1.03***	.97***	−.09*	1.02***	1.02***	1.00***	1.00***	1.06***	.03
	(.03)	(.03)	(.04)	(.05)	(.03)	(.05)	(.03)	(.03)	(.03)	(.04)	(.04)	(.05)
SMB_τ	.15**	.12*	.23***	.12	.002	−.15	.10*	.07	.25***	−.01	.42***	.31***
	(.06)	(.07)	(.09)	(.11)	(.08)	(.10)	(.06)	(.08)	(.07)	(.09)	(.09)	(.11)
HML_τ	−.05	−.10	−.49***	−.45***	−.27***	−.22	−.005	−.06	−.09	−.49***	−.55***	−.54***
	(.10)	(.10)	(.14)	(.17)	(.10)	(.16)	(.10)	(.10)	(.09)	(.10)	(.14)	(.17)
RMW_τ	.15	.08	−.32**	−.17	−.02	−.16	.22**	.06	.02	−.32**	−.17	−.38*
	(.10)	(.11)	(.14)	(.17)	(.13)	(.17)	(.10)	(.09)	(.10)	(.15)	(.15)	(.20)
CMA_τ	.23**	.24**	.26*	−.03	−.02	−.25	.20*	.35***	.28**	−.07	−.03	−.23
	(.11)	(.12)	(.15)	(.19)	(.12)	(.17)	(.11)	(.13)	(.11)	(.14)	(.16)	(.21)
Observations	228	228	228	228	228	228	228	228	228	228	228	228
Adjusted R^2	.92	.91	.84	.81	.88	.09	.93	.90	.89	.85	.84	.20
5^{th}- 1^{st}			**.54*** (.17) Annualized: 6.5%**						**.65*** (.21) Annualized: 7.8%**			
Panel B: Equal-weighted:												
Intercept	−.12	.15	.25***	−.03	.17*	.29***	−.15*	.04	.11	.17*	.25**	.40***
	(.08)	(.10)	(.10)	(.10)	(.10)	(.11)	(.08)	(.09)	(.09)	(.10)	(.11)	(.12)
$R_{m,\tau} - R_{f,\tau}$	1.02***	1.00***	.99***	1.01***	.99***	−.02	1.01***	.98***	.98***	1.05***	1.00***	−.02
	(.02)	(.03)	(.02)	(.02)	(.02)	(.03)	(.02)	(.02)	(.02)	(.02)	(.03)	(.03)
SMB_τ	.78***	.88***	.74***	.91***	.82***	.03	.70***	.84***	.82***	.87***	.92***	.23***
	(.04)	(.04)	(.05)	(.05)	(.05)	(.06)	(.04)	(.04)	(.04)	(.05)	(.06)	(.07)
HML_τ	.27***	.14*	.17**	.25***	.13**	−.13*	.31***	.27***	.20***	.06	.13	−.18**
	(.06)	(.08)	(.07)	(.07)	(.06)	(.07)	(.05)	(.08)	(.06)	(.08)	(.09)	(.08)
RMW_τ	.20***	.10	.01	.19**	.11	−.09	.26***	.14**	.11	.04	.08	−.18*
	(.06)	(.08)	(.08)	(.08)	(.07)	(.09)	(.06)	(.07)	(.07)	(.09)	(.09)	(.10)
CMA_τ	.09	−.02	−.04	−.07	−.08	−.17*	.08	.01	−.09	.01	−.13	−.21*
	(.08)	(.11)	(.08)	(.10)	(.09)	(.09)	(.07)	(.11)	(.09)	(.10)	(.11)	(.11)
Observations	228	228	228	228	228	228	228	228	228	228	228	228
Adjusted R^2	.96	.94	.95	.95	.94	.08	.95	.95	.95	.94	.93	.18
5^{th}- 1^{st}			**.29** (.13) Annualized: 3.5%**						**.40*** (.14) Annualized: 4.8%**			

This table reports monthly abnormal returns of portfolios built around two measures of the future intangible asset value of *PE* following equation (12). All variables are defined in Appendix B.1. Portfolios are formed at the end of June each year *t+1* by assigning firms into five quintiles based on *PEFV* or *PEFV*PE/TA*. *PEFV* and *PEFV*PE/TA* are measured at firms' fiscal year ending in year *t*. Robust standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

We form portfolios based on two measures for the future intangible asset value of *PE*: *PEFV* and *PEFV*PE/TA* (Banker et al., 2019). The latter interacts *PEFV* with current *PE* (*PEFV*PE/TA*) to implicitly combine the historically estimated intangible asset investment with the current opportunity set, where *PE* is the proxy for opportunity. *PEFV*PE/TA* is similar to the measure of capitalized R&D used in Chan et al. (2001) in that it creates a proxy for an intangible asset that depreciates over time. Since the portfolio analyses are supposed to capture abnormal return variation dependent on the variation in *PEFV*, we rely on the sample of firm-years with *PEFV* larger than zero (11,009 observations) to create these portfolios.[102]

Table 16 reports the results for the main quintile portfolio analyses in the 12 months after portfolio formation. Panel A shows value-weighted returns for both *PEFV* and *PEFV*PE/TA* and Panel B shows equal-weighted returns. In all specifications, the abnormal returns after controlling for the risk factor model are negative for the first quintile portfolios and significantly positive for the fifth quintile. The long-short returns are statistically and economically significant in all specifications. The annualized long-short value-weighted (equal-weighted) returns are 6.5% (3.5%) for *PEFV* and 7.8% (4.8%) for *PEFV*PE/TA*. The pattern of exposure to the risk factors indicates that high human capital creation firms are smaller and are growth (rather than value) firms with less robust profitability and more aggressive investments.[103] These results provide evidence that the market does not fully capture firms' variation in human capital creation, and this failure to impound the impact of human capital is strongest when examining the combination of the historic human capital investment and the opportunity to invest in human capital.

3.4.4 Additional analyses

Alternative portfolio formations and measures of abnormal returns

[102] We build the portfolios for the first time at the end of June 2000, such that we have a minimum of 40 firms per portfolio. Accordingly, we have 228 months in the portfolio analyses from July 2000 to June 2019.

[103] There is a strong pattern of lower exposure to the value factor for higher quintiles, with even strongly negative exposure to the factor for high quintiles when value-weighting the returns. This is in line with portfolio results for firms with high employee satisfaction reported by Edmans (2011). We remove the value factor from the model in the next table.

Our methodology to measure the human capital creation is FF12-industry-specific regarding the optimal lag structure. However, we build the portfolios by sorting the firms across all industries. As a primary robustness analysis, we follow Eisfeldt and Papanikolaou (2013) and build the portfolios per industry. Panel A of Table 17 shows the results for value-weighted *PEFV*PE/TA* portfolios.[104] We continue to get significant abnormal returns, but the magnitude is smaller than before.

Further, our sample consists of firms from countries with many different currencies, all of which are converted so that the analysis takes the perspective of an investor denominating returns in U.S. dollars. Some of these currencies (i.e., the Hungarian Forinth) are illiquid, which may lead to strong fluctuations in the exchange rate between the respective currency and the dollar. Such strong fluctuations may have an impact on the return measurement, impacting the results of the portfolio analyses. To mitigate this concern, we reduce the sample to firms from countries with highly liquid traded currencies (i.e., the Euro and the British Pound) and redo the portfolio formation with this subsample of firms.[105] Panel B of Table 17 shows that we find even stronger abnormal returns when doing so. In untabulated results, we also find stronger results when we remove penny stocks from the portfolios before calculating the returns as in Cohen et al. (2013).

The five-factor model that we use in our main analysis should be most suitable to analyze the risk return profile of portfolios based on an investment characteristic like human capital. It is intuitive that this produces negative long-short exposure to HML_τ, RMW_τ, and CMA_τ. Nevertheless, we remove HML_τ in Panel C and continue to find significant abnormal returns. Furthermore, our main model does not control for momentum in stock returns. We therefore corroborate our findings with a six factor model that adds the momentum factor (MOM_τ) to the main model in Panel D.

Finally, we disregard the optimal lag structure per industry and apply the same number of lags to firms across all industries. This allows firms from all industries to compete for

[104] As Table 16 shows, value-weighted *PEFV*PE/TA* portfolios generally produce stronger returns. We focus on this specification in the robustness analyses. We get similar but mostly weaker results when equal-weighting the returns or looking at *PEFV* only.

[105] In this analysis we focus on firms from the UK (British Pound) and from the countries that adopted the Euro in 1999, i.e., Austria, Belgium, Germany, Finland, France, Ireland, Italy, Luxembourg, Portugal, Spain and The Netherlands.

Table 17: Robustness analyses regarding portfolio returns

	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	LS
Panel A: Portfolio assignment per FF12-industry						
Intercept	.001	.14	.04	.45**	.36**	.36*
	(.09)	(.15)	(.11)	(.17)	(.16)	(.20)
Observations	228	228	228	228	228	228
Adjusted R^2	.93	.86	.89	.84	.87	.15
5th- 1st	**.36* (.19) Annualized: 4.3%**					
Panel B: Firms from countries with liquid currencies						
Intercept	−.04	−.05	−.08	.41**	.66***	.70***
	(.10)	(.13)	(.13)	(.18)	(.22)	(.25)
Observations	228	228	228	228	228	228
Adjusted R^2	.92	.88	.89	.83	.80	.24
5th- 1st	**.70*** (.24) Annualized: 8.4%**					
Panel C: Factor model without value factor						
Intercept	−.06	.04	.01	.44**	.47**	.53**
	(.09)	(.12)	(.12)	(.17)	(.20)	(.23)
Observations	228	228	228	228	228	228
Adjusted R^2	.93	.90	.89	.84	.82	.14
5th- 1st	**.53** (.22) Annualized: 6.4%**					
Panel D: Factor model with momentum factor						
Intercept	−.02	.09	.02	.58***	.56***	.58***
	(.09)	(.12)	(.13)	(.16)	(.18)	(.21)
Observations	228	228	228	228	228	228
Adjusted R^2	.93	.90	.89	.85	.84	.23
5th- 1st	**.58*** (.20) Annualized: 7.0%**					
Panel E: Two year lag structure across all industries						
Intercept	−.07	.02	.001	.48***	.22	.29*
	(.09)	(.12)	(.11)	(.13)	(.15)	(.18)
Observations	228	228	228	228	228	228
Adjusted R^2	.93	.89	.90	.90	.87	.15
5th- 1st	**.29* (.18) Annualized: 3.5%**					
Panel F: Three year lag structure across all industries						
Intercept	−.02	.01	−.07	.25*	.41**	.43**
	(.10)	(.15)	(.14)	(.14)	(.17)	(.19)
Observations	216	216	216	216	216	216
Adjusted R^2	.94	.91	.90	.89	.85	.06
5th- 1st	**.43** (.20) Annualized: 5.1%**					

This table reports monthly abnormal returns of value-weighted portfolios based on *PEFV*PE/TA* for several robustness analyses. Panel A shows results for assigning firms to portfolios per industry. Panel B reports results of reducing the sample to firms from countries with highly liquid traded currencies (i.e., EUR and GBP). Panel C excludes the value factor from the factor model. Panel D supplements the factor model with the momentum factor. Panel E (Panel F) shows results for applying the same lag structure of two (three) years to firms from all industries. Coefficients on the risk factors are not reported. Robust standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

high *PEFV* on equal grounds. We consider two and three lags and present results for the portfolio returns in Panel E and Panel F. We still get significant long-short returns for both the two- and three-lags specifications. Future research may consider the three lag model as a viable alternative to the industry-specific optimality.

Long-term portfolio returns

To further investigate duration and persistence of the abnormal returns, we analyze portfolio returns up to three years after portfolio formation. In Table 18, we observe that the sort on *PEFV*PE/TA* still produces abnormal long-short returns in the second year after portfolio formation of 5.1%, down from 7.8% in the first year. The returns eventually turn insignificant in the third year (3.1%). Untabulated results for *PEFV* show that the annualized figures evolve from 6.5% to 4.6% to insignificant 1.2%. *PEFV* estimates firms' (historic) investment in human capital. It appears meaningful that sorting on this variation leads to abnormal returns in the earlier years after a high investment and decreases in later years.[106]

Cross-sectional future returns

In our main tests, we report abnormal returns in line with standard approaches in the accounting and finance literature. However, portfolio models do not control for other effects on returns, such as firm-specific momentum, accruals, and other investment characteristics. We corroborate our findings with analyses of cross-sectional future returns. Table 19 reports that monthly returns for the one-year-ahead period, using Fama-Macbeth regressions, are statistically significant and positively associated with our measures of human capital creation even after controlling for R&D and SG&A, and when analyzing returns in excess of the industry-mean return (Fama and MacBeth, 1973).

Alternative approach to capture human capital creation

PEFV, the study's main measure of human capital creation implicit in *PE*, identifies firms that generate more future benefits from investing in their personnel. To this end, our measure commingles measuring how much firms engage in *ex ante* uncertain investments

[106]Untabulated results show strong persistence for a simple sort on *PE/TA* where the magnitude stays rather stable in the three years after measurement. As *PE/TA* is strongly auto-correlated, these results indicate that this measure might be a proxy for some systematic risk characteristic that is not captured by the five factor model. We leave this conjecture for future research.

Table 18: Long-term portfolio returns

	Year 1				
	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
Intercept	−.06	.06	.03	.55***	.59***
	(.09)	(.12)	(.12)	(.16)	(.18)
Observations	228	228	228	228	228
Adjusted R^2	.93	.90	.89	.85	.84
5^{th}- 1^{st}	**.65*** (.21) Annualized: 7.8%**				
	Year 2				
	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
Intercept	−.03	.04	.14	.55***	.40**
	(.11)	(.14)	(.14)	(.16)	(.16)
Observations	216	216	216	216	216
Adjusted R^2	.93	.88	.89	.83	.87
5^{th}- 1^{st}	**.43** (.19) Annualized: 5.1%**				
	Year 3				
	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
Intercept	−.0004	.14	−.04	.37**	.25
	(.10)	(.14)	(.12)	(.17)	(.16)
Observations	204	204	204	204	204
Adjusted R^2	.93	.89	.92	.89	.85
5^{th}- 1^{st}	**.25 (.19) Annualized: 3.1%**				

This table reports monthly abnormal returns of value-weighted portfolios based on *PEFV*PE/TA* up to the third year after portfolio formation following the same procedure as in Table 16. Coefficients on the risk factors are not reported. Robust standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

in personnel with how well this investment turns into benefits. An alternative approach would be to restrict the analysis to the *ex ante* investment to also include investment that was initially intended to, but ultimately did not produce future benefits (Kanodia et al., 2004). To show that our results are not sensitive to potential measurement bias inherent in *PEFV*, we adapt the methodology developed by Enache and Srivastava (2017) to our personnel expense setting, employing the following regressions:

$$PE/TA_{i,t} = \alpha + \beta_1 Sales/TA_{i,t} + \beta_2 SalesDecrease_{i,t} + \beta_3 Loss_{i,t} + \varepsilon_{i,t}, \tag{13}$$

and

$$PEInvest_{i,t} = PE/TA_{i,t} - \beta_{1,Ind,t} Sales/TA_{i,t}. \tag{14}$$

Table 19: Cross-sectional future monthly returns

	Dependent variable:					
	$(R_{i,\tau} - R_{f,\tau})_{t+1}$				$(R_{i,\tau} - R_{Ind,\tau})_{t+1}$	
	(1)	(2)	(3)	(4)	(5)	(6)
PEFV-Quintile $_{i,t}$	**.085*****		**.081*****		**.080*****	
	(.026)		**(.030)**		**(.030)**	
PEFV*PE/TA-Quintile $_{i,t}$		**.210*****		**.168*****		**.167*****
		(.029)		**(.036)**		**(.036)**
$Momentum_{-1,0}$	$-.032^{***}$	$-.033^{***}$	$-.034^{***}$	$-.034^{***}$	$-.036^{***}$	$-.037^{***}$
	(.007)	(.007)	(.008)	(.008)	(.008)	(.008)
$Momentum_{-12,-1}$	$.007^{*}$	.006	.004	.004	.004	.003
	(.004)	(.004)	(.005)	(.005)	(.005)	(.005)
$Accruals_{i,t}$	.015	$-.008$	$-.200$	$-.200$	$-.186$	$-.199$
	(.145)	(.144)	(.181)	(.182)	(.182)	(.182)
$AssetGrowth_{i,t}$	$-.157$	$-.153$	$-.423^{*}$	$-.409^{*}$	$-.410^{*}$	$-.391^{*}$
	(.181)	(.182)	(.216)	(.216)	(.210)	(.209)
$log(BE/ME)_{i,t}$	$.380^{***}$	$.460^{***}$	$.604^{***}$	$.640^{***}$	$.607^{***}$	$.645^{***}$
	(.075)	(.075)	(.092)	(.093)	(.087)	(.089)
$log(ME)_{i,t}$	$.287^{***}$	$.329^{***}$	$.309^{***}$	$.332^{***}$	$.306^{***}$	$.329^{***}$
	(.036)	(.038)	(.041)	(.042)	(.041)	(.042)
$SGA/TA_{i,t}$			$.568^{*}$	.354	$.519^{*}$	.299
			(.304)	(.309)	(.309)	(.315)
$RND/TA_{i,t}$			5.272^{***}	4.641^{***}	5.195^{***}	4.676^{***}
			(1.561)	(1.551)	(1.441)	(1.458)
$EBITDA/TA_{i,t}$			2.424^{***}	2.420^{***}	2.395^{***}	2.411^{***}
			(.685)	(.689)	(.663)	(.669)
Intercept	-1.241^{***}	-1.844^{***}	-1.608^{***}	-1.940^{***}	-2.325^{***}	-2.667^{***}
	(.437)	(.430)	(.501)	(.498)	(.446)	(.454)
Observations	109,908	109,908	72,824	72,824	72,824	72,824
R^2	.290	.291	.310	.311	.093	.094

This table reports results from average Fama and MacBeth (1973) regression coefficients for monthly returns regressed on various firm and return characteristics. The monthly returns are from July *t+1* to June *t+2*. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Using regression (13), we regress firms' total personnel expenditure on total sales scaled by average total asset, which is a proxy for current output, per industry and year. We also include dummies for firm-years with decreases in sales and negative earnings. We then use the industry-year-specific betas to subtract the portion of *PE* that supports current operations (i.e., the portion that varies with current sales) from total *PE*, leaving the portion of *PE* that should generate benefits in future periods ("*PEInvest*" in equation (14)). As before, we build portfolios around *PEInvest* in June of year *t+1* and measure abnormal returns after controlling for the five factor model from July of year *t+1* to June of year *t+2*. Table 20 shows economically and statistically significant abnormal returns

(5.8%) when assigning portfolios cross-sectionally (Panel A) and statistically insignificant returns (3.0%) when assigning portfolios per industry (Panel B).[107] Interestingly, in untabulated results, we find that the abnormal returns for this approach increase in the second year after portfolio formation (i.e., 6.4% and 5.6% for the two approaches) which is different from the pattern in Table 18. This is in line with *PEInvest* serving as a proxy for initial investment in human capital, whereas *PEFV* already captures the efficacy of that investment.

Table 20: Portfolios for alternative methodology to extract investment portion of *PE*

	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
Panel A: Portfolios around *PEInvest*					
Intercept	.01	.18*	.05	.12	.49***
	(.11)	(.10)	(.10)	(.11)	(.15)
Observations	258	258	258	258	258
Adjusted R^2	.93	.94	.93	.91	.89
5^{th}- 1^{st}	**.48*** (.18) Annualized: 5.8%**				
Panel B: Portfolio assignment per FF12-industry					
Intercept	−.03	.12	.04	.40***	.22*
	(.11)	(.09)	(.09)	(.11)	(.13)
Observations	258	258	258	258	258
Adjusted R^2	.94	.94	.95	.92	.89
5^{th}- 1^{st}	**.25 (.17) Annualized: 3.0%**				

This table reports monthly abnormal returns of value-weighted portfolios based on *PEInvest*. Coefficients on the risk factors are not reported. Robust standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

3.5 Conclusion

We develop a strategy to examine aspects of the intangible human capital investment embedded in a firm's personnel expense. We find that our proxy for human capital investment efficacy, *PEFV*, is positively associated with firm characteristics, such as growth

[107]For these analyses, we use our initial sample of 64,579 firm-years before requiring data availability in previous years. In line with our main methodology, we run the industry-year-specific models on the FF12-industry-level. Accordingly, we build the portfolios within FF12-industries in Panel B. We obtain similar results when we use FF48-industries as in Enache and Srivastava (2017). As *PEInvest* can be negative for some firm-years, we also obtain similar results when focusing on the positive firm-years in line with our main methodology.

opportunities and size, consistent with investment in the construct we seek to measure. Still, disclosures around human capital are limited and opaque. Given the magnitude of the underlying expenditure, we explore whether this opacity hinders price discovery. We show that the contemporaneous stock market prices *PE*'s current portion negatively and its future value portion positively. We next document that risk-adjusted abnormal returns can be earned on portfolios formed on two aspects of the future intangible asset value of *PE*: the component of *PE* most likely to represent an investment in human capital, and that component interacted with the opportunity set of potential human capital investment. These findings are robust to model selection and measurement choice.

Our findings are potentially informative to regulators examining how to improve disclosures around human capital. In addition, these insights on the future value generating ability of *PE* lead to questions for future research: Does the legal environment affect how returns to human capital creation are realized (e.g., Shleifer and Vishny, 1997)? Can firms acquire the human capital creating ability of target firms, and does it matter whether merging firms' human capital creating abilities are related (Lee et al., 2018)? Moreover, there are opportunities for research in other contexts. Does *PE* have higher cost stickiness when there is a higher potential to create future values from it (Chen et al., 2012)? Do firms with high human capital creating ability grant more long-term executive compensation incentives (Banker et al., 2011), and is executive compensation shielded from the negative effects of expensing personnel expenditures when they create higher future values (Huson et al., 2012)?

4 *Far*-Out-of-Sample Accounting Misconduct Prediction – Application to Non-U.S. Cases

Abstract

Most jurisdictions around the world do not have strong institutions in place to identify and enforce misconducted accounting. In this paper, we investigate whether financial accounting misconduct models established in the literature can be applied to *far*-out-of-sample prediction. We use the latest and most powerful models that are trained, tuned, and tested on U.S. data, which contains hundreds of identified misconduct instances. We apply these predictors to a European sample of firms that contains 21 accounting misconduct cases relating to 59 misconducted firm-years. We find that simple Logistic Regression models as well as sophisticated RUSBoost Ensemble Learning assign high ex-ante probabilities of misconduct to a large portion of European cases during misconduct years. Comparing the out-of-sample classification performance of European cases to U.S. benchmarks shows our approach to be highly predictive. These results have important implications for global enforcement institutions on how to design a preselection model for overseen firms prior to investigations. Hence, our results enhance academics', regulators', and investors' understanding of misconduct outside of the U.S.

Keywords: Misconduct prediction, Accounting fraud, Machine learning, Ensemble learning
JEL classification: C53, M41

Authors: Dan Amiram (Tel Aviv University), Zahn Bozanic (Florida State University), Matthias Regier, Johannes Roscher (Technische Universität München)
Lead author: Matthias Regier

Status: Working Paper

Acknowledgements We thank Claudia Imperatore (discussant), Arthur Stenzel, and workshop participants at University of St. Gallen, Technical University of Munich, and the 2022 International Accounting Section Midyear Meeting for helpful feedback. We are grateful for excellent research assistance by Pascal Bundschuh, Devy Kusuma Cendana, and Angélica Silva Russi. The lead author initiated this project while being affiliated with Harvard Business School as Visiting Fellow in 2021.

4.1 Introduction

Financial accounting misconduct undermines capital markets' role of efficiently allocating corporate (financial) resources and further impairs stakeholders' trust in corporations (Amiram et al., 2018).[108] The impact of misconduct is so broad that even the stakeholders of competitor firms can be negatively affected (e.g., Gleason et al., 2008). However, most jurisdictions around the world do not have strong institutions in place to identify and enforce deceitful accounting. In this paper, we investigate whether misconduct prediction models can be applied *far*-out-of-sample. Out-of-sample tests usually refer to testing a model's prediction accuracy for the subset of a sample that is not used for training. In the misconduct prediction literature, it is common to partition a sample of U.S. firm-years containing misconduct instances into earlier years used for training and later years used for testing the performance of the trained models. We use the term *far*-out-of-sample to imply an application of U.S.-trained models in a different market.

To do so, we train established models using data from the largest among the most developed financial markets (e.g., Rajan and Zingales, 1998; Wurgler, 2000) in the world, which is the United States, and apply them in markets with weaker accounting misconduct enforcement.[109] While improving processes for misconduct prediction is relevant for many market participants, our approach is best understood by taking on the enforcer's perspective. We envision an enforcement entity that preselects the firms with the highest predicted misconduct probability out of all firms it oversees for more in-depth investigation.[110]

Basing such preselection on predicted probabilities from a quantitative model is not uncommon. For instance, for almost a decade the U.S. Securities and Exchange Commission

[108] While most of the related literature uses the term *fraud*, we use the term *misconduct* throughout the paper in line with Amiram et al. (2018) suggesting this to be a better characterization of both the underlying concept and the way it is empirically identified.

[109] U.S. firms are likely subject to the toughest enforcement (e.g., Brown et al., 2014; Prentice, 2005). We test our approach for European firms that are generally exposed to similar institutional settings in terms of auditing of financial statements and enforcement of compliance that are, however, weaker. For instance, Brown et al. (2014)'s proxies for *audit* and *enforcement* show that (with the exception of the United Kingdom) major European countries have weaker institutional settings.

[110] While most countries have national stock market authorities that may consider implementing such an approach, supranational entities like the recently founded European Securities and Markets Authority (ESMA) may provide such an enforcement mechanism across countries.

(SEC) has implemented an Accounting Quality Model (AQM) as a tool for preselection prior to engaging in more rigorous examinations.[111] In the spirit of AQM, our paper takes this approach a step further by assessing whether U.S.-trained models are informative for misconduct instances outside U.S. markets. In so doing, we address a significant problem in that non-U.S. enforcement entities, because of either lack of regulatory resources or priority, do not possess their own rich datasets of historic cases to learn from and train models.[112] Over time, these entities may identify (more) cases (earlier) and refine the adopted U.S. models accordingly. If enforcement entities are well-equipped to identify cases, they may at minimum increase misconduct deterrence since corporations appear to take enforcement resources and efforts into account (e.g., Kedia and Rajgopal, 2011).

The possible motivations for accounting misconduct can be characterized by the following: i.) perceived pressure (e.g., by financial markets) or incentives (e.g., compensation plans) to engage in misconduct, ii.) the perceived opportunity to do so (e.g., lack of internal or external controls), and iii.) the rationalization employed to justify the behavior in front of those involved or those who might eventually find out.[113] Stronger enforcement, through either ex ante deterrence or ex post restitution and/or incarceration, both of which may be facilitated by employing our approach, should mainly affect opportunities and rationalization. However, understanding the incentives is an important aspect of such strong institutions. In line with this interdependence, the strong U.S. enforcer usually also identifies the original motives behind misconducted corporate accounting.[114]

The reasons for the enforcement inferiority of non-U.S. enforcers are manifold. Regulatory institutions outside of the U.S. are less effective at detection and deterrence than the SEC due to fewer resources and/or limited power. Appendix C.3 describes the en-

[111]We refer to the U.S. SEC's Accounting Quality Model risk score, termed 'RoboCop' by the business press, developed by the Division of Risk, Strategy and Financial Innovation's Office of Quantitative Research in 2012. RoboCop focuses on discretionary accruals to identify earnings management.

[112]While there are ideas for leading indicators of misconduct, such as deviations from distributional properties of numbers (Amiram et al., 2015), which would not require rich training datasets, basically all advances in this field rely on training models with historic U.S. misconduct cases.

[113]Despite frequent criticism, these three elements originally developed by Cressey (1953) and later termed 'fraud triangle' have been the basis for theoretical work in forensic accounting research for decades (e.g., Huber, 2017).

[114]For instance, almost every Auditing and Accounting Enforcement Release issued by the U.S. SEC, contains a section on the motives that drove the behavior of the people involved.

forcement background for Germany and the UK, the two largest economies in our *far-out-of-sample* population. German enforcers typically lack both resources and power to adequately investigate potential financial accounting misconduct. While resources may be relatively greater in the UK, their enforcement priorities are on prosecuting criminal behavior outside of accounting misconduct and/or focus more of their enforcement efforts on financial services businesses.

We manually collect our sample of non-U.S. accounting misconduct cases. Starting with accounting scandals listed in Hail et al. (2018), we identify misconduct cases through an exhaustive keyword search in country-specific business newspapers in major European countries. We collect all affected firm-years after 2005's IFRS adoption and end the sample in 2017. We identify 21 firms with 59 misconduct firm-years from eight countries (Germany, France, Italy, Spain, Sweden, Switzerland, The Netherlands, and the UK). After training the model using U.S. cases, we then predict misconduct probabilities out-of-sample for all firms listed in any of those eight countries during the sample period.

In line with the objectives of our envisioned use case, we restrict the variables fed into the prediction models to readily available financial statement items. We use a common set of 28 raw items that firms report and database providers collect around the globe. We report results for two types of established misconduct prediction models utilizing these items. The first is ratio-based Logistic Regression (LR) (Dechow et al., 2011), which is most commonly used in accounting literature, and the second is a more advanced learning algorithm called RUSBoost Ensemble Learning, recently suggested to have similar or superior predictive power (Bao et al., 2020).[115] While the first uses theoretically motivated or empirically established constructs such as the accounting flexibility in accruals (e.g., Richardson et al., 2005) or operating assets (e.g., Barton and Simko, 2002) to calculate predictive ratios from the 28 raw items, the second feeds the raw items into the algorithm and allows decision trees to combine them in search of the optimal model.

At a cutoff at the top 10% of predicted misconduct probabilities, the European out-of-

[115]The RUSBoost vs. LR model performance superiority published in Bao et al. (2020) is criticized to be overstated due to an inconsistency in the misconduct identification data (Bao et al., 2022; Walker, 2021a,b, 2022). None of the results building on Bao et al. (2020) presented in the paper are subject to this inconsistency.

sample results for the LR models have a Sensitivity of between 14 and 22%, slightly higher than the results for the U.S. benchmark prediction, which is between 12% and 21%.[116] As it is unlikely that the *far*-out-of-sample application works better than the U.S.-out-of-sample prediction, we infer that our identified cases are probably rather severe and easy to detect. The results for the RUSBoost approach are even stronger for both the U.S. benchmark prediction and our *far*-out-of-sample application with Sensitivity figures around and above 30%. We conclude that both straightforward as well as sophisticated models relying solely on data that is easy to obtain could already do a decent job in predicting misconduct *far*-out-of-sample.

Since we use two different databases to collect the data for the predictors for the training (Compustat) and the European testing (Compustat Global) samples, we address inconsistencies in the two by comparing the figures for firms that are covered by both.[117] When we remove raw items with large deviations from the RUSBoost and affected ratios from the LR models, we observe several developments. As one would expect, the benchmark performance for the U.S. sample mostly declines, because the reduced models are likely less good than the models identified by prior research. For the ratio-based LR models, the prediction performance for the European out-of-sample application stays about the same with both slight improvements and declines. For the raw item-based RUSBoost models, however, the European prediction performance improves. This indicates that the established prediction approaches, in particular most recent advancements like RUSBoost Ensemble Learning, are quite robust to many design choices and data inconsistencies that enforcement entities may face. These results build further confidence in that the *far*-out-of-sample application is a promising approach that can be further adapted in line with

[116] Misconduct predictions are calculated for all firms or firm-years in a sample so that the predictions figures are interpreted in relative terms. To evaluate the performance, one has to assume a cutoff above which (below which) predictions are classified as misconduct (no misconduct). We mainly use 10% as the cutoff in our analyses and provide the reasoning in a later section. The Sensitivity metric is the percentage of actual misconduct observations flagged correctly, i.e., the correct positives divided by the correct positives plus the false negatives.

[117] As the focal issue of our paper is to test the *far*-out-of-sample applicability, we use Compustat for the U.S. data to be able to perform the model training exactly as in earlier papers. Consequently, we have to use the Compustat Global data for the non-U.S. testing to have the best possible comparability. In a real approach, it might be more practical to use Refinitiv data (earlier Thomson Reuters Datastream) that is consistent across U.S. and non-U.S. firms for both training and testing. Despite available coding schemes to align Refinitiv and Compustat, the trained models would likely differ somewhat.

specific data availabilities.

This paper's primary contribution lies in its regulatory implications. We are the first to show that U.S.-trained misconduct prediction models may be applied *far*-out-of-sample to preselect firms for further investigation. This insight offers a potential detection tool to markets that may not have as rich of an enforcement environment as the U.S. Our analysis indicates that the established prediction approaches, in particular recent advancements like RUSBoost, are robust to many design choices and data limitations, which helps to improve applicability and generalizability to other jurisdictions. Second, our results contribute an understanding of the factors determining financial accounting misconduct across different countries. The strength of our results in terms of *far*-out-of-sample prediction performance shows that misconduct has global common factors. This insight again supports the generalizability of relationships identified on U.S. data to other markets and jurisdictions. Third, our work further offers *far*-out-of-sample applications to other prediction contexts. For example, our approach may be used to improve predictions of other rare events like corporate bankruptcy (e.g., Agarwal and Taffler, 2008; Charitou et al., 2004; Jones et al., 2017) or bank failure (e.g., Jin et al., 2011; Tam and Kiang, 1992). In this regard, our approach may become an important tool for global prediction problems in an increasingly interconnected world.

The remainder of this paper is organized as follows. Section 2 of the chapter (i.e., 4.2) describes the collection of the sample of European misconduct cases. Section 3 explains the data we use, which prediction models we apply, and how we classify performance. Section 4 presents the main results and Section 5 reports several robustness analyses. Section 6 concludes the paper.

4.2 Sample of European accounting misconduct cases

Regulators worldwide are in a "never-ending battle with the creative accountant" (Jones, 2010, p. 25). Despite regulators' best efforts in strong enforcement regimes, high-profile cases like Enron or WorldCom occurred and may occur again. Notwithstanding researchers' and regulators' inability to detect all misconduct cases, it appears that U.S.

enforcers, such as the SEC and the Department of Justice, investigate and process a good number of cases each year (e.g., Bao et al., 2020).[118] Outside of the U.S., no jurisdiction has similarly powerful institutions that are as effective in enforcing financial reporting misconduct.[119] Because misconduct case identification occurs through a variety of potential channels (e.g., whistleblowing, analyst research, or during bankruptcy handling), we attempt to systematically identify as many cases as possible. To do so, we begin with the cases collected in Hail et al. (2018). They identify corporate (accounting) scandals through key word search in leading (business) newspapers in many countries through 2015. We collect their identified cases and then augment their sample with more recent misconduct cases following their methodology for the years 2015 to 2021.

Our analysis focuses on European countries in the period after the 2005 mandatory IFRS adoption. We manually analyze every case identified as an accounting scandal in the Hail et al. (2018) online appendix and by our extension into the most recent years through reading the collected source articles as well as additional (secondary) literature available.[120] We focus only on financial reporting misconduct and disregard cases where firms misused stakeholders' assets but did not manipulate financial statements.[121] Moreover, consistent with prior misconduct literature (e.g., Dechow et al., 2011), we focus on listed, non-financial firms and require fiscal year end accounting numbers to be affected.[122] For each case, we identify the years when the misconduct scheme was most likely implemented (e.g., as sentenced by a court ruling). While some cases relate to a specific manipulation in a specific fiscal year, other cases relate to manipulations implemented over several years. Cases enter our final sample when they relate to fiscal year

[118]'Partial observability' is an endemic problem of financial accounting misconduct research, see Amiram et al. (2018) for a discussion of this issue.

[119]Appendix C.3 illustrates the differences in available resources and forensic capabilities between the U.S. and the two largest economies in our European misconduct sample, namely Germany and the UK.

[120]Additional literature comprises additional news articles, documents of legal proceedings, analyst reports, or auditors' investigations.

[121]For instance, in the case of Panaxia Security AB, the security firm misused the cash that it handled for its customers but did not engage in reporting misconduct.

[122]There are several cases where only quarterly or semi-annually issued reports were affected, such as M&C Saatchi PLC, Tesco PLC, or EMI Group Limited. Furthermore, in the case of Europacorp SA, only financial communication regarding the financial outlook and depreciation considerations but no financial statement figures were affected.

2005 or later. We end the sample in 2017.[123]

Our final sample of cases contains 21 firms (with 59 misconducted firm-years) from eight countries. Table 21 lists the names of the identified firms and the corresponding firm-years per country. Eight firms are located in the United Kingdom, three each in Germany and Sweden, two each in France and The Netherlands, and one each in Italy, Spain and Switzerland. Appendix C.4 lists and summarizes the cases that are included in the final sample along with how we arrived at the identification of the misconduct years.

Table 21: Sample of European accounting misconduct cases

Country	Names of the identified firms	Number of firms (firm-years)
Germany	Conergy AG (2006), MIFA Mitteldeutsche Fahrradwerke AG (2012), Wirecard AG (2011-2017)	3 (9)
France	Carrère Group (2005-2006), Safran SA (2005)	2 (3)
Italy	Mariella Burani Fashion Group SPA (2007-2008)	1 (2)
Spain	Pescanova SA (2009-2011)	1 (3)
Sweden	Eltel AB (2015), Eniro AB (2013), Telefonaktiebolaget LM Ericsson (2012-2016)	3 (7)
Switzerland	Panalpina Welttransport AG (2005)	1 (1)
The Netherlands	Innoconcepts NV (2005-2008), Steinhoff Investment Holdings Ltd (2009-2017)	2 (13)
United Kingdom (including one case from Ireland)	Autonomy Corporation PLC (2009-2010), Globo PLC (2008-2014), NMC Health PLC (2014-2017), Patisserie Holdings PLC (2017), Phoenix IT Group (2009-2011), Quindell Business Process Services Ltd / Watchstone Group PLC (2013), Speedy Hire PLC (2011-2012), Worldspreads Group PLC (2009)	8 (21)
Total		21 (59)

This table presents the names of the firms included in the final sample of European accounting misconduct cases. The affected years in brackets relate to the fiscal year end accounts that were manipulated. The misconduct schemes may have started earlier than 2005 or continued after 2017, which is not indicated in this table. Appendix C.4 contains detailed summaries of the accounting misconduct cases in this sample.

[123] It takes some time until misconduct cases are enlightened so that we could include them in our sample. Ending the sample in a later year would therefore increase the risk of including false negatives when predicting misconduct probabilities across European firms including the identified cases. Moreover, at the time of the writing of the paper, 2017 was the most recent year with complete annual financial figures available in Compustat Global.

4.3 Data and methodology

4.3.1 U.S. misconduct sample and European prediction sample

Our training sample methodology mainly draws upon the work by Dechow et al. (2011) and Bao et al. (2020). Similar to these studies, we use U.S. SEC Accounting and Auditing Enforcement Releases (AAERs) to identify misconduct in the training sample.[124] Shifts in U.S. firms' misconduct behavior along with historical changes in the nature of SEC enforcement suggest the first year with strong enforcement to be 1991 (e.g., Atkins and Bondi, 2008; Bao et al., 2020; Beasley et al., 1999, 2010; Erickson et al., 2006).[125] The shift in the U.S. regulators' accounting misconduct enforcement in 2008, coinciding with the global financial crisis, suggests that this is the last year where the SEC was very powerful in identifying misconduct (e.g., Bao et al., 2020; Ceresney, 2013; Rakoff, 2014).[126] We therefore consider the period from 1991 to 2008 to be the one most powerful period that could be utilized for model training purposes.[127]

As our European misconduct cases span the years from 2005 to 2017 and we do not want to create a look-ahead bias, we use the year 2005 in the U.S. data to benchmark our prediction results in our main analyses.[128] Requiring a two-year gap between the last training year and the test year (Dyck et al., 2010) means that the sample used for

[124] The AAERs are provided by the University of California-Berkeley Center for Financial Reporting and Management database, which according to Karpoff et al. (2017) ranks first in comprehensively identifying financial accounting misconduct cases.

[125] Atkins and Bondi (2008) argue that the purpose and power of the SEC shifted towards imposing more punitive actions in the 1990s. Moreover, the use of stock options and other forms of pay-for-performance in executive compensation increased substantially during the 1990s (Erickson et al., 2006), coinciding with more frequent citations of insider trading as a possible motive for the misconduct in the published AAERs (Beasley et al., 1999, 2010). Finally, more subtle techniques of manipulating financial statements like understating expenses or liabilities (as compared to overstating revenues and assets) became more frequent during the 1990s (Beasley et al., 2010).

[126] Already in the aftermath of the 2001 terrorist attacks, the Federal Bureau of Investigation had shifted many agents that were assigned to financial fraud investigation to antiterrorism work (Rakoff, 2014). Moreover, following the financial crisis, the SEC shifted attention away from accounting misconduct to Ponzi-like schemes (Ceresney, 2013; Rakoff, 2014). Finally, in 2009, the Department of Justice made a decision to spread the misconduct investigation work among numerous unexperienced U.S. Attorney's Offices (Rakoff, 2014).

[127] To illustrate the development of the SEC's effectiveness in detecting misconduct, we may look at how many cases we have in the final sample used for training. Beginning with 23 in 1991, there is a somewhat gradual increase to the peak at 84 in 2000, followed by a somewhat gradual decline to 20 in 2008.

[128] Even though we train and test the prediction models with firms from different parts of the world, utilizing training data from years later than the testing years could entail a look-ahead bias in the sense that the models would learn from misconduct techniques of later years that would not have yet been detected in the years of the testing.

training ends in 2003. This means that in our main analyses we train with the U.S. data from 1991 to 2003, predict the U.S. year 2005 for benchmark purposes, and predict all of our European years from 2005 to 2017 based on the same identified predictors. We fully utilize the training years from 1991 to 2008 with a corresponding test year in 2010 in alternative analyses.

Table 22 Panel A lists the samples used in the empirical analyses. Out of the 88,265 firm-years in the U.S. sample from 1991 to 2005, we use 76,819 in the main Training sample until 2003 and 5,688 in the main Test sample of year 2005.[129] The Training sample contains 708 misconducted firm-years that relate to 316 firms. We address serial misconduct of the same firm in both the training and test periods by removing the 42 misconduct firms of the Test sample from the Training sample (Bao et al., 2020; Perols, 2011; Perols et al., 2017). We follow (Dechow et al., 2011) and use only non-financial firms for training and test and remove misconduct observations that relate to an understatement of firms' results.

Table 22: Different samples and descriptive statistics

Panel A: Sample selection

Sample	Firms (misconduct firms)	Firm-years (misconduct firm-years)
Sample of all publicly listed non-financial U.S. firms during the period 1991–2005 used for training and testing in Bao et al. (2020): U.S. sample	13,656 (316)	88,265 (708)
thereof Training sample during the period 1991–2003: (U.S.) Training sample	12,876 (297)	76,819 (612)
thereof Test sample of year 2005: (U.S.) Test sample	5,688 (42)	5,688 (42)
Sample of all firms from same countries and SIC-3 industries as misconduct firms with total assets above 5 million U.S. dollars active during the period 2005–2017 used for misconduct prediction: European sample	659 (21))	7,653 (59)

Panel B: Descriptive statistics

Sample	Variable	N	Min	25%	50%	Mean	75%	Max
U.S. Training sample	at	76,819	0.0	17.4	84.1	1,463.6	454.2	270,717.4
	sale	76,819	-21.8	13.4	81.2	1,235.7	450.6	257,157.0
U.S. Test sample	at	5,688	0.0	20.4	139.7	2,754.3	908.6	244,587.0
	sale	5,688	0.0	9.4	108.2	2,262.8	811.1	328,213.0
European sample	at	7,653	5.1	39.2	128.1	2,150.4	537.1	136,976.4
	sale	7,653	0.0	36.9	121.1	1,441.6	512.9	81,653.6

[129] This is in line with the firm-years used in Bao et al. (2020).

Our European prediction sample covers all non-financial firms listed in any of the eight countries and operating in one of the SIC-3 industries represented in the list of misconduct cases from 2005 to 2017. We choose this industry level as the resulting percentage of misconduct instances in the total number of firm-years is quite in the same range as in the U.S. Test sample to ease comparisons in the prediction performance. The more granular SIC-4 level samples too few firms. The broader SIC-2 level samples too many, which would require to add further variables into the matching. We obtain financial data for the European sample from Compustat Global and convert the figures into U.S. dollars using the respective exchange rates (i.e., EUR, GBP, SEK, CHF, etc.) of the U.S. Federal Reserve Bank. We compare the figures of all our European (non-misconduct) sample firms that are covered by both Compustat and Compustat Global to match the variable definitions in the two databases as close as possible. We address remaining deviations in a later section. We remove firms that have total assets (*at*) of less than 5 million U.S. dollars. We further remove firm-years where either total net sales (*sale*), depreciation and amortization (*dp*), or gross PPE (*ppegt*) is missing. All other variables that we use are set to zero if missing.[130] The final European sample contains 7,653 firm-years of 659 firms including the 21 misconduct firms (59 firm-years) described above. The percentage of misconduct instances in the European sample is 0.77% (i.e., 59/7,653), which is very close to 0.74% for the U.S. Test sample of year 2005 (i.e., 42/5,688).

4.3.2 Misconduct prediction models

Accounting misconduct prediction models use two main kinds of data: numerical and, to a lesser extent, textual. Recent advancements to analyze textual data from corporate filings or conference calls using techniques such as dictionary-based sentiment analysis (e.g., Larcker and Zakolyukina, 2012; Loughran and McDonald, 2011) or purely statistical approaches (e.g., Brown et al., 2020; Purda and Skillicorn, 2015) are promising.

[130] We thereby accept few missing values for *ivst*, *pstk*, *cogs*, and *xint* and many missing values for *sstk* (missing for more than one third of firm-years), because all these variables may reasonably be zero in many firms or firm-years. The variable *dltis* is not populated at all after fiscal year 2004, we therefore remove it in a later section. We also set the financial ratios to zero when they cannot be calculated due to missing values in the raw items. All our inferences hold when we restrict the sample to firm-years with data availability for the four variables with few missing values as well as all financial ratios.

However, both written and spoken language is very much subject to cultural, national, and geographical differences. While it seems far-fetched to translate several pieces of text to a common language for the purposes of misconduct prediction, even though many corporate filings and communication in non-U.S. markets are available in English, they are mostly prepared by non-native speakers, undermining comparability. Moreover, there is very limited commonality regarding the availability of textual corporate data, such as Management Discussion and Analysis sections, between the U.S. and other markets. We therefore focus on misconduct prediction approaches using numerical data.

Beyond financial statement data, researchers have employed a variety of less conventional data - such as insider trading cues (Summers and Sweeney, 1998), quantifications of the relationship with the auditor (Bell and Carcello, 2000), or operational measures such as facilities growth (Brazel et al., 2009) - to assist in predicting misconduct. The complexity of the models ranges from rather straightforward Logistic Regression (e.g., Dechow et al., 2011) to Neural Network Technology (e.g., Green and Choi, 1997). As our envisioned use case implies that enforcement entities apply U.S.-trained models in other markets, we aim for practicability (e.g., Amiram et al., 2018) in both the data fed into the models and the computational understanding and power required to run them. Therefore, we focus on an established literature stream that restricts the data used to a common, readily available set of financial statement variables (Bao et al., 2020; Cecchini et al., 2010; Dechow et al., 2011), which are also comparably available outside the U.S., in different statistical approaches.[131]

Logistic Regression

The two main approaches we use to predict misconduct are a Logistic Regression approach based on financial ratios and a RUSBoost Ensemble Learning approach based on the raw data items that comprise the ratios. We begin our analyses with the straightforward Logistic Regression approach offered by Dechow et al. (2011). After investigating numerous plausible misconduct predictors, Dechow et al. (2011) provide a model that

[131] The analysis of financial statement data has been used to predict firms' financial distress for more than a century (e.g., Beaver et al., 2010). Therefore, we argue that it is relatively straightforward to obtain such data and use it in financial accounting misconduct prediction models.

solely relies on financial statement variables, mostly ratios, which are relatively easy to calculate.[132] Their backward elimination approach arrives at the following seven variables, which we call the basic Dechow et al. (2011) LR model: *rsst_acc* measures the summed changes in non-cash working capital, net non-current operating assets and net financial assets scaled by average total assets (i.e., RSST accruals as in Richardson et al. (2005)). *ch_rec* is the deflated change in receivables and *ch_inv* is the deflated change in inventory. *soft_assets* is the percentage of soft assets (i.e., non-tangible and non-cash assets) to total assets. *ch_cs* measures the change in cash sales and *ch_roa* measures the change in return on assets. Finally, *issue* is a binary indicator of whether the firm issues equity or debt. Appendix C.1 lists all raw financial data items and Appendix C.2 describes the ratio calculation.

Bao et al. (2020) extend the Dechow et al. (2011) model to also include another seven ratios (i.e., 14 variables in total): Working capital accruals (*wc_acc*), change in cash margin (*ch_cm*), book-to-market ratio (*bm*)[133], depreciation index (*dpi*), retained earnings over total assets (*reoa*), EBIT over total assets (*EBIT*), and change in free cash flows (*ch_fcf*). We call this extended model the Bao et al. (2020) LR model.

RUSBoost Ensemble Learning

Logistic Regression is restricted by the assumed linearity between the variables fed into the model and the log odds of the misconduct outcome.[134] Bao et al. (2020) make a compelling case for leaving the determination of associations between the variables to a more powerful and robust classification algorithm. To this end, they also argue to feed the model with raw data items only (as tabulated in Appendix C.1) to allow more flexibility. RUSBoost combines the selection algorithm AdaBoost with a random under-sampling procedure (RUS). During the iterative training process, AdaBoost uses a sequence of

[132]Here are two examples of variables that are not considered: 1) The percentage of short-term executive compensation incentives cannot be found in firms' financial statements, i.e., it is not readily available. 2) While the calculations of (all kinds of) accruals is rather easy, determining discretionary accruals is not.

[133]Bao et al. (2020) argue that adding the book-to-market ratio to the financial statement variables does not harm the intent to only use readily available data because Compustat provides the required market data. As the calculation of the fiscal year end share price (*prcc_f*) and the corresponding common shares outstanding (*csho*) is somewhat more complicated in Compustat Global, we omit the respective variables in a later section.

[134]Although Logistic Regression does not assume dependent and independent variables to be linearly related, it still requires that the independent variables are linearly related to the log odds.

weak classifiers (i.e., small decision trees) that have a performance only slightly better than random choice to classify the firm-years into misconduct or non-misconduct. To gradually improve the model, the weights of the sample are adapted in each iteration to give wrong classifications more importance. Eventually, the algorithm constructs a strong classifier through a weighted average of all the weak classifiers based on their error rate for the training set. To address the rarity of misconduct observations in any training dataset, the RUS procedure extends the Adaboost algorithm with random under-sampling of the non-misconduct observations to balance the number of misconduct and non-misconduct observations in each iteration (Perols, 2011). We use the same tuning specifications (i.e., we run 3,000 iterations with a maximum tree size of 5 and set the seed of the random number generator to 0).

While one benefit of Logistic Regression fitting is that it requires only little computational power (i.e., all models in the paper can be run in a couple of seconds using standard statistical software on personal computers), running 3,000 iterations to train the RUSBoost model does require a couple of minutes. We still argue that such an approach could easily be implemented into and maintained as part of an accounting misconduct investigation process. We further consider another more sophisticated approach using the same raw items as in the RUSBoost model, namely Support Vector Machines with Financial Kernels (Cecchini et al., 2010). This approach produces results that are similar, yet slightly weaker throughout. As this approach is also conceptually and computationally more demanding, we do not discuss the results in detail. Appendix C.5 reports baseline results for this approach.

4.3.3 Classification performance evaluation

We use several measures to evaluate the performance of our *far*-out-of-sample classification approaches. Doing so requires assuming a cutoff above which (below which) predictions on the firm-year level are classified as misconducted (not misconducted). Earlier literature offers two opposing ends of the spectrum for this cutoff. On one end, Dechow et al. (2011) label all observations as misconducted if the scaled misconduct probability exceeds the unconditional expectation i.e., the unconditional probability of a randomly selected observation of the sample being misconducted. This procedure effectively labels more than one third of observations as misconducted. Other literature shows a cutoff as low as 50% next to higher cutoffs (e.g., Brown et al., 2020; Larcker and Zakolyukina, 2012). On the other end of the spectrum, Bao et al. (2020) classify only the top 1% of observations regarding misconduct probability as misconducted. This cutoff corresponds with the literature's assumption that likely about one percent of firms engage in (detected or undetected) misconduct behavior in a given year. Our envisioned out-of-sample application should serve as a meaningful preselection for firms to be investigated more in-depth. Such a preselection would have to balance the likelihood of actual misconducting firms being among the selected firms (i.e., lower cutoff) with being able to cope with the resulting workload (i.e., higher cutoff). We therefore consider the top 10% as a meaningful cutoff.[135] We further restrict the cutoff to the top 5% in later analyses.

Next to the choice of the cutoff, there are different plausible ways to benchmark the predicted misconduct probabilities. First, assuming an institution that oversees a rather small population of firms, it may be suitable to rank the misconduct probabilities relative to a full year of U.S. misconduct probabilities. This allows relatively more or fewer firms to be flagged than when being restricted to a ranking within the small population itself. However, such a relative ranking may be vulnerable to biases when the predicted misconduct probabilities are generally higher or lower in the out-of-sample application, for instance, due to differences in the reporting frameworks (e.g., Barth et al., 2012). We therefore also calculate the percentile ranking within the European sample itself as a second alternative. Third, we calculate the percentiles within each financial year of the European sample, which is closest to our envisioned yearly preselection of firms that are then investigated more in-depth. Performing these alternative ranking procedures next to each other allows drawing general inferences regarding properties of the models when applied *far*-out-of-sample on European firms. While the first alternative may be

[135] The German Financial Reporting Enforcement Panel (FREP), for instance, seeks to carry out examinations in about 80 firms (ESMA, 2020), which corresponds to about 17% of the total number of German listed firms at the time of the writing of this paper. Historical numbers of yearly FREP reviews confirm that range (Pasch, 2017).

very practical, the second and third alternative are likely better indicators of prediction performance.

The first and most straightforward metric, Sensitivity, also known as True Positive Rate, measures the fraction of true misconduct firm-year observations that are correctly identified (e.g., Bao et al., 2020; Dechow et al., 2011). The related metric Precision measures the portion of predicted misconduct observations correctly labelled as misconducted (e.g., Bao et al., 2020; Larcker and Zakolyukina, 2012). Choosing a lower cutoff usually increases the Sensitivity at the cost of lower Precision. In line with the practical use case that we envision, we also report the number of firms (# Firms) where any of the misconducted years shows a probability that is higher than the cutoff.[136] This assumes that it suffices to flag misconduct firms correctly once, so that the following in-depth investigation can figure out the extent of the manipulation.

We further use the Normalized Discounted Cumulative Gain at the position k (NDCG@k, i.e., NDCG@10%) to provide a metric that considers Sensitivity and Precision of the prediction altogether (e.g., Bao et al., 2020; Brown et al., 2020). For this metric, we rank the firm-year observations in the test dataset by descending misconduct probabilities down until the last firm-year observation that is ranked in the top 10% (for a cutoff at the top 10%) of all observations in a test year, i.e., the "ranking list". DCG@k equals the Discounted Cumulative Gain of observations in the ranking list at ranking position k. The calculation of DCG@k thereby considers two main premises. Firstly, true misconduct observations in the ranking list are scored higher than non-misconduct observations. Secondly, true misconduct observations are scored higher when they appear higher in the ranking list. The ideal DCG@k value is the value that is equivalent to all the true misconduct observations being ranked at the top of the ranking list. NDCG@k normalizes the DCG@k by the ideal value and can thus take values between zero and one with higher values representing better performance.

Finally, we report a metric that is insensitive to the choice of the cutoff. The area

[136] Be aware that our European sample prediction results pool the prediction for the years 2005 to 2017, while our U.S. Test benchmark prediction always relates to just one year (i.e., 2005 or 2010), so that the number of firms is equal to the number of firm-years.

under the curve (AUC) summarizes the entire two-dimensional area under the receiver operating characteristics (ROC) curve (Fawcett, 2006). The ROC curve of a classifier is derived by plotting the True Positive Rate (i.e., the Sensitivity) on the y-axis and the False Positive Rate on the x-axis at different classification cutoffs.[137] The AUC ranges in value from zero to one, yet the minimum reference value is 0.5 as the ROC of a random classifier already has an AUC value of 0.5.

4.4 Empirical results

4.4.1 Descriptive statistics

Table 22 Panel B shows descriptive statistics regarding firm size in terms of total assets (*at*) and sales (*sale*) of the firms in our samples. Firms in the main U.S. Test sample of 2005 are somewhat larger than firm-years over the period of the Training sample from 1991 to 2003. This is reasonable as the covered firms grew over the years both nominal and real. As for the nominal growth, it is uncommon to correct for inflation in the misconduct literature. Firms in the European sample from 2005 to 2017 are slightly larger than firms in the U.S. Training but somewhat smaller than in 2005's U.S. Test sample. We consider these descriptives to indicate that our process to build the European sample within the SIC-3 industry categories of the misconduct firms arrives at a cross-section of firms that is generally comparable to the U.S. cross-section utilized for training and benchmark testing.

4.4.2 Logistic Regression

We begin our prediction analyses with the two Logistic Regression approaches with seven and 14 ratios, respectively. Table 23 shows the results of running the regression on the U.S. Training sample to calculate the coefficients for prediction. It is intuitive that RSST accruals (*rsst_acc*) and the change in receivables (*ch_rec*) are significantly positively associated with misconduct, the change in inventory (*ch_inv*) is insignificant. The coefficients for the percentage of soft assets (*soft_assets*), the change in cash sales (*ch_cs*), and the

[137]The False Positive Rate measures the portion of actual non-misconduct observations that are incorrectly classified as misconducted. The AUC metric is sometimes abbreviated with AUROC (i.e., the Area Under the Receiver Operating characteristics Curve).

change in return on assets (*ch_roa*), suggest that misconduct firms have a less tangible asset base, grow faster, and are facing decreasing profitability, respectively. The coefficient on whether the firm issues capital or debt (*issue*) is strongly positively associated with misconduct. When extending the model to include 14 ratios, the change in the cash margin (*ch_cm*) seems to be negatively associated with misconduct while retained earnings over total assets (*reoa*) is positively associated with misconduct. The overall explanatory power of the two LR models measured with McFadden's R^2 is rather low, in line with explaining a rare binary event being difficult.

Table 23: Coefficients of Logistic Regression models

Variable	Basic Dechow et al. (2011) LR model	Bao et al. (2020) LR model
Intercept	-7.623 (0.277) ***	-7.424 (0.296) ***
rsst_acc	0.267 (0.132) **	-0.002 (0.32)
ch_rec	1.404 (0.478) ***	1.514 (0.546) ***
ch_inv	1.053 (0.641)	1.111 (0.742)
soft_assets	2.174 (0.199) ***	2.223 (0.203) ***
ch_cs	0.066 (0.031) **	0.064 (0.033) *
ch_roa	-0.300 (0.148) **	-0.202 (0.235)
issue	1.354 (0.248) ***	1.261 (0.248) ***
wc_acc		-0.294 (0.432)
ch_cm		-0.032 (0.019) *
bm		-0.036 (0.044)
dpi		-0.054 (0.096)
reoa		0.146 (0.047) ***
EBIT		-0.139 (0.16)
ch_fcf		-0.227 (0.265)
$R^2_{McFadden}$	0.040	0.046
Observations	76,819	76,819

This table presents the regression coefficients of Logistic Regressions of the dummy indicating misconduct or non-misconduct on either seven or 14 financial ratios for the U.S. Training sample from 1991 to 2003. Standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table 24 presents the European prediction results for the main U.S. Test sample benchmark in the year 2005 and the alternative benchmark in the year 2010. The respective results for the European sample pool the firm-years from 2005 to 2017. Panel A (Panel B) shows the results for the basic Dechow et al. (2011) LR model (the Bao et al. (2020) LR model). We observe a Sensitivity in the European sample (when percentiles are ranked relative to the U.S. distribution) of 18.64%, i.e., 11 of the 59 misconduct firm-years are among the top 10% in terms of predicted misconduct probability (not tabulated). A total of seven firms of the 21 misconduct firms are flagged correctly in at least one of the misconduct years. We observe quite similar results when looking at the alternative percentile calculations within the European sample only and within the European sample per year (Sensitivity for both: 16.95%), where a total of six and eight firms are flagged correctly at some point, respectively. Naturally, the corresponding Precision is only 1.23%, 1.38%, and 1.37%, respectively, since a cutoff at the top 10% also flags many non-misconduct firm-years (i.e., false positives). The NDCG@k (AUC) metric is 0.099 (0.606) for the percentiles relative to the U.S. and within Europe and 0.092 (0.608) when ranking independently per year.[138] The fact that the performance measures are quite similar across different ranking methods indicates that the distribution of misconduct probabilities in the European sample is similar to the U.S. Test sample, i.e., the estimated probabilities for the European sample are neither generally higher nor generally lower than those for the U.S. sample.

With the exception of the AUC, the metrics calculated for the main U.S. Test sample are comparably low. For instance, the Sensitivity is just 11.90%, which means that only five of the 42 misconduct firms are flagged correctly. When comparing the performance metrics, we infer that the basic Dechow et al. (2011) LR model is a promising approach for out-of-sample applications in misconduct detection. However, the performance figures should not be interpreted as to that the out-of-sample application produces stronger results than when predicting within the U.S. firm universe. Our collected cases are likely more severe than the average case included in the U.S. Training and Test samples since they were identified by less strong enforcement entities, making their detection easier.

When extending the training period until the year 2008 and running the U.S. benchmark test for the year 2010, we get quite similar results for the European out-of-sample prediction. Interestingly, the results for the U.S. Test sample in year 2010 are somewhat better than for the year 2005. Some of this deviation is likely pure chance. However, in line with the SEC becoming less powerful in those years, it identified less misconduct cases

[138] Those two metrics must be the same for the ranking approaches relative to the U.S. and within Europe and only differ when ranking per year. In the latter case, we calculate these two metrics per year (i.e., 2005 to 2017) and report their average figures.

Table 24: Logistic Regression model prediction results

Panel A: Basic Dechow et al. (2011) LR model

	Performance metrics with cutoff at top 10%				
	Sensitivity	Precision	# Firms	NDCG@k	AUC
U.S. Training sample 1991-2003					
U.S. Test sample - 2005	11.90%	0.88%	5/42	0.059	0.623
European sample (2005-2017)					
Relative to U.S. Test sample	18.64%	1.23%	7/21	0.099	0.606
Relative to itself	16.95%	1.38%	6/21	0.099	0.606
Relative to itself per year	16.95%	1.37%	8/21	0.092	0.608
U.S. Training sample 1991-2008					
U.S. Test sample - 2010	16.67%	0.77%	4/24	0.067	0.647
European sample (2005-2017)					
Relative to U.S. Test sample	22.03%	1.07%	9/21	0.089	0.605
Relative to itself	15.25%	1.24%	6/21	0.089	0.605
Relative to itself per year	13.56%	1.09%	6/21	0.082	0.606

Panel B: Bao et al. (2020) LR model

	Performance metrics with cutoff at top 10%				
	Sensitivity	Precision	# Firms	NDCG@k	AUC
U.S. Training sample 1991-2003					
U.S. Test sample - 2005	11.90%	0.88%	5/42	0.061	0.641
European sample (2005-2017)					
Relative to U.S. Test sample	20.34%	1.11%	8/21	0.085	0.618
Relative to itself	15.25%	1.24%	6/21	0.085	0.618
Relative to itself per year	15.25%	1.23%	7/21	0.088	0.621
U.S. Training sample 1991-2008					
U.S. Test sample - 2010	20.83%	0.96%	5/24	0.087	0.690
European sample (2005-2017)					
Relative to U.S. Test sample	22.03%	0.94%	9/21	0.090	0.620
Relative to itself	16.95%	1.38%	6/21	0.090	0.620
Relative to itself per year	13.56%	1.09%	6/21	0.078	0.628

This table presents the classification performance evaluation for two LR models with seven or 14 financial ratios. The training period is 1999 to 2003 for the main approach without looking ahead of the European sample beginning in 2005 and 1999 to 2008 for the alternative approach that utilizes all available years with good misconduct data in the U.S. sample until 2008. The three alternative ways to calculate the percentiles of the European misconduct prediction are relative to the U.S. Test sample percentiles, relative to its own distribution of misconduct predictions pooling all years, and relative to its own distribution of misconduct predictions per year from 2005 to 2017.

in 2010 (24) than in 2005 (42). It may be that the fewer identified cases in a given year are more severe, making the detection with the model easier. Panel B of Table 24 shows the results for the Bao et al. (2020) LR model that increases the basic Dechow et al. (2011) LR model to 14 ratios. While the results appear to be somewhat better for the U.S. Test, in particular when looking at the alternative sample in year 2010, we do not observe an improvement for the European sample. The performance metrics are again quite similar across the different ranking methods. This model may be a meaningful alternative, however, at this point, we cannot confirm that it adds significant detection power to the *far*-out-of-sample approach.

4.4.3 RUSBoost Ensemble Learning

We next turn to the prediction results of the potentially more powerful RUSBoost Ensemble Learning approach. We feed the algorithm with the 28 raw data items that are used to calculate the 14 ratios. Table 25 shows the relative importance of the items in the final strong classifier, which is constructed by the weighted majority vote of the single predictions of the entire ensemble of 3,000 weak classifiers. Two of the three most important items are the number of common shares outstanding (*csho*) and the sale of common and preferred stock (*sstk*), both of which are affected when firms issue equity. Firms may have greater incentives to manipulate and may also have a higher likelihood to be selected for investigation by the SEC when they are involved in financing activities on capital markets (Dechow et al., 2011). Further, *act*, *ppegt*, *che*, *ap*, *re* and *prcc_f* do also seem to be quite important. Some of these variables are not perfectly matched between Compustat and Compustat Global, we therefore remove them in later analyses.

Table 26 reports the performance evaluation of the prediction results. The Sensitivity when calculating the European percentiles relative to the distribution of the main U.S. Test sample in year 2005 shows an astonishing 30.51%. 18 of the 59 firm-years (or 10 of the 21 firms) are flagged correctly at the chosen cutoff of 10%. This again compares quite favorably to the results of the U.S. Test sample, which also shows a higher value (Sensitivity: 26.19%). The figures for Precision, NDCG@k and the AUC are also slightly

Table 25: Feature importance of the data items in the RUSBoost model

Variable	Importance	Variable	Importance	Variable	Importance	Variable	Importance
csho	1.639	*prcc_f*	1.254	*at*	1.076	*sale*	0.668
act	1.484	*ceq*	1.199	*ib*	0.884	*ivao*	0.638
sstk	1.423	*dlc*	1.169	*dltis*	0.872	*lt*	0.624
ppegt	1.411	*invt*	1.125	*dltt*	0.848	*txp*	0.588
che	1.405	*rect*	1.122	*lct*	0.817	*ivst*	0.548
ap	1.399	*dp*	1.121	*xint*	0.755	*ni*	0.408
re	1.270	*cogs*	1.093	*txt*	0.733	*pstk*	0.275

This table presents the feature importance in descending order for the RUSBoost Ensemble Learning for the U.S. Training sample from 1991 to 2003.

higher than for the U.S. benchmark prediction and substantially higher than for the *far-out-of-sample* predictions of the LR models. The performance figures are comparably high when ranking the probabilities within the European sample or within each year of the European sample. The results for the alternative U.S. Test sample in year 2010 are similar yet even slightly stronger.

Table 26: RUSBoost prediction results

	Performance metrics with cutoff at top 10%				
	Sensitivity	Precision	# Firms	NDCG@k	AUC
U.S. Training sample 1991-2003					
U.S. Test sample - 2005	26.19%	2.04%	11/42	0.125	0.726
European sample (2005-2017)					
Relative to U.S. Test sample	30.51%	2.49%	10/21	0.169	0.726
Relative to itself	30.51%	2.48%	10/21	0.169	0.726
Relative to itself per year	32.20%	2.60%	11/21	0.183	0.728
U.S. Training sample 1991-2008					
U.S. Test sample - 2010	29.17%	1.41%	7/24	0.119	0.640
European sample (2005-2017)					
Relative to U.S. Test sample	37.29%	2.32%	12/21	0.190	0.713
Relative to itself	33.90%	2.75%	11/21	0.190	0.713
Relative to itself per year	35.59%	2.87%	12/21	0.209	0.717

This table presents the classification performance evaluation for the RUSBoost Ensemble Learning model with the main and the alternative training samples and the three alternative ways to rank the European percentiles of the distribution of the misconduct prediction.

4.5 Robustness tests

4.5.1 Conformity between Compustat and Compustat Global

Our robustness tests focus on data limitations stemming from the use of Compustat Global, subjecting the approach to a more conservative cutoff, and sensitivity to the choice of the training period. We first address potential biases introduced by remaining differences in the databases used for training (Compustat) and our out-of-sample application (Compustat Global). Out of our final European sample containing 659 firms (7,653 firm-years), 19 firms (191 firm-years) are also covered by Compustat. We can therefore compare the figures for the raw data items as provided in Compustat with our currency-translated matches from Compustat Global. We calculate the difference per raw item and scale it by total assets if it is a balance sheet, income statement or cash flow statement item or by itself if it is a market value item. We assess whether the absolute difference is off by at least 10%.

The two raw items with the most severe inconsistencies are the two market value items number of shares (*csho*) and their end of year price (*prcc_f*). While Compustat provides them in the annual file, using Compustat Global requires drawing them from the daily file through aggregating over potentially several share classes, which is common in Europe. *csho* and *prcc_f* are off by at least 10% in about one third of the firm-years.[139] Therefore, in our first robustness test, we remove both items from the list of items fed into the RUSBoost approach. Next, we remove the debt issuance item *dltis*, which is not populated at all in Compustat Global after fiscal year 2004. We further remove the items that are at least 10% off in at least 10% of firm-years. This concerns the retained earnings (*re*) and the cost of goods sold (*cogs*). We rerun the RUSBoost model without these five items. These items affect four of the ratios used in the Bao et al. (2020) LR model (i.e., *issue*, *ch_cm*, *bm*, and *reoa*), accordingly, we also rerun this model without them.

Table 27 reports the results. Panel A reports results for the Bao et al. (2020) LR model

[139] Their implied market value of equity (i.e., *csho* times *prcc_f*) is off by at least 10% in only about 9% of firm-years, though, meaning that some of the inconsistencies must stem from our approach of drawing the figures from the daily file through aggregation, which may lead to reporting higher numbers of shares at lower prices or vice versa.

without the four affected ratios. As one would expect, the performance results for the U.S. Test sample are slightly lower (e.g., Sensitivity of 16.67% for the 2010 U.S. Test sample compared to 20.83% for the full set of variables), in line with less explanatory power when removing the affected ratios. The out-of-sample prediction evaluation metrics using both the main and the alternative U.S. Training and Test samples do show some changes, the overall performance does not seem to suffer though. This makes us confident that our strong *far*-out-of-sample results are not driven by potential biases stemming from data inconsistencies.

The results of the reduced RUSBoost model show a more pronounced pattern. First, the degradation for the 2010 U.S. Test sample appears to be larger.[140] At the same time, there seems to be a substantial improvement in the performance of the European out-of-sample predictions. This implies that data inconsistencies do affect the raw item-based RUSBoost, but also that it still performs well when removing affected items for *far*-out-of-sample applications. We generally consider these results to imply strong practicability for enforcement agencies outside the U.S. that might not tolerate potential data inconsistencies or might choose to remove specific items or ratios for other reasons. It appears that both the LR approach and the RUSBoost approach are quite robust to doing so.

4.5.2 Further robustness tests

More conservative cutoff at the top 5%

The results presented so far assume that a cutoff at the top 10% of firms in terms of predicted misconduct probability implies a feasible workload for the respective enforcement entity. We report results for a more conservative cutoff at the top 5% (e.g., Brown et al., 2020) in Table 28. We focus on the two reduced models from the previous section. We further focus on the main U.S. Test benchmark of year 2005 with a corresponding U.S. Training sample until 2003. Panel A (Panel B) shows the reduced LR model (the reduced RUSBoost model).

It appears in Panel A that the more conservative cutoff has quite an effect on the

[140] The results for the 2005 U.S. Test sample slightly improve which we rather assign to chance.

Table 27: Robustness tests regarding data conformity

Panel A: Bao et al. (2020) LR model without *issue*, *ch_cm*, *bm*, and *reoa*

	Performance metrics with cutoff at top 10%				
	Sensitivity	Precision	# Firms	NDCG@k	AUC
U.S. Training sample 1991-2003					
U.S. Test sample - 2005	11.90%	0.88%	5/42	0.058	0.632
European sample (2005-2017)					
Relative to U.S. Test sample	22.03%	0.92%	9/21	0.080	0.573
Relative to itself	11.86%	0.96%	6/21	0.080	0.573
Relative to itself per year	13.56%	1.09%	7/21	0.085	0.593
U.S. Training sample 1991-2008					
U.S. Test sample - 2010	16.67%	0.77%	4/24	0.067	0.676
European sample (2005-2017)					
Relative to U.S. Test sample	30.51%	0.91%	10/21	0.083	0.576
Relative to itself	15.25%	1.24%	7/21	0.083	0.576
Relative to itself per year	15.25%	1.23%	6/21	0.083	0.599

Panel B: RUSBoost model without *re*, *cogs*, *dltis*, *prcc_f*, and *csho*

	Performance metrics with cutoff at top 10%				
	Sensitivity	Precision	# Firms	NDCG@k	AUC
U.S. Training sample 1991-2003					
U.S. Test sample - 2005	28.57%	2.22%	12/42	0.135	0.697
European sample (2005-2017)					
Relative to U.S. Test sample	40.68%	2.58%	12/21	0.168	0.759
Relative to itself	30.51%	2.48%	10/21	0.168	0.759
Relative to itself per year	33.90%	2.73%	11/21	0.182	0.760
U.S. Training sample 1991-2008					
U.S. Test sample - 2010	20.83%	1.01%	5/24	0.089	0.682
European sample (2005-2017)					
Relative to U.S. Test sample	50.85%	3.06%	13/21	0.224	0.734
Relative to itself	42.37%	3.44%	12/21	0.224	0.734
Relative to itself per year	42.37%	3.42%	12/21	0.236	0.738

This table presents the classification performance evaluation for the two types of models after reducing them in line with the data conformity between Compustat and Compustat Global. The alternative training samples and percentile calculations follow the same methodology as before.

Table 28: Further robustness tests

	Performance metrics with specific cutoff				
	Sensitivity	Precision	# Firms	NDCG@k	AUC
Panel A: Cutoff at top 5% for reduced Bao et al. (2020) LR model					
U.S. Training sample 1991-2003					
U.S. Test sample - 2005	4.76%	0.70%	2/42	0.028	0.632
European sample (2005-2017)					
Relative to U.S. Test sample	8.47%	0.91%	4/21	0.058	0.573
Relative to itself	5.08%	0.87%	3/21	0.058	0.573
Relative to itself per year	8.47%	1.43%	4/21	0.071	0.593
Panel B: Cutoff at top 5% for reduced RUSBoost model					
U.S. Training sample 1991-2003					
U.S. Test sample - 2005	11.90%	1.96%	5/42	0.064	0.697
European sample (2005-2017)					
Relative to U.S. Test sample	18.64%	2.59%	8/21	0.076	0.759
Relative to itself	11.86%	2.03%	5/21	0.076	0.759
Relative to itself per year	15.25%	2.57%	7/21	0.083	0.760
Panel C: Recent training data for reduced Bao et al. (2020) LR model					
U.S. Training sample 1994-2003					
U.S. Test sample - 2005	11.90%	0.88%	5/42	0.058	0.635
European sample (2005-2017)					
Relative to U.S. Test sample	27.12%	1.03%	10/21	0.079	0.577
Relative to itself	11.86%	0.96%	6/21	0.079	0.577
Relative to itself per year	13.56%	1.09%	7/21	0.086	0.595
Panel D: Recent training data for reduced RUSBoost model					
U.S. Training sample 1994-2003					
U.S. Test sample - 2005	19.05%	1.48%	8/42	0.101	0.685
European sample (2005-2017)					
Relative to U.S. Test sample	33.90%	2.29%	11/21	0.166	0.755
Relative to itself	28.81%	2.34%	10/21	0.166	0.755
Relative to itself per year	32.20%	2.60%	10/21	0.160	0.754
Panel E: Without dotcom for reduced Bao et al. (2020) LR model					
U.S. Training sample 1991-2003 without dotcom years (1999-2002)					
U.S. Test sample - 2005	9.52%	0.70%	4/42	0.045	0.620
European sample (2005-2017)					
Relative to U.S. Test sample	20.33%	0.98%	9/21	0.074	0.568
Relative to itself	11.86%	0.96%	6/21	0.074	0.568
Relative to itself per year	11.86%	0.96%	6/21	0.086	0.579
Panel F: Without dotcom for reduced RUSBoost model					
U.S. Training sample 1991-2003 without dotcom years (1999-2002)					
U.S. Test sample - 2005	16.67%	1.30%	7/42	0.089	0.676
European sample (2005-2017)					
Relative to U.S. Test sample	42.37%	2.83%	10/21	0.179	0.728
Relative to itself	35.59%	2.89%	9/21	0.179	0.728
Relative to itself per year	38.98%	3.14%	10/21	0.179	0.727

This table presents the classification performance evaluation for several further robustness tests performed on the two types of models in their reduced form presented in Table 27. Panel A and Panel B assume a higher cutoff at the top 5%. Panel C and Panel D focus on the most recent ten-year-period of the U.S. Training sample. Panel E and Panel F remove the dotcom years from 1999 to 2002 from the U.S. Training sample.

reduced Bao et al. (2020) LR model. The Sensitivity for the main U.S. Test sample goes down from 11.90% to 4.76%, the Sensitivity for the European sample shrinks from 22.03% to 8.47% (relative to the U.S. Test sample). While the Precision seems to be largely unaffected, the NDCG@k figures also decrease substantially. The AUC figures are independent of the choice of the cutoff.

As reported in Panel B, the prediction performance for the RUSBoost model also decreases substantially at a cutoff of 5%, for instance, the Sensitivity for the European prediction relative to the U.S. Test decreases from 40.68% to 18.64%. The Precision for the European predictions also decreases slightly. We regard these specifications to be the most conservative ones presented throughout the paper and document that even with this cutoff, five to eight out of 21 misconduct firms are flagged correctly at some point.

Focus on more recent data to train the models

Next, we address the concern that an agency that trains a model with the historic U.S. data might be skeptical about relying on rather old training data. We believe that the preferred modus operandi rather implies using the most recent training years on a rolling window basis when doing the yearly preselection. We therefore remove as many old years of data so that we end up with the minimum requirement of ten years suggested in Bao et al. (2020). Consequently, we shrink the main U.S. Training sample to financial years 1994 to 2003. Panel C and Panel D of Table 28 show the results for the two reduced LR and RUSBoost models from before, the cutoff is 10%.

It appears that results for both models are largely unaffected by reducing the training period to the more recent years. The reported Sensitivity, Precision, number of correctly flagged firms, NDCG@k, and AUC figures are very similar to the ones reported in Table 27. It again appears as if the RUSBoost approach has more predictive power than the well-established LR approach.

Remove the dotcom misconduct cases from the years 1999 to 2002 from the training sample

Finally, we note that the years from the building to the burst of the dotcom bubble (i.e., 1999 until 2002) have substantially more cases in the U.S. Training sample than earlier or later years. The misconduct schemes implemented in the affected firms in these years were somewhat special in the sense that whole business models in hyped new economy contexts were made up to obtain funding at capital markets. However, also the enforcement handling with hindsight of the bubble's burst was somewhat special (e.g., Penman, 2003).[141] One may therefore prefer to remove those cases from the training sample. We do so in Panel E and Panel F of Table 28 for the reduced models from before.

While we do observe that this sample restriction somewhat affects the prediction performance for the U.S. Test sample, the results for the European out-of-sample application seem again largely unaffected. This reinforces our confidence that our strong results for the *far*-out-of-sample application are not model artefacts. Still, the application of the established models appears to be very robust to design choices one may make to address institutional differences.

4.6 Conclusion

This paper is a first attempt to assess whether and how U.S.-trained accounting misconduct prediction models can be applied *far*-out-of-sample in markets that do not have nearly sufficient data-points to do their own training. Feeding only readily available financial statement data into well-established models performs notably well. In most of our tests, the prediction performance metrics are even higher than for the U.S. benchmarks, which likely means that our identified cases are very severe and easy to detect. It could also mean that non-U.S. firms mimic earlier U.S. firms' misconduct schemes and implement them in similar ways, making it easier for models trained on those U.S. firms to flag correctly. Such behavior would constitute even stronger arguments for the application of our suggested approach.

It is unclear, how megatrends like technological advancement and increases in societal wealth affect the incidence of financial misconduct in the years to come. Some researchers expect a long-term decrease, on net (e.g. Karpoff, 2021). We view our approach to be a core building block to lever various technological advancement. For instance, future

[141] Misconduct detection may be different and generally more likely during a macroeconomic bust (e.g., Amiram et al., 2018).

research may develop ways to apply also non-financial numerical data (e.g., Brazel et al., 2009) or even textual misconduct models (e.g., Brown et al., 2020; Larcker and Zakolyukina, 2012; Purda and Skillicorn, 2015) *far*-out-of-sample.

5 Conclusion

This chapter concludes my dissertation. As each of the included essays already contains a section with concluding remarks, I try to bring in a higher-level and at times more critical perspective in the following paragraphs.

On the broadest level, my dissertation improves our understanding of several aspects of corporate governance. Firstly, widening our understanding of corporate investment, my essays show that more long-term executive compensation improves investment efficiency and that modern firms' investments in human capital may be measurable but are not fully captured by market participants. Secondly, regarding the economics of regulation, my dissertation is informative for mandating executive compensation, improving the disclosure around intangible capital creation, and developing accounting misconduct prediction processes in non-U.S. enforcement institutions. Thirdly, my essays relate to other stakeholders' interests in the sense that employees and their interest groups get arguments for discussions to advance how firms think and report about investing in human capital and that the general public (i.e., all kinds of different interest groups) learns how enforcement of accounting misconduct can leverage existing data to become more effective, likely deterring some misconduct ex-ante.

In **Essay 1**, I theoretically and empirically connect mandated longer duration in executive compensation and firms' investment efficiency. Mandated longer duration increases the length of executives' evaluation window, which could force them to internalize the unprofitability of today's empire building that materializes in the future, reducing over-investment. At the same time, the longer evaluation window may mitigate myopic disincentives for current-period investments with future payoffs or incentives for over-signalling, reducing under- or over-investment. However, such a regulation may also interfere with efficient contracting or it may be immaterial if it is too vague.

Subjecting this to a series of tests, I find statistically and economically significant improvements in investment efficiency following the regulation. The results are robust to a wide range of (abnormal) investment specifications and different plausible control groups. Cross-sectional analyses suggest that the documented effects are driven by firms with weak corporate governance characteristics, in particular, firms whose compensation committee had a low degree of independence before the regulation. It further appears that these improvements are mainly driven by reductions in over-investment. Overall, my findings support the notion that long-term orientation in executive compensation serves a role in mitigating frictions that hamper efficient investment decisions. Governmental and regulatory intervention into corporate governance in general and executive compensation in particular is frequently discussed and demanded. The essay underscores how regulation potentially affects real firm behavior and emphasizes the need to report and investigate the economic impacts of such intervention.

This essay heavily builds on and benefits from its strong identification of the independent construct of longer incentive duration. This is in line with a general trend towards (hoping for) causal inference despite using observational data in accounting research (e.g., Gow et al., 2016). The importance of quasi-experimental exogenous identification is nicely illustrated by the fact that the reliance on the regulatory intervention allows this essay to investigate executive compensation without even controlling for actual annual compensation. It further allows to remain somewhat agnostic as to how specific firms' systems were before the act as well as how specific firms changed their systems. Thereby the unique properties of the identified setting extend accounting research despite the problems around the transparency of executive compensation in European firms mentioned in the introduction of the dissertation. However, not every construct of interest can be identified through a quasi-experimental shock. Therefore, in particular for new phenomena, in-depth descriptive studies (like the second essay of the dissertation) are more suitable.

Despite the strong identification of the treatment group (i.e., publicly listed German firms) in the VorstAG setting, identifying an ideal control group is less straightforward. The essay therefore motivates several different control groups, i.e., all other European firms, a matched subset of these firms, and not (directly) affected private German firms. Despite allowing fewer different model operationalizations, this third control group is likely quite strong as it limits concerns that German firms increased their investment efficiency compared to their European peers for reasons other than or in addition to the compensation treatment.[142]

Finally, while the first essay clearly establishes that mandated longer duration reduces abnormal investment in general, it may be that the effect is fully driven by reductions in over-investment. The results suggest that under-investment possibly even increased. As accounting research does not yet contrast the two forms of inefficient behavior, I cannot assess whether (some) under-investment can be tolerated when inhibiting (a lot of) over-investment. Further, we do not know whether the investment efficiency construct allows firms enough leeway to invest in radical innovation or postpone investment to leapfrog development stages. Related to this and highlighted by the second essay of the dissertation, it may be that some very important forms of investment that the modern firm makes is not adequately captured by the established investment measures. In other words, I caution the reader to be aware that the interpretation of the first essay's results is limited by the established methodologies.

After the first essay relies on established proxies of corporate investment, **Essay 2** questions them and tries to make them more suitable for today's organizational characteristics. We develop a strategy to examine aspects of the intangible human capital investment embedded in a firm's PE. As disclosures around human capital are limited and opaque, we explore whether this opacity hinders price discovery. In our main analyses, we show that risk-adjusted abnormal returns can be earned on portfolios formed on our human capital creation proxy. These findings are robust to model selection and measurement choice.

Our findings naturally motivate several further research questions: First, does the legal environment affect how returns to corporate human capital creation are realized (e.g., Shleifer and Vishny, 1997)? For instance, does the mobility of employees (including the firm's intellectual property) (Klasa et al., 2018) affect how and when stock prices reflect investments in human capital? And relatedly, can firms acquire the human capital

[142] In line with non-U.S. empirical accounting research, I use the Thomson Reuters (now Refinitiv) Datastream data for the main sample of listed European firms. I add data from Bureau van Dijk's Dafne database to analyze the third control group of large private German firms. Only few papers utilize this database for empirical research (Engel and Middendorf (2009) being one example), although it should provide ample opportunities to research concepts related to the economies of pubic listing, reporting requirements, ownership structures, and resulting corporate governance forces.

creating ability of target firms and does the degree of relatedness of merging firms' human capital creation matter for acquisition and merger performance (Lee et al., 2018)?

Second, there are opportunities for research utilizing our proxy to better understand processes around human capital creation internal to the firm. Important questions revolve around implicit guarantees to employees and the stickiness of firms' PE commitments, for instance, does PE behave more asymmetrically in expansion versus recession scenarios when the firm has a higher potential to create future values from it (Chen et al., 2012)? Relatedly, do compensation committees explicitly account for human capital creation in designing the contracts? For instance, do firms with high human capital creating ability grant more long-term executive compensation to incentivize such investments (Banker et al., 2011), and is executive compensation shielded from the negative effects of expensing personnel expenditures when they create higher future values (Huson et al., 2012)?

The human capital creation proxy that we develop is clearly crude. However, it is a first attempt at measuring what is likely the most important asset that today's firms (try to) create. Providing such an imperfect attempt should be helpful for later and better proxies that future research might develop. Moreover, it may be informative for (all kinds of potential) regulators in showing that financial reporting may need more transparency regarding how firms invest in their personnel. One very straightforward issue that this essay makes transparent is that U.S. firms do not even have to report PE while our results demonstrate that it apparently has an important information content.

One major limitation of our suggested methodology is that it requires quite a long timeline of data per firm to be estimated. One may say that our methodology rather captures human capital creation in firms that survive long enough to be included in the sample. While those firms probably are (or become) the most important players in the economy, the analysis is likely subject to a substantial survivorship bias. In this regard, simpler proxies (like the last one presented in the essay), may be more practical for application outside of academic research.[143]

[143] While we report the results for this last proxy on the same final sample as the main analyses, many of the sample selection steps would not be necessary, making it possibly more useful for other applications.

Essay 3 tackles a rather different aspect of corporate governance and mainly speaks to regulators. It is a first attempt to assess whether and how U.S.-trained accounting misconduct prediction models can be applied *far*-out-of-sample in markets that do not have nearly sufficient data-points to do their own training. Feeding only readily available financial statement data into well-established models performs notably - if not astonishingly - well. In most of our tests, the prediction performance metrics are even higher than for the U.S. benchmarks, which likely means that our identified cases are very severe and easy to detect. It could even be that non-U.S. firms mimic earlier U.S. firms' misconduct schemes and implement them in similar ways, making it easier for models trained on those U.S. firms to flag correctly. We argue that such behavior would constitute even stronger arguments for the application of our suggested approach.

As mentioned earlier, this essay may be characterized as *untheoretical* in the sense that it provides little theoretical contribution. While the absence of theoretical contribution does not yet indicate practical contribution, the design of the approach presented in the essay clearly aims for offering a practicable workflow. In light of the strong results regarding the suggested *far*-out-of-sample application, it appears irresponsible to not consider implementing such an approach as a non-U.S. enforcement institutions' decision maker. We further argue that our suggested approach may at some point allow to apply also non-financial numerical data (e.g., Brazel et al., 2009) or even textual misconduct models (e.g., Brown et al., 2020; Larcker and Zakolyukina, 2012; Purda and Skillicorn, 2015) out-of-sample. Finally, the *far*-out-of-sample approach may improve predictions of other rare events like corporate bankruptcy or bank failure (e.g., Charitou et al., 2004; Jin et al., 2011; Tam and Kiang, 1992).

A major limitation can be seen in the timing of the suggested analyses. In 2021 (i.e., the time of the writing of the essay), the data availability allowed to analyze data until fiscal year 2017, which corresponds to a four year gap. This may be in conflict with enforcers trying to dissect and probe the latest corporate filings at any point in time. While digitalization trends in corporate reporting like XBRL (i.e., Extensible Business Reporting Language) as well as database providers becoming quicker at updating annual accounts will reduce that gap substantially, such developments may take some time. Until then, this time lag constitutes an impediment to effective enforcement as for many misconduct schemes the probability of uncovering them likely decreases the more time passes since its installation.[144]

Relatedly, our envisioned enforcement institution might not prefer to follow our suggested combination of Compustat and Compustat Global databases.[145] In a real approach, it might be even more practical to use Refinitiv data (earlier Thomson Reuters Datastream) that is consistent across U.S. and non-U.S. firms. There are robust coding schemes to align the Refinitiv items with the typically used Compustat items, so that one likely would not even have to reduce the models in line with the data inconsistencies as suggested in our essay. While providing results for such an approach would go beyond the scope of our essay without adding insight to its core research question, it is fair to assume but yet to show that there is a way to use Refinitiv that produces results that are similarly useful as those presented in the essay.

[144]Not all misconduct schemes become so large that they must inevitably result in being detected. Even in the case of Wirecard, that ex-post appears as if it would have been impossible to hide for a sustained period of time (see Appendix C.4), the fraudsters had reasonable plans such as the merger with another entity that would possibly have allowed to bury the misconduct (Storbeck, 2020).

[145]Note that U.S. studies almost always use Compustat data (i.e., focus on North America), however, (multi-country) non-U.S. studies usually use Refinitiv data (earlier Thomson Reuters Datastream). We use Compustat for the U.S. data to be able to perform the model training exactly as in earlier papers. This way, the focal issue in our approach really is to test the performance of these trained models *far*-out-of-sample. Consequently, we have to use the (upcoming but still less common) Compustat Global data for the non-U.S. testing to have the best possible comparability.

Bibliography

Agarwal, V. and Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking and Finance*, 32(8):1541–1551.

Amat, N., Mercader, R., Crespo, P., and Feijoo, A. N. (2013). La fiabilidad de la información financiera y su impacto para analistas e inversores. el caso de pescanova. *Revista de Contabilidad y Dirección*, 16:291–314.

Amiram, D., Bozanic, Z., Cox, J. D., Dupont, Q., Karpoff, J. M., and Sloan, R. (2018). Financial reporting fraud and other forms of misconduct: a multidisciplinary review of the literature. *Review of Accounting Studies*, 23(2):732–783.

Amiram, D., Bozanic, Z., and Rouen, E. (2015). Financial statement errors: Evidence from the distributional properties of financial statement numbers. *Review of Accounting Studies*, 20(4):1540–1593.

Armstrong, C. S., Barth, M. E., Jagolinzer, A. D., and Riedl, E. J. (2010). Market reaction to the adoption of ifrs in europe. *The Accounting Review*, 85(1):31–61.

Atkins, P. S. and Bondi, B. J. (2008). Evaluating the mission: A critical review of the history and evolution of the sec enforcement program. *Fordham Journal of Corporate & Financial Law*, 13:367–417.

Baker, C. R., Ding, Y., Stolowy, H., et al. (2003). *Income Statements "by Nature" and Analysis of Company Performance: An Application to US Airline Companies*. Groupe HEC.

Balakrishnan, K., Watts, R., and Zuo, L. (2016). The effect of accounting conservatism on corporate investment during the global financial crisis. *Journal of Business Finance & Accounting*, 43(5-6):513–542.

Ballester, M., Livnat, J., and Sinha, N. (2002). Labor costs and investments in human capital. *Journal of Accounting, Auditing & Finance*, 17(4):351–373.

Banker, R. D., Huang, R., and Natarajan, R. (2011). Equity incentives and long-term value created by sg&a expenditure. *Contemporary Accounting Research*, 28(3):794–830.

Banker, R. D., Huang, R., Natarajan, R., and Zhao, S. (2019). Market valuation of intangible asset: Evidence on sg&a expenditure. *The Accounting Review*, 94(6):61–90.

Bao, Y., Ke, B., Li, B., Yu, Y. J., and Zhang, J. (2020). Detecting accounting fraud in publicly traded us firms using a machine learning approach. *Journal of Accounting Research*, 58(1):199–235.

Bao, Y., Ke, B., Li, B., Yu, Y. J., and Zhang, J. (2022). Erratum. *Journal of Accounting Research*, 60(4):1635–1646.

Barth, M. E., Landsman, W. R., Lang, M., and Williams, C. (2012). Are ifrs-based and us gaap-based accounting amounts comparable? *Journal of Accounting and Economics*, 54(1):68–93.

Barton, J. and Simko, P. J. (2002). The balance sheet as an earnings management constraint. *The Accounting Review*, 77(s-1):1–27.

Beasley, M. S., Carcello, J. V., and Hermanson, D. R. (1999). Fraudulent financial reporting: 1987-1997: An analysis of us public companies *Sponsored by the Committee of Sponsoring Organizations of the Treadway Commission (COSO).*

Beasley, M. S., Hermanson, D. R., Carcello, J. V., and Neal, T. L. (2010). Fraudulent financial reporting: 1998-2007: An analysis of us public companies *Sponsored by the Committee of Sponsoring Organizations of the Treadway Commission (COSO).*

Beaver, W. H., Correia, M., and McNichols, M. F. (2010). Financial statement analysis and the prediction of financial distress. *Foundations and Trends® in Accounting*, 5(2):99–173.

Bebchuk, L. A. and Fried, J. M. (2003). Executive compensation as an agency problem. *Journal of Economic Perspectives*, 17(3):71–92.

Bebchuk, L. A. and Fried, J. M. (2010). Paying for long-term performance. *University of Pennsylvania Law Review*, pages 1915–1959.

Bebchuk, L. A. and Stole, L. A. (1993). Do short-term objectives lead to under-or overinvestment in long-term projects? *The Journal of Finance*, 48(2):719–729.

Bell, T. B. and Carcello, J. V. (2000). A decision aid for assessing the likelihood of fraudulent financial reporting. *Auditing: A Journal of Practice & Theory*, 19(1):169–184.

Berle, A. A. J. and Means, G. C. (1932). *The modern corporation and private property.* New York: Macmillan Company.

Bertrand, M. and Mullainathan, S. (2003). Enjoying the quiet life? corporate governance and managerial preferences. *Journal of Political Economy*, 111(5):1043–1075.

Biddle, G. C., Callahan, C. M., Hong, H. A., and Knowles, R. L. (2016). Do adoptions of international financial reporting standards enhance capital investment efficiency? *SSRN Working Paper.*

Biddle, G. C. and Hilary, G. (2006). Accounting quality and firm-level capital investment. *The Accounting Review*, 81(5):963–982.

Biddle, G. C., Hilary, G., and Verdi, R. S. (2009). How does financial reporting quality relate to investment efficiency? *Journal of Accounting and Economics*, 48(2-3):112–131.

Blackburne, T., Bozanic, Z., Johnson, B. A., and Roulstone, D. T. (2020). The regulatory observer effect: Evidence from sec investigations. *SSRN Working Paper.*

Blair, M. M. (2003). Firm-specific human capital and theories of the firm. *SSRN Working Paper.*

Brazel, J. F., Jones, K. L., and Zimbelman, M. F. (2009). Using nonfinancial measures to assess fraud risk. *Journal of Accounting Research*, 47(5):1135–1166.

Brown, N. C., Crowley, R. M., and Elliott, W. B. (2020). What are you saying? using topic to detect financial misreporting. *Journal of Accounting Research*, 58(1):237–291.

Brown, P., Preiato, J., and Tarca, A. (2014). Measuring country differences in enforcement of accounting standards: An audit and enforcement proxy. *Journal of Business Finance & Accounting*, 41(1-2):1–52.

Byard, D., Li, Y., and Yu, Y. (2011). The effect of mandatory ifrs adoption on financial analysts' information environment. *Journal of Accounting Research*, 49(1):69–96.

Cassell, C. A., Huang, S. X., Sanchez, J. M., and Stuart, M. D. (2012). Seeking safety: The relation between ceo inside debt holdings and the riskiness of firm investment and financial policies. *Journal of Financial Economics*, 103(3):588–610.

Cecchini, M., Aytug, H., Koehler, G. J., and Pathak, P. (2010). Detecting management fraud in public companies. *Management Science*, 56(7):1146–1160.

Ceresney, A. (2013). Financial Reporting and Accounting Fraud. *Speech at the American Law Institute Continuing Legal Education, Washington, D.C., Sept. 19, 2013.*

Cernat, L. (2009). The emerging european corporate governance model: Anglo-saxon, continental, or still the century of diversity? In *European Corporate Governance*, pages 160–172. Routledge.

Chan, L. K., Lakonishok, J., and Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *The Journal of Finance*, 56(6):2431–2456.

Chan, S. H., Martin, J. D., and Kensinger, J. W. (1990). Corporate research and development expenditures and share value. *Journal of Financial Economics*, 26(2):255–276.

Charitou, A., Neophytou, E., and Charalambous, C. (2004). Predicting corporate failure: empirical evidence for the uk. *European Accounting Review*, 13(3):465–497.

Chen, C., Young, D., and Zhuang, Z. (2013). Externalities of mandatory ifrs adoption: Evidence from cross-border spillover effects of financial information on investment efficiency. *The Accounting Review*, 88(3):881–914.

Chen, C. X., Lu, H., and Sougiannis, T. (2012). The agency problem, corporate governance, and the asymmetrical behavior of selling, general, and administrative costs. *Contemporary Accounting Research*, 29(1):252–282.

Chen, F., Hope, O.-K., Li, Q., and Wang, X. (2011). Financial reporting quality and investment efficiency of private firms in emerging markets. *The Accounting Review*, 86(4):1255–1288.

Chen, W., Hribar, P., and Melessa, S. (2018). Incorrect inferences when using residuals as dependent variables. *Journal of Accounting Research*, 56(3):751–796.

Cheng, M., Dhaliwal, D., and Zhang, Y. (2013). Does investment efficiency improve after the disclosure of material weaknesses in internal control over financial reporting? *Journal of Accounting and Economics*, 56(1):1–18.

Christensen, H. B., Floyd, E., Liu, L. Y., and Maffett, M. (2017). The real effects of mandated information on social responsibility in financial reports: Evidence from mine-safety records. *Journal of Accounting and Economics*, 64(2-3):284–304.

Christensen, H. B., Hail, L., and Leuz, C. (2013). Mandatory ifrs reporting and changes in enforcement. *Journal of Accounting and Economics*, 56(2-3):147–177.

Ciftci, M. and Cready, W. M. (2011). Scale effects of r&d as reflected in earnings and returns. *Journal of Accounting and Economics*, 52(1):62–80.

Cohen, L., Diether, K., and Malloy, C. (2013). Misvaluing innovation. *The Review of Financial Studies*, 26(3):635–666.

Cressey, D. R. (1953). *Other people's money; a study of the social psychology of embezzlement.* Free press.

Dechow, P. M., Ge, W., Larson, C. R., and Sloan, R. G. (2011). Predicting material accounting misstatements. *Contemporary Accounting Research*, 28(1):17–82.

Deutscher-Bundestag (2009). *Bundestagsdrucksache [Proceedings of the German Parliament] 16 /13433 (17 June 2009)*. Deutscher Bundestag.

Donangelo, A., Gourio, F., Kehrig, M., and Palacios, M. (2019). The cross-section of labor leverage and equity returns. *Journal of Financial Economics*, 132(2):497 – 518.

Dyck, A., Morse, A., and Zingales, L. (2010). Who blows the whistle on corporate fraud? *The Journal of Finance*, 65(6):2213–2253.

Eberhart, A. C., Maxwell, W. F., and Siddique, A. R. (2004). An examination of long-term abnormal stock returns and operating performance following r&d increases. *The Journal of Finance*, 59(2):623–650.

Edmans, A. (2011). Does the stock market fully value intangibles? employee satisfaction and equity prices. *Journal of Financial Economics*, 101(3):621–640.

Edmans, A., Fang, V. W., and Lewellen, K. A. (2017a). Equity vesting and investment. *Review of Financial Studies*, 30(7):2229–2271.

Edmans, A., Gabaix, X., and Jenter, D. (2017b). Executive compensation: A survey of theory and evidence. In *The Handbook of the Economics of Corporate Governance*, volume 1, pages 383–539. Elsevier.

Edmans, A., Gosling, T., and Jenter, D. (2021). Ceo compensation: Evidence from the field. *SSRN Working Paper*.

Edmans, A. and Liu, Q. (2011). Inside debt. *Review of Finance*, 15(1):75–102.

Eisdorfer, A., Giaccotto, C., and White, R. (2013). Capital structure, executive compensation, and investment efficiency. *Journal of Banking and Finance*, 37(2):549–562.

Eisfeldt, A. L. and Papanikolaou, D. (2013). Organization capital and the cross-section of expected returns. *The Journal of Finance*, 68(4):1365–1406.

Enache, L. and Srivastava, A. (2017). Should intangible investments be reported separately or commingled with operating expenses? new evidence. *Management Science*, 64(7):3446–3468.

Engel, D. and Middendorf, T. (2009). Investment, internal funds and public banking in germany. *Journal of Banking and Finance*, 33(11):2132–2139.

Erickson, M., Hanlon, M., and Maydew, E. L. (2006). Is there a link between executive equity incentives and accounting fraud? *Journal of Accounting Research*, 44(1):113–143.

Ernstberger, J., Link, B., Stich, M., and Vogler, O. (2017). The real effects of mandatory quarterly reporting. *The Accounting Review*, 92(5):33–60.

Ernstberger, J., Stich, M., and Vogler, O. (2012). Economic consequences of accounting enforcement reforms: The case of germany. *European Accounting Review*, 21(2):217–251.

ESMA (2020). Fast track peer review on the application of the guidelines on the enforcement of financial information (ESMA/2014/1293) by BaFin and FREP in the context of Wirecard. *European Securities and Markets Authority Peer Review Report*.

European Commission (2019). Intra-eu labour mobility at a glance: Main findings of the 2019 annual report on intra-eu labour mobility *European Commission Directorate-General for Employment, Social Affairs and Inclusion*.

Evans, J. H. I., Gao, Z., Hwang, Y., and Wu, W.-T. (2017). Performance periods in ceo performance-based equity awards: theory and evidence. *The Accounting Review*, 93(2):161–190.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fama, E. F. and French, K. R. (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics*, 123(3):441–463.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874.

Flamholtz, E. (1971). A model for human resource valuation: A stochastic process with service rewards. *The Accounting Review*, 46(2):253–267.

Flammer, C. and Bansal, P. (2017). Does a long-term orientation create value? evidence from a regression discontinuity. *Strategic Management Journal*, 38(9):1827–1847.

Foye, J., Mramor, D., and Pahor, M. (2013). A respecified fama french three-factor model for the new european union member states. *Journal of International Financial Management & Accounting*, 24(1):3–25.

Freeman, R. E. and Reed, D. L. (1983). Stockholders and stakeholders: A new perspective on corporate governance. *California Management Review*, 25(3):88–106.

French, K. R. (2018). Fama French 12-industry classification, available at: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french//Data_Library/det_12_ind_port.html`. urldate = 2018-12-10.

French, K. R. (2019). Fama French monthly European five factor files, available at: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Developed`. urldate = 2019-12-28.

Fudenberg, D., Holmstrom, B., and Milgrom, P. (1990). Short-term contracts and long-term agency relationships. *Journal of Economic Theory*, 51(1):1–31.

Gao, R. and Yu, X. (2018). How to measure capital investment efficiency: a literature synthesis. *Accounting and Finance*, 60(1):299–334.

Gillan, S. L., Hartzell, J. C., Koch, A., and Starks, L. T. (2018). Getting the incentives right: Backfilling and biases in executive compensation data. *Review of Financial Studies*, 31(4):1460–1498.

Gleason, C. A., Jenkins, N. T., and Johnson, W. B. (2008). The contagion effects of accounting restatements. *The Accounting Review*, 83(1):83–110.

Gopalan, R., Milbourn, T., Song, F., and Thakor, A. V. (2014). Duration of executive compensation. *The Journal of Finance*, 69(6):2777–2817.

Götz, A. and Friese, N. (2010). Empirische analyse der vorstandsvergütung im dax und mdax nach einführung des vorstandsvergütungsangemessenheitsgesetzes [empirical analysis of executive compensation in dax and mdax following the adoption of the appropriate director compensation act]. *Corporate Finance biz*, 1(6):410–420.

Götz, A. and Friese, N. (2011). Vorstandsvergütung im dax und mdax: Weiterführung der empirischen analyse 2010 nach einführung des vorstandsvergütungsangemessenheitsgesetzes [executive compensation in dax and mdax: continuation of the 2010 empirical analysis following the adoption of the appropriate director compensation act]. *Corporate Finance biz*, 2(8):498–508.

Gow, I. D., Larcker, D. F., and Reiss, P. C. (2016). Causal inference in accounting research. *Journal of Accounting Research*, 54(2):477–523.

Graham, J. R., Harvey, C. R., and Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics*, 40(1-3):3–73.

Gray, S. J. (1988). Towards a theory of cultural influence on the development of accounting systems internationally. *Abacus*, 24(1):1–15.

Green, B. P. and Choi, J. H. (1997). Assessing the risk of management fraud through neural network technology. *Auditing: A Journal of Practice & Theory*, 16:14–28.

Grossman, S. J. and Hart, O. D. (1980). Disclosure laws and takeover bids. *The Journal of Finance*, 35(2):323–334.

Hail, L., Tahoun, A., and Wang, C. (2018). Corporate scandals and regulation. *Journal of Accounting Research*, 56(2):617–671.

He, G. (2015). The effect of ceo inside debt holdings on financial reporting quality. *Review of Accounting Studies*, 20(1):501–536.

He, S. and Narayanamoorthy, G. G. (2020). Earnings acceleration and stock returns. *Journal of Accounting and Economics*, 69(1):article 101238.

Hitz, J.-M. and Müller-Bloch, S. (2015). Market reactions to the regulation of executive compensation. *European Accounting Review*, 24(4):659–684.

Holmstrom, B. and Costa, J. R. I. (1986). Managerial incentives and capital management. *Quarterly Journal of Economics*, 101(4):835–860.

Hsu, A. W.-h., Jung, B., and Pourjalali, H. (2015). Does international accounting standard no. 27 improve investment efficiency? *Journal of Accounting, Auditing & Finance*, 30(4):484–508.

Huber, W. (2017). Forensic accounting, fraud theory, and the end of the fraud triangle. *Journal of Theoretical Accounting Research*, 12(2):28–48.

Human-Capital-Management-Coalition (2019). Human Capital Management Coalition's final comment submission to the U.S. Securities and Exchange Commission regarding its proposed rulemaking, "Modernization of Regulation S-K Items 101, 103, and 105" (Exchange Act Release No. 86614), available at: `https://uawtrust.org/AdminCenter/Library.Files/Media/501/AboutUs/HCMCoalition/hcmc-commentsubmission-oct2019.pdf`. urldate = 2021-03-25.

Huson, M. R., Tian, Y., Wiedman, C. I., and Wier, H. A. (2012). Compensation committees' treatment of earnings components in ceos' terminal years. *The Accounting Review*, 87(1):231–259.

Jaggia, P. B. and Thakor, A. V. (1994). Firm-specific human capital and optimal capital structure. *International Economic Review*, pages 283–308.

Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review*, 76(2):323–329.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305–360.

Jensen, M. C. and Murphy, K. J. (1990). Ceo incentives—it's not how much you pay, but how. *Harvard Business Review*, 68(3):138–153.

Jin, J. Y., Kanagaretnam, K., and Lobo, G. J. (2011). Ability of accounting and audit quality variables to predict bank failure during the financial crisis. *Journal of Banking and Finance*, 35(11):2811–2819.

Jochem, T., Ladika, T., and Sautner, Z. (2018). The retention effects of unvested equity: Evidence from accelerated option vesting. *Review of Financial Studies*, 31(11):4142–4186.

John, K. and Senbet, L. W. (1998). Corporate governance and board effectiveness. *Journal of Banking and Finance*, 22(4):371–403.

Jones, M. J. (2010). *Creative accounting, fraud and international accounting scandals*. John Wiley & Sons.

Jones, S., Johnstone, D., and Wilson, R. (2017). Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks. *Journal of Business Finance & Accounting*, 44(1-2):3–34.

Kanodia, C., Sapra, H., and Venugopalan, R. (2004). Should intangibles be measured: What are the economic trade-offs? *Journal of Accounting Research*, 42(1):89–120.

Karpoff, J. M. (2021). The future of financial fraud. *Journal of Corporate Finance*, 66:101694.

Karpoff, J. M., Koester, A., Lee, D. S., and Martin, G. S. (2017). Proxies and databases in financial misconduct research. *The Accounting Review*, 92(6):129–163.

Kedia, S. and Rajgopal, S. (2011). Do the sec's enforcement preferences affect corporate misconduct? *Journal of Accounting and Economics*, 51(3):259–278.

Keynes, J. M. (1936). *The General Theory of Employment, Interest, and Money*. London, UK: Macmillan.

Klasa, S., Ortiz-Molina, H., Serfling, M., and Srinivasan, S. (2018). Protection of trade secrets and capital structure decisions. *Journal of Financial Economics*, 128(2):266–286.

Knyazeva, A., Knyazeva, D., and Masulis, R. W. (2013). The supply of corporate directors and board independence. *Review of Financial Studies*, 26(6):1561–1605.

Koh, P.-S. and Reeb, D. M. (2015). Missing r&d. *Journal of Accounting and Economics*, 60(1):73–94.

Kothari, S. P. and Zimmerman, J. L. (1995). Price and return models. *Journal of Accounting and Economics*, 20(2):155–192.

Kumar, P. and Sivaramakrishnan, K. (2008). Who monitors the monitor? the effect of board independence on executive compensation and firm value. *Review of Financial Studies*, 21(3):1371–1401.

Kumar, R. and Sopariwala, P. R. (1992). The effect of adoption of long-term performance plans on stock prices and accounting numbers. *Journal of Financial and Quantitative Analysis*, 27(4):561–573.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. (2000). Investor protection and corporate governance. *Journal of Financial Economics*, 58(1-2):3–27.

Ladika, T. and Sautner, Z. (2020). Managerial short-termism and investment: Evidence from accelerated option vesting. *Review of Finance*, 24(2):305–344.

Lai, S.-M., Liu, C.-L., and Wang, T. (2014). Increased disclosure and investment efficiency. *Asia-Pacific Journal of Accounting and Economics*, 21(3):308–327.

Lara, J. M. G., Osma, B. G., and Penalva, F. (2016). Accounting conservatism and firm investment efficiency. *Journal of Accounting and Economics*, 61(1):221–238.

Larcker, D. F. (1983). The association between performance plan adoption and corporate capital investment. *Journal of Accounting and Economics*, 5:3–30.

Larcker, D. F. and Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, 50(2):495–540.

Lee, K. H., Mauer, D. C., and Xu, E. Q. (2018). Human capital relatedness and mergers and acquisitions. *Journal of Financial Economics*, 129(1):111–135.

Leuz, C. and Wysocki, P. D. (2016). The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of Accounting Research*, 54(2):525–622.

Lev, B. (2019). Ending the accounting-for-intangibles status quo. *European Accounting Review*, 28(4):713–736.

Lev, B. and Schwartz, A. (1971). On the use of the economic concept of human capital in financial statements. *The Accounting Review*, 46(1):103–112.

Lev, B. and Sougiannis, T. (1996). The capitalization, amortization, and value-relevance of r&d. *Journal of Accounting and Economics*, 21(1):107–138.

Li, L., Monroe, G. S., and Coulton, J. (2018). Managerial litigation risk and corporate investment efficiency: Evidence from derivative lawsuits. *SSRN Working Paper*.

Li, Z. and Wang, L. (2016). Executive compensation incentives contingent on long-term accounting performance. *Review of Financial Studies*, 29(6):1586–1633.

Lin, C., Officer, M. S., Schmid, T., and Zou, H. (2019). Is skin in the game a game changer? evidence from mandatory changes of d&o insurance policies. *Journal of Accounting and Economics*, 68(1):Article 101225.

Litsoukov, N. (2015). BaFin and the SEC: An international comparison of supervisory practice. *BaFin - Expert Articles*.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

MacMinn, R. D. and Han, L.-M. (1990). Limited liability, corporate value, and the demand for liability insurance. *Journal of Risk and Insurance*, pages 581–607.

Marshall, H., Hewton, J., Altman, P. I., Asaro, M. A., James, J. B. J., Butler, P. W., Connolly, C. F., Díaz, E., Goldstein, K. R., Modesti, C. B., and Moyne, P. D. (2021). The FCA And SEC Annual Reports – A Statistical Comparison. *Financial Regulatory Alert Akin Gump Strauss Hauer & Feld LLP*.

Maurer, M. (2021). Companies Offer Investors a Glimpse at Employee Turnover. *The Wall Street Journal*, available at: `https://www.wsj.com/articles/companies-offer-investors-\a-glimpse-at-employee-turnover-11616405401?` urldate = 2021-03-25.

McNichols, M. F. and Stubben, S. R. (2008). Does earnings management affect firms' investment decisions? *The Accounting Review*, 83(6):1571–1603.

Modigliani, F. and Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review*, 48(3):261–297.

Myers, S. C. and 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.

O'Byrne, S. F. and Rajgopal, S. (2022). Employee value added: A new measure of gain-sharing between labor and capital. *Journal of Applied Corporate Finance*, 34(2):30–44.

Pantzalis, C. and Park, J. C. (2009). Equity market valuation of human capital and stock returns. *Journal of Banking and Finance*, 33(9):1610–1623.

Parrino, R. (1997). Ceo turnover and outside succession a cross-sectional analysis. *Journal of Financial Economics*, 46(2):165–197.

Pasch, L. (2017). What drives misstatements in financial statements? evidence from germany. *SSRN Working Paper*.

Penman, S. H. (2003). The quality of financial statements: Perspectives from the recent stock market bubble. *Accounting Horizons*, 17:77–96.

Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice & Theory*, 30(2):19–50.

Perols, J. L., Bowen, R. M., Zimmermann, C., and Samba, B. (2017). Finding needles in a haystack: Using data analytics to improve fraud prediction. *The Accounting Review*, 92(2):221–245.

Petrin, M. (2015). Executive compensation in the united kingdom–past, present, and future. *SSRN Working Paper*.

Prentice, R. A. (2005). The inevitability of a strong sec. *Cornell Law Review*, 91:775.

Purda, L. and Skillicorn, D. (2015). Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research*, 32(3):1193–1223.

Rajan, R. and Zingales, L. (1998). Financial dependence and growth. *American Economic Review*, 88(3):559–586.

Rakoff, J. S. (2014). The financial crisis: why have no high-level executives been prosecuted? *The New York Review of Books*, 9(2):7.

Richardson, S. A., Sloan, R. G., Soliman, M. T., and Tuna, I. (2005). Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics*, 39(3):437–485.

Roe, M., Spamann, H., Fried, J., and Wang, C. (2020). The european commission's sustainable corporate governance report: A critique. *SSRN Working Paper*.

Rose-Ackerman, S. (1991). Risk taking and ruin: Bankruptcy and investment choice. *The Journal of Legal Studies*, 20(2):277–310.

Rosett, J. G. (2003). Labour leverage, equity risk and corporate policy choice. *European Accounting Review*, 12(4):699–732.

Ross, S. A. (2004). Compensation, incentives, and the duality of risk aversion and riskiness. *The Journal of Finance*, 59(1):207–225.

Rouen, E. (2019). The problem with accounting for employees as costs instead of assets. *Harvard Business Review*.

Rouen, E. (2020). Rethinking measurement of pay disparity and its relation to firm performance. *The Accounting Review*, 95(1):343–378.

Roychowdhury, S., Shroff, N., and Verdi, R. S. (2019). The effects of financial reporting and disclosure on corporate investment: A review. *Journal of Accounting and Economics*, 68(2-3):101246.

Schiemann, F. and Guenther, T. (2013). Earnings predictability, value relevance, and employee expenses. *International Journal of Accounting*, 48(2):149–172.

Schroth, J. (2018). Managerial compensation and stock price manipulation. *Journal of Accounting Research*, 56(5):1335–1381.

Shailer, G. (2018). Corporate governance. *In: Poff, D., Michalos, A. (eds) Encyclopedia of Business and Professional Ethics*, pages 1–6.

Shleifer, A. and Vishny, R. W. (1997). A survey of corporate governance. *The Journal of Finance*, 52(2):737–783.

Stein, J. C. (1988). Takeover threats and managerial myopia. *Journal of Political Economy*, 96(1):61–80.

Stein, J. C. (2003). Agency, information and corporate investment. In *Handbook of the Economics of Finance*, volume 1, pages 111–165. Elsevier.

Stiglitz, J. E. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Economic Review*, 71(3):393–410.

Storbeck, O. (2020). How a dream to buy Deutsche Bank turned into a nightmare. *Australian Financial Review*, available at: `https://www.afr.com/companies/financial-services/how-a-dream-to-buy-deutsche-bank-turned-into-a-nightmare-20200825-p55pad`. urldate = 2022-12-04.

Summers, S. L. and Sweeney, J. T. (1998). Fraudulently misstated financial statements and insider trading: An empirical analysis. *The Accounting Review*, 73(1):131–146.

Tam, K. Y. and Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science*, 38(7):926–947.

Verdi, R. S. (2006). Financial reporting quality and investment efficiency. *SSRN Working Paper*.

Viswanathan, V. and Chopra, K. (2015). Study on factors affecting job satisfaction, loyalty, and commitment among managerial staff with reference to e-publishing organisations in ites/bpm industry. *International Journal for Research in Applied Science & Engineering Technology*, 3(7):458–465.

von Thadden, E.-L. (1995). Long-term contracts, short-term investment and monitoring. *Review of Economic Studies*, 62(4):557–575.

Walker, S. (2021a). Critique of an article on machine learning in the detection of accounting fraud. *Econ Journal Watch*, 18(2):61–70.

Walker, S. (2021b). Rejoinder to the critique of an article on machine learning in the detection of accounting fraud. *Econ Journal Watch*, 18(2):230–234.

Walker, S. (2022). Erroneous erratum to accounting fraud article. *Econ Journal Watch*, 19(2):190–203.

Wurgler, J. (2000). Financial markets and the allocation of capital. *Journal of Financial Economics*, 58(1-2):187–214.

Zingales, L. (2000). In search of new foundations. *The Journal of Finance*, 55(4):1623–1653.

Zingales, L. (2009). The future of securities regulation. *Journal of Accounting Research*, 47(2):391–425.

Appendix A

Appendix A.1 - Variable definitions

Variable	Definition and Thomson Reuters Datastream mnemonic
<u>Investment variables:</u>	
C/PPE_t	CAPEX (WC04601) scaled by average PPE (WC02501)
C/PPE_{Rev}	Residual from regressing C/PPE based on the accelerator model, Equation (4)
C/PPE_Q	Residual from regressing C/PPE based on Tobin's q model, Equation (5)
I/TA_t	CAPEX (WC04601) plus R&D (WC01201) over average total assets (WC02999), whereby missing values for R&D are set to 0 when CAPEX are available
I/TA_{Rev}	Residual from regressing I/TA based on the accelerator model, Equation (4)
I/TA_Q	Residual from regressing I/TA based on Tobin's q model, Equation (5)
<u>Investment model variables:</u>	
$RevCh_{t-1}$	Change in revenues (WC01001) scaled by previous year's revenues
$RevChN_{t-1}$	Indicator variable set to 1 when the change in revenues is negative, 0 otherwise
Q_{t-1}	Beginning-of-year enterprise value (WC18100) divided by beginning-of-year total assets (WC02999)
CFO_t	Cash flow from operating activities (WC04860) over average total assets (WC02999)
<u>Control variables:</u>	
$Lever_{t-1}$	Leverage is calculated as beginning-of-year total debt (WC03255) over beginning of year total assets (WC02999)
$Size_{t-1}$	Natural logarithm of a firm's number of employees (WC07011) as of the beginning of the year
Z_t	Altman's Z score is calculated as (1.2 times beginning-of-year working capital (WC03151) + 1.4 times retained earnings (WC03495) + 3.3 times EBIT (WC18191) + 0.999 times revenues (WC01001) / beginning-of-year total assets (WC02999) + 0.6 times (beginning-of-year market capitalization (WC08001) / beginning-of-year total liabilities (WC03351))
$Regul$	Indicator variable set to 1 for regulated industries such as utilities (SIC 49) and airlines & railroads (SIC 40-47)
$Homog$	Variable measuring industry homogeneity based on the data by Parrino (1997), for industries with missing data I calculate the average of the next higher SIC level
$Tangb_{t-1}$	Asset tangibility is calculated as beginning-of-year PPE (WC02501) over beginning-of-year total assets (WC02999)
$Slack_{t-1}$	Financial slack is calculated as beginning-of-year Cash (WC02003), if missing, Cash generic (WC02005), over beginning-of-year total assets (WC02999)
<u>Cross-sectional variables:</u>	
Gov	Dummy variable coded to 1 for firms that are in the lowest quartile of German firms' 2007 corporate governance pillar score (CGSCORE)
CCI	Dummy variable coded to 1 for firms that are in the lowest quartile of German firms' 2007 compensation committee independence (CGBFO04V)

Appendix A.2 - Changes in remuneration systems - exemplary extract from online Supplement

This appendix lists hand-collected indications of how German firms implemented the VorstAG as stipulated in their annual reports, exemplary for three firms. The online Supplement at the journal's Taylor and Francis website provides the full list of descriptions. The change categories numbered from 1 to 6 refer to the summary of frequencies tabulated in Appendix A.3.

BASF SE: In the second half of 2009, the compensation system for Board members was aligned with the German Act on the Appropriateness of Management Board Remuneration (...) Contracts with all Board members, regardless of existing contractual terms, were consensually and uniformly converted to this system effective as of January 1, 2010. (...) The main objectives of this were to give greater relevance to the component based on long-term success and to make compensation for the Board of Executive Directors even more performance-oriented. (...) The information presented below relates to the compensation system valid until December 31, 2009 and the respective compensation of Board members for 2009. (...) The return on assets (ROA) is used to determine the variable compensation of all employee groups and also determines - 2009 for the last time as sole basis - the Board members' annual variable compensation (variable bonus). (AR 2009) The annual variable compensation (variable bonus) of the Board of Executive Directors is based on the performance of the entire Board and the return on assets. (...) The Supervisory Board assesses the achievement of goals in relation to the last three years. (AR 2010) *Change categories: 3,1*

Deutsche Post AG: The remuneration component linked to the company's annual profits now also includes a sustainability component in line with the provisions of the (...) (VorstAG) (...) the annual performance related remuneration will in future no longer be paid in full for the year on the basis of having reached the agreed targets. Instead, 50 % of the annual performance-related remuneration will flow into a new medium-term component with a three-year calculation period (performance phase of one year, sustainability phase of two years). (AR 2009) *Change category: 2*

Wincor Nixdorf AG: We plan to submit a new share option program for members of the Board of Directors and others with a subscription entitlement for approval by the AGM in 2010. The changes in the new program, which is based on the existing one, are intended, in part, to ensure compliance with (...) (AR 2008/2009) In accordance with the requirements of Germany's VorstAG Act, the vesting period for share options granted under the 2010 share option program has been extended from two to four years. (AR 2009/2010) *Change category: 4*

Appendix A.3 - Summary of changes in remuneration systems

Categories of change in remuneration system	Frequency of change Main change	All changes
1 Multi-year performance assessment[a]	15	18
2 Deferral of a portion of variable components[a]	11	17
3 Gear performance-related components to long-term determinants[b]	7	8
4 Increase in vesting period of stocks or stock options[c]	4	12
5 Adoption of a bonus-malus-system[d]	4	6
6 Requirement of investment in own stocks or stock options[e]	0	11
() Apparently no or likely very minor change	5	5
Total	46	77

This appendix lists the categories of the changes in the remuneration systems observed in the 46 treated German firms. *Main change* refers to the frequency of a particular category being regarded as the main change in the executive compensation system (i.e., the first change category listed per firm in the online Supplement) and *All changes* refers to the total frequency of the particular category.

[a]Mostly refers to the adoption of such a multi-year assessment system or a deferral of awarded variable compensation, and sometimes implies that an existing system was used for a larger portion of the components.

[b]Refers to changing the determinants or lifting the ratio of long-term to short-term components.

[c]Mostly refers to an increase in the vesting period to four years. This is often implemented in addition to a main change in a component of the system.

[d]Only refers to instances where this is precisely stipulated; however, it is likely that some more multi-year assessment or deferral schemes contain a bonus-malus or conditional payback provision.

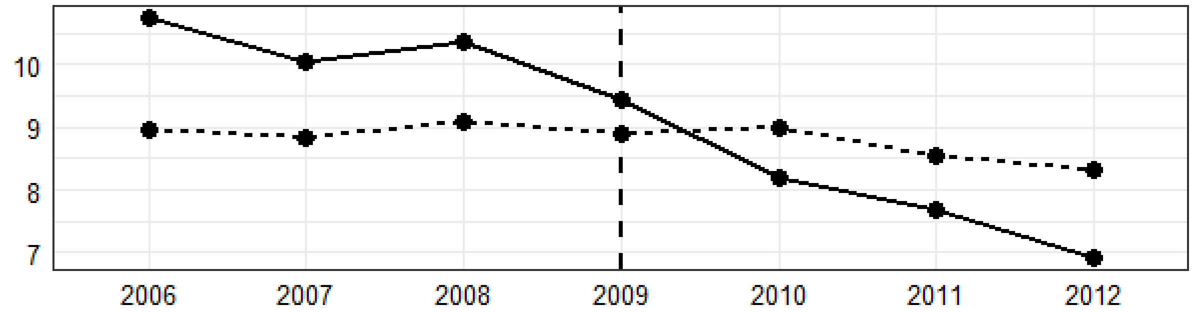
[e]Is never implemented as the main change and is mostly implemented in addition to deferral provisions.

Appendix A.4 - List of private German firms

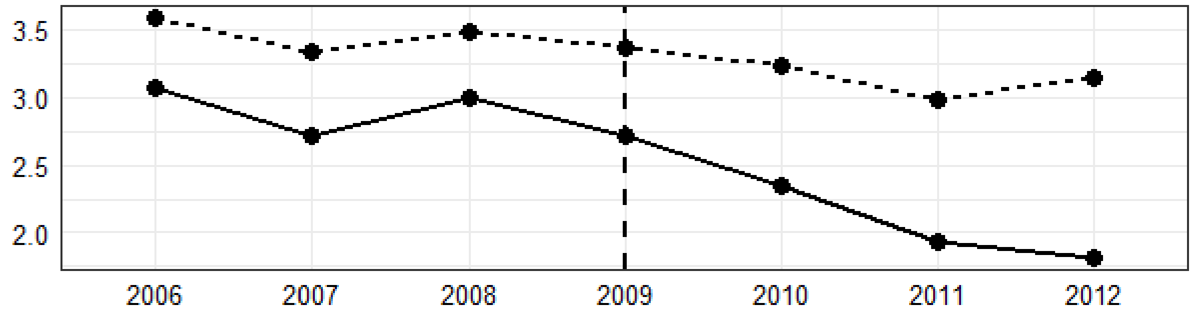
Adolf Würth GmbH & Co. KG, DATEV eG, Dehner Gartencenter GmbH & Co. KG, Diehl Stiftung & Co. KG, Dr. Johannes Heidenhain GmbH, E.G.O. Blanc und Fischer & Co. GmbH, Eberspächer Gruppe GmbH & Co. KG, EHG Service GmbH, Fiege Logistik Holding Stiftung & Co. KG, Friedhelm Loh Stiftung & Co. KG, Fritz Dräxlmaier GmbH & Co. KG, Gebr. Knauf KG, Georgsmarienhütte Holding GmbH, Häfele GmbH & Co. KG, Hettich Holding GmbH & Co. oHG, Hoyer GmbH Internationale Fachspedition, I.K. Hofmann GmbH, K+K Klaas & Kock B.V. & Co. KG, KAEFER Isoliertechnik GmbH & Co. KG, Katharina Kasper Holding GmbH, L. Possehl & Co. mbH, Lohmann GmbH & Co. KG, MAHLE GmbH, Mann + Hummel Holding GmbH, MESSER Group GmbH, Piepenbrock Unternehmensgruppe GmbH + Co. KG, Rohde & Schwarz GmbH & Co. KG, SRH-Holding SdbR, TRUMPF GmbH + Co. KG, Vorwerk & Co. KG, W-E-G GmbH & Co. KG, Wilh. Werhahn KG

Appendix A.5 - Pattern of the average dependent variable per treatment and control group

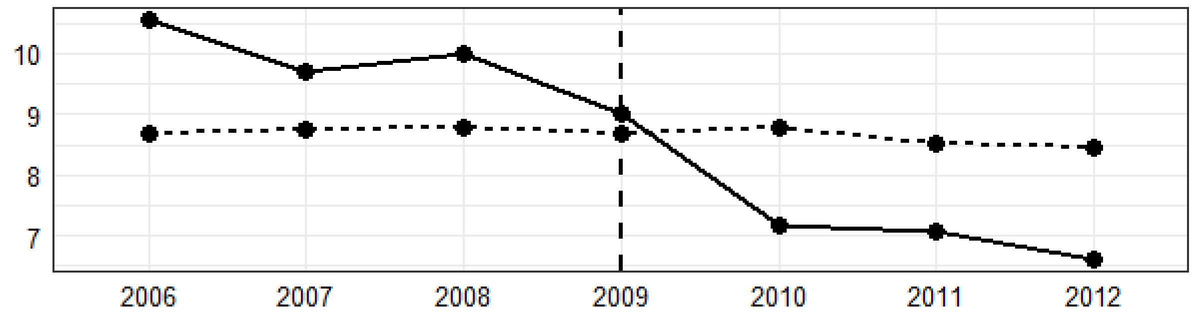
Panel A: $|C/PPE_{Rev}|$



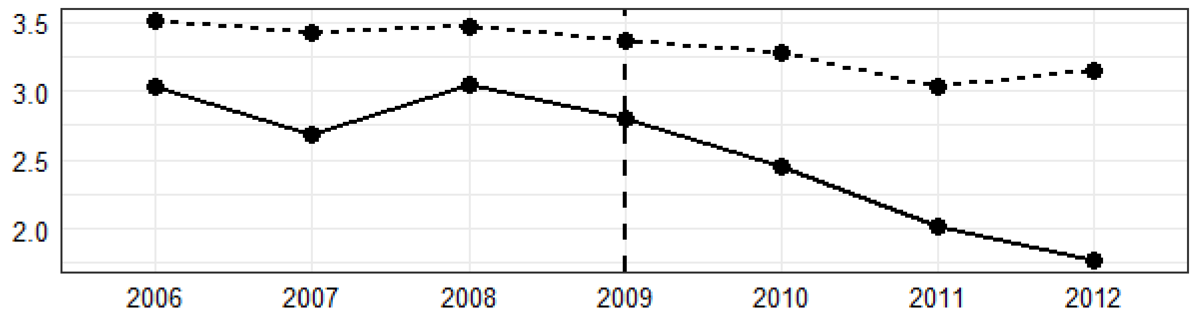
Panel B: $|I/TA_{Rev}|$



Panel C: $|C/PPE_Q|$



Panel D: $|I/TA_Q|$



This appendix plots the average abnormal investment dependent variable per treatment (solid line) and control group (dashed line) to visualize the treatment effect by year over the sample period. Panels A and B map the coefficients for $|C/PPE_{Rev}|$ and $|I/TA_{Rev}|$, which are the absolute values of the regression residuals based on the accelerator model using the two alternative investment measures CAPEX scaled by average PPE and the sum of CAPEX and R&D scaled by average total assets. Panels C and D show results for both investment measures and the alternative Tobin's q approach (i.e., $|C/PPE_Q|$ and $|I/TA_Q|$).

Appendix A.6 - Alternative differentiation between over-investment and under-investment

Panel A: Full sample

Dependent variable:	$\lvert C/PPE_{Rev} \rvert$	$\lvert C/PPE_{Q} \rvert$	$\lvert I/TA_{Rev} \rvert$	$\lvert I/TA_{Q} \rvert$
	(1)	(2)	(3)	(4)
Treat*Post*ResidN-Pre	**4.342*****	**3.878*****	**0.582**	**0.878****
	(1.492)	**(1.197)**	**(0.487)**	**(0.431)**
Treat*Post	**-4.272*****	**-4.310*****	**-0.632***	**-0.758****
	(0.960)	**(0.833)**	**(0.353)**	**(0.384)**
*Post*ResidN-Pre*	1.720**	1.157*	0.981***	1.038***
	(0.828)	(0.644)	(0.217)	(0.180)
Post	−0.721	−0.460	−0.728***	−0.685***
	(0.687)	(0.579)	(0.194)	(0.185)
Firm FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	2,477	2,477	2,477	2,477
Adjusted R^2	0.441	0.469	0.570	0.598

Panel B: PSM subsample

Dependent variable:	$\lvert C/PPE_{Rev} \rvert$	$\lvert C/PPE_{Q} \rvert$	$\lvert I/TA_{Rev} \rvert$	$\lvert I/TA_{Q} \rvert$
	(1)	(2)	(3)	(4)
Treat*Post*ResidN-Pre	**5.728*****	**5.131*****	**0.637**	**0.920****
	(1.703)	**(1.462)**	**(0.490)**	**(0.455)**
Treat*Post	**-5.529*****	**-5.571*****	**-0.945*****	**-1.144*****
	(1.393)	**(1.172)**	**(0.359)**	**(0.369)**
*Post*ResidN-Pre*	0.795	−0.170	0.900***	0.960***
	(1.304)	(1.031)	(0.295)	(0.250)
Post	0.453	0.707	−0.401	−0.350
	(1.264)	(1.035)	(0.294)	(0.252)
Firm FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	523	523	523	523
Adjusted R^2	0.626	0.650	0.493	0.527

This appendix presents the results of the treatment effect of longer duration on former over-investors and former under-investors. *ResidN-Pre* is coded 1 if the average investment residual of a firm in the pre-period is negative and 0 otherwise. The coefficient of the DID interaction (i.e., the second line) measures the reduction in abnormal investment of former over-investors following treatment. The total of the coefficient of the three-way interaction (i.e., the first line) and the DID interaction measures the reduction in abnormal investment of former under-investors following treatment. In both panels, the four columns correspond to the combinations of the two alternative investment measures with the two alternative investment models. First-step investment model regressors, other control variables, intercept, and fixed effect dummies are not tabulated. All variables are defined in Appendix A.1. Standard errors are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Appendix B

Appendix B.1 - Variable definitions

Variable	Definition and Thomson Reuters Datastream mnemonic
$TA_{i,t}$	Average of beginning $(t-1)$ and end of year (t) total assets (WC02999)
$PE_{i,t}$	Personnel expense for all employees and officers (WC01084)
$OI_{i,t}^{BDNAPE}$	Operating income (WC01250) before depreciation & amortization (WC01151) and PE used in optimal lag structure and future value regressions
$OI_{i,t}^{BPE}$	Operating income before PE used in contemporaneous price analyses
$PE/TA_{i,t}$	$PE_{i,t}$ scaled by average total assets before instrumental variable approach
$(PE/TA_{predicted})_{i,t}$	Value predicted through instrumental variable approach
$OI/TA_{i,t}$	$OI_{i,t}^{BDNAPE}$ scaled by average total assets
$\#E_{i,t}$	End of year number of employees (WC07011)
$PEFV_{i,t}$	The personnel expenditure future value, which is the firm-year-specific sum of the discounted coefficients on lagged PE
$PEFV - Decile_{i,t}$	Deciles of $PEFV_{i,t}$ scaled to range from zero to one
$PEFV * PE/TA_{i,t}$	$PEFV_{i,t}$ multiplied with $PE/TA_{i,t}$ used in portfolio analyses
suffix $-PS_{i,t}$	End of year shares outstanding (indirect calculation dividing market capitalization (WC08001) by share price (P))
$OIPS_{i,t}$	$OI_{i,t}^{BPE}$ divided by shares outstanding (in US$)
$PEPS_{i,t}$	$PE_{i,t}$ divided by shares outstanding (in US$)
$RNDPS_{i,t}$	R&D expenses (WC01201, set to zero if missing) per share (in US$)
$SGAPS_{i,t}$	SG&A expenses (WC01101) excluding R&D per share (in US$)
$P_{i,t}$	End of year stock price (P, in US$)
$EPS_{i,t}$	Mean consensus earnings per share forecast (EPS1MN, in US$)
$FE_{i,t}$	Actual earnings per share (EPSIBES, in US$) minus mean consensus earnings per share forecast, also used in absolute terms ($\lvert FE_{i,t}\rvert$)
$MTB_{i,t}$	End of year market-to-book ratio (MTBV)
$Tangibility_{i,t}$	End of year property, plant & equipment (WC02501) scaled by total assets
$MeanPay_{i,t}$	PE (in US$) divided by number of employees
$TrainingDays_{i,t}(\%)$	Employee training hours (SOTDDP018) divided by 8 (hours) and 230 (working days) multiplied by 100
$R_{i,\tau}$	Firm-level return in month τ (obtained with mnemonic RI, in US$)
$R_{p,\tau}$	Return of portfolio p in month τ
$R_{f,\tau}$ and $R_{m,\tau}$	Monthly risk-free and market return (from K. French's library)
SMB_τ, HML_τ, RMW_τ, CMA_τ, and MOM_τ	Monthly size, value, operating profitability, investment aggressiveness, and momentum factor return (from K. French's library)
$R_{Ind,\tau}$	Average industry-level (FF12) return in month τ
$Momentum$	Momentum for each month τ, measured as the cumulative return from $\tau-1$ to τ ($Momentum_{-1,0}$) and $\tau-12$ to $\tau-2$ ($Momentum_{-12,-2}$), respectively
$Accruals_{i,t}$	Accruals measured as net income (WC01651) less net cash from operations (WC04860) scaled by book equity (total assets - total liabilities (WC02003))
$AssetGrowth_{i,t}$	Change in total assets from $t-1$ to t scaled by $t-1$
$log(BE/ME)_{i,t}$	Natural logarithm of book equity divided by market capitalization
$log(ME)_{i,t}$	Natural logarithm of market capitalization (MV) as of June $t+1$ (in US$)
$EBITDA/TA_{i,t}$	EBITDA (WC18198) scaled by average total assets

Appendix B.2 - *SGAFV* robustness analysis

	Dependent variable:			
	$P_{i,t}/P_{i,t-1}$			
	(1)	(2)	(3)	(4)
$OIPS_{i,t}/P_{i,t-1}$	0.147	0.153	0.153	0.158
	(0.152)	(0.154)	(0.146)	(0.148)
$PEPS_{i,t}/P_{i,t-1}$	−0.043	−0.056	−0.139	−0.146
	(0.155)	(0.155)	(0.140)	(0.141)
$\boldsymbol{PEFV_{i,t}/P_{i,t-1}}$		**0.058******		**0.051*****
		(0.022)		**(0.022)**
$\boldsymbol{SGAFV_{i,t}/P_{i,t-1}}$	**0.100****	**0.025**	**0.086**	**0.022**
	(0.054)	**(0.041)**	**(0.053)**	**(0.041)**
$EPS_{i,t}/P_{i,t-1}$	2.226***	2.237***	2.325***	2.332***
	(0.292)	(0.293)	(0.268)	(0.270)
$SGAPS_{i,t}/P_{i,t-1}$			0.123***	0.117***
			(0.042)	(0.040)
$RNDPS_{i,t}/P_{i,t-1}$			1.748***	1.703***
			(0.576)	(0.574)
Intercept	0.699***	0.685***	0.694***	0.682***
	(0.056)	(0.058)	(0.056)	(0.058)
F12 dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Observations	2,793	2,793	2,793	2,793
Adjusted R^2	0.382	0.385	0.394	0.397

This table reports the results of OLS regression of contemporaneous stock price on *PEFV* and *SGAFV* to test whether *PEFV* is incremental to *SGAFV*. Two-way-cluster robust standard errors, clustering at the firm and year levels, are shown in parentheses. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

We calculate *SGAFV* for firm-years within our sample of *PEFV* firm-years with sufficient SG&A data. We use the same instrumental variables approach as in our *PEFV* calculation. We further use the same optimal lag structure on the FF12-industry-level. For the regressions in this table, we focus on the firm-years where both *PEFV* and *SGAFV* are larger than zero.

Appendix C

Appendix C.1 - Descriptions of 28 raw financial data items

Variable	Description	Variable	Description
Balance sheet items:		Income statement items:	
che	Cash and short-term investments	*sale*	Sales/turnover (net)
rect	Receivables - total	*cogs*	Cost of goods sold
invt	Inventories - total	*dp*	Depreciation and amortization
ivst	Short-term investments - total	*xint*	Interest and related expense - total
act	Current assets - total	*txt*	Income taxes - total
ppegt	Prop., plant and equipm. - total (gross)	*ib*	Income before extraordinary items
ivao	Investment and advances - other	*ni*	Net income (loss)
at	Assets - total	Cash flow statement items:	
ap	Accounts payable - Trade	*sstk*	Sale of common and preferred stock
dlc	Debt in current liabilities - total	*dltis*	Long-term debt - issuance
txp	Income taxes payable	Market value items:	
lct	Current liabilities - total	*prcc_f*	Price close - annual - fiscal
dltt	Long-term debt - total	*csho*	Common shares outstanding
lt	Liabilities - total		
ceq	Common/ordinary equity - total		
pstk	Preferred/preference stock (capital) - total		
re	Retained earnings		

We match *ni* in Compustat with *nicon* in Compustat Global. We replace *ni* with *ib*, if *ni* is missing. Further, Compustat Global does not contain a match for *prcc_f*. We obtain *prcc_f* and *csho* via aggregating *cshoc* * *prccd* over possibly several share classes from the Compustat Global daily file for the respective fiscal year end date.

Appendix C.2 - Descriptions of 14 financial ratios calculated with 28 raw items

Variable	Calculation	Description
Ratios used in the Basic Dechow et al. (2011) LR model:		
rsst_acc	$(wc - lag\_wc + nco - lag\_nco + fin - lag\_fin) * 2 / (at + lag\_at)$	RSST accruals (Richardson et al., 2005)
ch_rec	$(rect - lag\_rect) * 2 / (at + lag\_at)$	Change in receivables
ch_inv	$(inv - lag\_inv) * 2 / (at + lag\_at)$	Change in inventory
soft_assets	$(at - ppent - che) / at$	Soft assets [%] (based on findings in Barton and Simko (2002))
ch_cs	$(cs - lag\_cs) / lag\_cs$	Change in cash sales
ch_roa	$roa - lag\_roa$	Change in return on assets
issue	IF $sstk > 0 \mid dltis > 0$ THEN $issue = 1$; ELSE $issue = 0$	Actual issuance
Additional ratios used in the Bao et al. (2020) LR model:		
wc_acc	$(wc - lag\_wc) * 2 / (at + lag\_at)$	Working capital accruals
ch_cm	$(cm - lag\_cm) / lag\_cm$	Change in cash margin
bm	$ceq / (prcc\_f * csho)$	Book-to-market ratio
dpi	$(lag\_dp / (lag\_dp + lag\_ppent)) / (dp / (dp + ppent))$	Depreciation index (Beneish, 1999)
reoa	re / at	Retained earnings over total assets (Summers and Sweeney, 1998)
EBIT	$(ni + xint + txt) / at$	EBIT over total assets (Summers and Sweeney, 1998)
ch_fcf	$ch\_ib - rsst\_acc$	Change in free cash flows
Supplementary calculations:		
wc	$(act - che) - (lct - dlc - txp)$	Non-cash working capital
nco	$(at - act - ivao) - (lt - lct - dltt)$	Non-current oprating assets
fin	$(ivst + ivao) - (dltt + dlc + pstk)$	Net financial assets
cs	$sale - (rect - lag\_rect)$	Cash sales
roa	$ni * 2 / (at + lag\_at)$	Return on assets
cm	$1 - (cogs - (invt - lag\_invt) + (ap - lag\_ap)) / (sale - (rect - lag\_rect))$	Cash margin
ch_ib	$(ib - lag\_ib) * 2 / (at + lag\_at)$	Change in income before extraordinary items
ppent	Property, plant and equipment - total (net)	

Appendix C.3 - Differences in available resources and forensic capabilities to detect accounting misconduct

In this appendix, we illustrate the available resources and forensic capabilities to detect and examine potential accounting misconduct of the responsible authorities in Germany and the UK and compare these institutions and their powers to the U.S. SEC. We begin with a description of the SEC enforcement process to then show where the non-U.S. enforcers fall short.

United States:

In the U.S., SEC's Division of Enforcement carries out investigations of different potential securities law violations. The investigative process begins with a variety of internal (e.g., corporate filings reviews) or external leads. External leads may be tips (e.g., from whistleblowers) or referrals (e.g. from other regulators or stock exchanges), where the SEC collects and triages hundreds of thousands each year. If staff believes that a suggested investigation would have the potential to address misconduct consistent with SEC's enforcement priorities, it becomes a Matter Under Inquiry (MUI). If (after reviewing readily available information) staff believed there are sufficient indications of misconduct, SEC opens an investigation. The policy of closing MUIs or converting them to an investigation is to do so within sixty days, the MUI phase may also be skipped. During the investigation, staff reviews relevant documents obtained from either the target or other public sources, interviews witnesses, reviews auditors' work papers or subpoena documents from other entities involved with the target. There are formal requirements when issuing a subpoena, compelling testimony, or administering oaths, however, they may again be skipped. If staff then preliminary determines that a violation of securities laws has occurred, the SEC votes on whether to authorize enforcement action (after notifying the target of its intent to pursue an enforcement action (i.e., via a Wells notice) and providing it with the opportunity to file a response (i.e., a Wells submission)). Enforcement action (e.g., Accounting and Auditing Enforcement Releases) can occur via civil actions (i.e., a complaint against the target in the U.S. District Court) or administrative proceedings (i.e., a complaint is adjudicated by an internal administrative law judge), where SEC's history of obtaining successful judgement is astonishing. We refer to Blackburne et al. (2020) for more details on the enforcement process.

Germany:

In Germany, enforcement of accounting misconduct is performed in a two-tier system. In the first tier, the Financial Reporting Enforcement Panel (FREP), a privately organized institution, examines the compliance of financial statements with applicable financial reporting frameworks (only until end of 2021). In the second tier, the German Federal Financial Supervisory Authority (Bundesanstalt für Finanzdienstleistungsaufsicht, BaFin), is responsible for investigation and enforcement in the case of infringements in the first tier. FREP examines the most recent filings of publicly listed firms based on random sampling, specific indications (e.g., media coverage), and at the request of the BaFin. The annually updated enforcement priorities focus on typical reporting errors, particularly in relation to newly applicable standards following updates to IFRS. In terms of capacity, FREP seeks to carry out 80 examinations per year (ESMA, 2020). FREP's strategy and historic action suggests that it rather focuses on error prevention regarding the use of reporting standards and not intentional misreporting for whatever motive may drive it. BaFin's role is somewhat more comparable to SEC's. Both supervise security issuers and transactions on financial markets. While the SEC mainly supervises brokers, investment advisers, clearing houses, and stock exchanges, BaFin even appears as a universal supervisor unifying securities, banks, and insurance supervision. However, although appearing to have more responsibility, paradoxically, BaFin has no legal authority to file civil actions. In addition, the SEC's investigatory powers are more comprehensive as they include the right to issue a subpoena, demand records, and summon suspects or witnesses to a hearing. When comparing the total fines ordered by the SEC and the BaFin in the fiscal year 2014, SEC's 4.2 billion U.S. dollars dwarf BaFin's 4 million Euros. However, these figures cannot be directly compared as the BaFin is not able to conclude comprehensive settlements and because serious capital market law violations are sanctioned under German criminal law and not by the BaFin (Litsoukov, 2015). In a report investigating the enforcement of financial information of BaFin and FREP in the context of the recent Wirecard case, the European Securities and Markets Authority (ESMA, 2020) concludes that the power and resources of these institutions are not comprehensive to accurately investigate potential accounting misconduct. Among other things, BaFin does not have the required power and authority to request relevant information from auditors or other third parties to validate suspicions of criminal activities. Moreover, confidentiality requirements prevent an effective exchange of information between the two main institutions.

United Kingdom:

In the UK, different government bodies regulate and investigate domestic accounting misconduct cases. The two most prominent institutions are the Financial Conduct Authority (FCA, formerly Financial Services Authority) and the Serious Fraud Office (SFO). In line with the UK being one of the world's most important financial center, the FCA regulates the conduct of all financial markets and financial services firms and performs investigations if a potential serious misconduct is detected. Both the FCA and the SEC generally have enormous powers to suspend or cease trading of publicly listed firms, charge substantial fines to firms, and impose sanctions on individuals. However, when comparing the total fines ordered in 2020, SEC's 4.7 billion U.S. dollars are worth about 18 times as much as FCA's 190 million British pounds. This comparison again lags as the FCA considers individuals' or entities' ability to pay more directly and because the different figures also reflect different legal cultures (Marshall et al., 2021). The SFO, on the other hand, only takes on a small number of large economic crime cases that are usually complex. The SFO often collaborates with other departments and ministries, even internationally, and also has the right to prosecute. Furthermore, it can offer firms to enter a Deferred Prosecution Agreement (DPA), where both parties would agree on compensations and settlements. One of the more recent DPAs is between the SFO and Tesco. While accounting misconduct may be the reason for SFO investigation, this office mainly prosecutes cases related to other criminal charges involving bribery, corruption, and organized crime. One may therefore infer that the FCA has too much of a focus on financial services businesses and the SFO is responsible for too complex and serious criminal behavior to properly enforce accounting misconduct in the broad universe of overseen firms.

Appendix C.4 - Summaries and details of identified misconduct cases

In this appendix, we summarize the cases included in the European accounting misconduct sample. For each case, we assess the (fiscal) years where the misconduct schemes were most likely implemented. As most of the newspaper articles referred to in the summaries are behind paywalls, they are available from the authors upon request.

Autonomy Corporation PLC - GB0055007982

Autonomy Corporation PLC is an England-based software company that had traded on the London Stock Exchange until it was acquired by HP in October 2011. In 2012, HP announced that it was recording a USD 8.8bn impairment charge due to accounting improprieties, misrepresentation, and disclosure failures of the company it had bought. The U.S. SEC found that the company had inflated revenues through resellers of more than 30 transactions totaling nearly USD 200m from the first quarter of 2009 to the second quarter of 2011. The SEC also found that Autonomy backdated at least five reseller transactions to hit revenue targets. Former CEO Christopher Egan had backdated four out of the five agreements. Egan was also involved in some round-trip transactions to enable resellers to pay the company. Egan was later ordered to cease and desist and was fined a total of USD 923,391 by the SEC (SEC, November 15, 2016).

Based on the SEC's findings, we view this case to entail financial accounting misconduct in the years from 2009 to 2010.

Carrère Group - FR0000044422

Carrère Group is the second-largest French television production and distribution company based in Aubervilliers, France. In 2010, the company was investigated by the French Stock Market Authority AMF for irregularities in the company's accounts from 2005 to 2006 concerning misevaluation of liabilities of EUR 1.8m, dubious lease contracts, and non-declared transaction of shares by its main shareholder and founder. In 2012, the company was fined EUR 4m by the AMF after finding that the owner of the company had sold shares worth EUR 1.8m in mid-2007 while having privileged information about its struggles. It also came to light that the company had been using fictitious orders in which it artificially inflated its turnover by EUR 5m in 2006 thanks to two orders of literary exploitation rights. Moreover, the company inflated the value of its catalogue in its accounts. Eventually, the catalogue valued at EUR 209m was depreciated by 115m in 2007. The company also failed to declare in its 2006 accounts that the covenants of a EUR 14m loan were no longer respected. Therefore, the loan should have been recorded in the accounts as a "current debt" and not as a "non-current debt" (La Tribune, July 24, 2012; Challenges, July 24, 2012).

We view this case to entail financial accounting misconduct in the period ranging from 2005 to 2006.

Conergy AG - DE000A1KRCK4

Conergy AG was a German manufacturer of photovoltaic power plants, based in Hamburg, Germany. The company filed for insolvency in 2013. Numerous former executives of the company are accused of accounting misconduct by public prosecutors. In the fiscal year 2006, the company incorrectly reported revenues of its subsidiaries, resulting in a fictitious profit of EUR 46m. If the revenues of its subsidiaries had been recognized in accordance with accounting standards, the company would have posted a loss of EUR 2m. While the state court of Hamburg dropped the charges against the former CEO Dieter Ammer in 2015, legal proceedings against other former executives have been continued (TAZ, April 22, 2015; FAZ, October 29, 2015).

We view this case to entail financial accounting misconduct in the fiscal year 2006.

Eltel AB - SE0006509949

Eltel AB is a Nordic field service provider for communication and power networks that was listed on the Stockholm Exchange in February 2015. During the leadership of CEO Håkan Kirstein, the company has discovered possible historical inaccuracies in some of the projects' reports during the previous CEO's tenure. Earlier aggressive profit recognition and premature revenue recognition have resulted in several profit warnings. The company's board and auditor PwC reported previous CEO Axel Hjärne to the police in May 2017 (resigned in June 2016) regarding suspicions of accounting violation and/or fraud regarding the company's financial statements (SvD, February 21, 2017; Expressen, May 31, 2017).

With limited references and news coverage, we conclude that accounting misconduct affected the company's financial statements for the year 2015.

Eniro AB - SE0011256312

Eniro AB is a Nordic tech company incorporated in Stockholm, Sweden. In August 2014, its board of directors launched an investigation to validate the consolidated financial statements. The investigation uncovered inaccuracies concerning accrual accounting in fiscal year 2013, resulting in early revenue recognition. According to the company's 2014 annual report, the "[...] accrual errors were in the Desktop search and Mobile search revenue categories in the Local search segment and pertained to the allocation of discounts in advertising packages. These discounts were incorrectly allocated only to revenue that is allocated over time (subscription fees). As a result, the revenue recognized at the time of sale was too high, and thus the reported income for December 2013 was too high." (p. 74) The accounting malpractices resulted in an overstatement of 2013 gross operating revenue as well as operating income by SEK 72m and an overstatement of retained earnings by SEK 55m. Former CEO Johan Lindgren's severance package and other benefits were rescinded and he was reported to the police (FT, September 5, 2014; SvD, June 26, 2015).

We view this case to entail financial accounting misconduct in fiscal year 2013.

Globo PLC - GB00B282VW04

Globo PLC is a technology company engaged in enterprise mobility management and application development. In July 2015, a report by Quintessential Capital Management – Investigative Research Team, investigated the company's revenue model and suggested that the group's revenue was based on fictitious sales invoices generated by shell companies that it created and controlled since the group's listing in 2007 (Quintessential Capital Management – Investigative Research Team, July 18, 2015). In October 2015, both its CEO and CFO resigned after they revealed to the directors that there had been falsification of data and misrepresentation of the company's financial situation. A month later, the company went into administration. (In December 2015, the UK's Financial Reporting Council launched an investigation in relation to the audits of the company's financial statements for the years 2013 and 2014. However, the auditor Grant Thornton was not found to have committed misconduct (FRC, July 30, 2018).)

We infer the accounting misconduct period to be from 2007 to 2014. The company's fiscal year 2007 is not available, our sample therefore only contains the years from 2008 on.

Innoconcepts NV - NL0000361145

Innoconcepts NV was a Dutch listed company based in Capelle aan den IJssel. It was declared bankrupt in 2010. In March 2015, the Public Prosecutor's Office announced that it had launched a criminal investigation into three former directors of the company. They are suspected of fraud in the annual financial statements 2005 through 2008, swindling and forgery. Two of them are also alleged to have traded in shares with insider information in 2007. The trustee wants to recover a large part of the estate deficit from accounting firm Deloitte. The deficit was estimated at EUR 63.5m. The parties involved, including the accountants, settled in December 2015 for an amount of EUR 18m. On November 11, 2020, the Justice Department demanded prison sentences of up to 36 months and fines of EUR 1.5m and EUR 3.4m against the three former directors (Dutch News, February 25, 2011; De Accountant, December 11, 2015).

We view this case to entail financial accounting misconduct in the period ranging from 2005 to 2008.

Mariella Burani Fashion Group SPA - IT0001403739

Mariella Burani Fashion Group SPA, based in Cavriago, Italy, was a clothing, jewelry, and watches company. In February 2010, the company declared bankruptcy. In February 2013, its founders and other shareholders, Walter, and Giovanni Burani, were sentenced to six years in jail for fraudulent bankruptcy of the company for manipulating the stock market through setting up a series of fraudulent deals for the purpose of inflating the balance sheet and supporting the share price between 2007 and 2009. Bourse watchdog Consob fined the company, former managers, and a subsidiary EUR 6.7m (Reuters, July 28, 2010; Fashion Network, February 5, 2013; Fashion Network, May 30, 2013).

We view this case to entail financial accounting misconduct in the period ranging from 2007 to 2008.

MIFA Mitteldeutsche Fahrradwerke AG - DE000A0B95Y8

MIFA Mitteldeutsche Fahrradwerke AG was an East German bicycle manufacturer based in Sangerhausen, Germany. The company overvalued the balance sheet items raw, auxiliary, and operating materials as well as finished products in the fiscal year 2012. Consequently, the reported inventories of the fiscal year 2013 had to be corrected downward by EUR 15m (FAZ, May 15, 2014; FAZ, August 23, 2014; Handelsblatt, August 26, 2014). (Today, the company is privately owned and part of Zweirad Union e-Mobility GmbH.)

We follow that the financial accounting misconduct affected fiscal year 2012.

NMC Health PLC - GB00B7FC0762

NMC Health PLC was a private healthcare company based in the UK. In December 2019, Muddy Waters Research released a report on alleged theft of corporate assets and intentional manipulation of accounts (Muddy Waters Research, December 17, 2019). The report provided numerous findings, including reports of inflated costs related to the redevelopment of NMC Royal Women's Hospital in 2014 and debt understatement in 2018 by not reporting its Aspen Healthcare acquisition as a finance lease. Following the release of this report, it was discovered that there was an additional debt of USD 2.7bn that had not been approved by the company's board, on top of the USD 2.1bn debt the company had reported in June 2019 (Guardian, March 10, 2020). In February 2020, the company's shares were suspended from trading. At the end of February 2020, the Financial Conduct Authority launched an investigation into the company's finances. In April, the company collapsed and the audit regulator Financial Reporting Council announced an investigation into the audit of the company's 2018 accounts by its auditor EY (Guardian, May 4, 2020).

Although the investigations by the UK investigators Financial Reporting Council and Financial Conduct Authority are still ongoing, we conclude that this case entails financial accounting misconduct that affected the company's financial statements in the years 2014 to 2018. Our sample only contains the years until 2017.

Panalpina Welttransport AG - CH0002168083

Panalpina Welttransport AG is a logistics company based in Basel, Switzerland. At a subsidiary, Panalpina Airfreight AG, a single manager responsible for the leasing of transportation capacities in airplanes incorrectly estimated transportation volumes over a period of 14 months, which resulted in substantial losses. To cover up his own mistakes, the manager had been falsifying bookings since autumn 2004. Consequently, the reported revenues of the fiscal years 2004 and 2005 had to be restated by EUR 7m and EUR 14.1m, respectively (Manager Magazin, January 4, 2006). (Today, the company is part of the Danish transport and logistics company DSV Panalpina A/S.)

We therefore infer that the financial accounting misconduct spanned the years 2004 to 2005. Our sample only contains the years from 2005.

Patisserie Holdings PLC - GB00BM4NV504

Patisserie Holdings PLC was a UK-based cake and casual dining company. The company's collapse in January 2019 was caused by an alleged accounting fraud that was valued at GBP 94m in 2017. The fraud involved the overstatement of its cash positions by GBP 30m, failure to disclose overdrafts of nearly GBP 10m and the overstatement of asset value. Following the scandal, the company was taken over by Causeway Capital, an Irish private equity firm. UK's Serious Fraud Office arrested and questioned the company's former CEO and five other people (BBC, March 15, 2019).

Hence, we follow that the accounting misconduct affected the financial reports of year 2017.

Pescanova SA - ES0169350016

Pescanova SA was a Spanish fishing company based in Redondela, Galicia. In 2013, Spain's stock market regulator issued a statement saying that the submitted 2012 financial results did not comply with required accounting standards, possibly opening the door to sanctions. According to the ruling of the Audiencia Nacional (Spain's National Court), that also sentences the former president of the company to prison time, auditor BDO Auditores S.L., hired by the company since 2002, had issued fraudulent favourable reports on the annual and consolidated accounts of the company for the years 2010, 2011, and 2012, which was never published. In March 2013 an insolvency process started. Later, the company revealed that chairman Manuel Fernandez de Sousa had sold half of his 14.4 percent stake in the company between December 2012 and February 2013, shortly before starting work on the insolvency process. The insolvency administration revealed three times the reported debt, negative equity of almost EUR 1.0bn, and a real result of almost EUR 800m in losses. Finally, in 2020, the Audiencia Nacional sentenced former president Manuel Fernández de Sousa-Faro to eight years in jail for fraud, distortion of economic and financial information, and for falsifying commercial documents (Amat et al. (2013); Reuters, April 16, 2013; Salmon Business, October 7, 2020; El País, October 7, 2020).

We view this case to entail financial accounting misconduct in the period ranging from 2009 to 2011.

Phoenix IT Group - GB00B0315W65

Phoenix IT Group is a UK-incorporated manager of IT infrastructure facilities. Through investigation performed by auditor PwC and commercial law firm Nabarri LLP in 2012, the company had been found to have inflated its net assets by approximately GBP 14m on a post-tax basis. PwC also found that the profits of Servo, a subsidiary of the company had been overstated since March 2009 (ChannelBiz, November 29, 2012) (In July 2015, Daisy Group acquired the company.)

We therefore conclude that the financial accounting misconduct most likely spanned the years from 2009 to 2011.

Quindell Business Process Services Ltd / Watchstone Group PLC - GB00BYNBFN51

Quindell Business Process Services Ltd is an investment holding company and was listed on the London Stock Exchange. After a review by accounting firm PwC, the company made changes to its accounting policies of its professional services division in 2013 because PwC deemed that the previous accounting practices at the company were not appropriate. Using the old accounting policies, the company had overstated its 2013 pre-tax profits by GBP 44.9m and would have overstated its 2014 pre-tax profits by GBP 312m (Insurance Times, August 5, 2015). In 2015, UK's Serious Fraud Office launched an investigation into the company's business and financial statements from 2013 and 2014. In 2018, UK's Financial Reporting Council fined auditor KPMG with GBP 4.5m, as KPMG admitted misconduct when auditing the company's 2013 financial statements (Independent, June 11, 2018). (Quindell was later rebranded as Watchstone Group PLC.)

As such, we view this case to entail financial accounting misconduct in the year 2013.

Safran SA - FR0000073272

Safran SA, located in Paris, is the second-largest aircraft equipment manufacturer, and operates in the fields of rocket engine, aerospace, and defense. In 2006, during the audit committee meeting of the company's supervisory board, the executive board and auditors Constantin, Deloitte, and KPMG reported on unexplained accounting entries discovered in the files of the subsidiary Sagem Défense Sécurité. The amount likely to impact the financial statements of the subsidiary is estimated, at this stage, at EUR 100m. Reported in press releases of December 8 and 12, 2006, auditors KPMG and EY reported to the audit committee their conclusions on the accounting impact of the unexplained entries. The level of corrections arrived at during their investigations is €134.5 million, broken down as follows:
- Impact on 2005 shareholders' equity (before tax): -€106.6 million,
- Impact on 2005 operating income: -€25.8 million,
- Impact on 2006 operating income: - €2.1 million,
(Les Echos, December 11, 2006; Flight Global, December 20, 2006; Décision de la Commission des sanctions, October 28, 2010).

We view this case to entail financial accounting misconduct in the year 2005.

Speedy Hire PLC - GB0000163088

Speedy Hire PLC is a construction equipment rental company, incorporated in Liverpool, UK. In November 29, 2013, accounting issues had been discovered at the company's Middle East international division and forced the company's CEO Steve Corcoran to step down (FT, May 13, 2014; Speedy Hire PLC, November 29, 2013). On February 28, 2014, the company announced that a subsequent internal investigation determined that the misstatements totaled GBP 4.8m. The accounting misconduct comprised the overstatement of revenues and the understatement of costs. Net assets had been overstated by GBP 0.6m in fiscal year 2011 and GBP 2.1m in fiscal year 2012, both of which ended on March 31 of the next year (Constructionnews, November 28, 2013; Reuters, February 14, 2014).

We therefore conclude that the accounting misconduct spanned the fiscal years 2011 to 2012.

Steinhoff Investment Holdings Ltd - NL0011375019

Steinhoff Investment Holdings Ltd is an international retail holding company focused on furniture and household goods. The company is headquartered in Stellenbosch, South Africa, incorporated in The Netherlands, and publicly listed in Germany and South Africa. A small group of its executives and employees fabricated revenues and boosted asset values by systematically conducting fictitious transactions. According to a forensic investigation led by audit firm PwC in 2019, these transactions totaled EUR 6.5bn and affected the fiscal years 2009 to 2017. The subsidiaries Talgarth Group, Campion/Fulcrum Group, and Tulett Holdings are the entities of the company involved in the accounting misconduct. Most of the deceptive transactions conducted in these subsidiaries induced fictitious receivables or loans on the level of the group (PwC Forensic Investigation, March 15, 2019; Reuters, March 15, 2019; Zeit, March 4, 2021).

Based on the results of the forensic investigation of audit firm PwC, we conclude that the financial accounting misconduct likely spanned the years 2009 to 2017 (and may have become most serious in 2015 and 2016).

Telefonaktiebolaget LM Ericsson - SE0000108656

Telefonaktiebolaget LM Ericsson is a company that provides infrastructure, services and software to the telecommunication industry, and other sectors. In 2019, the company agreed to pay over USD 1bn to resolve the investigation of the U.S. Department of Justice into violations of the. The company was charged with conspiracies to violate the anti-bribery, books and records, and internal controls provisions of the Foreign Corrupt Practices Act. Ericsson has admitted that it has conducted bribery, falsifying of its books and records and failure to implement internal accounting controls in Djibouti, China, Vietnam, Indonesia, and Kuwait. Sham contracts, fake invoices, improper recording of payments were the main ways the company had falsified its books until 2016. In each country, the misrepresentation of books ranges from USD 450,000 to USD 45m (Department of Justice, December 6, 2019).

While it is unclear in which years the company definitely committed accounting misconduct, based on the findings of the U.S. Department of Justice that the company committed fraud in China from 2013 to 2016, and in Vietnam and Indonesia from 2012 to 2015, we conclude that the misconduct period most likely spanned the years 2012 to 2016.

Wirecard AG - DE0007472060

Wirecard AG was a German payment processor and financial services provider incorporated in Aschheim, Germany. The company's business model was to facilitate debit and credit card transactions between customers and retailers independent of sales channels and payment instruments by collecting shoppers' payment details. To settle payments outside the European Union, the company depended on a network of external acquiring partners. The revenue generated by its third-party acquiring partners was paid into Escrow accounts, managed, and overseen by trustees. Beginning in 2015, the company differentiated between receivables and liabilities from domestic and third-party acquiring business operations in its external reporting and most of the company's consolidated operating income was associated with these third-party acquiring business operations in the period from 2016 to 2018. On June 22, 2020, the management issued an ad hoc notification to its investors, confirming that the previously reported funds from its third-party acquiring business operations, totaling EUR 1.9bn, did not exist (ESMA, 2020). The first doubts about the company's accounting regarding the payables and receivables of its third-party acquiring business were raised by the Financial Times in July 2015 (Financial Times (Part 1), July 23, 2015; Financial Times (Part 2), July 23, 2015). According to a testimony of Oliver Bellenhaus, a former employee responsible for the company's Dubai subsidiary, the first corporate funds had been shifted out of the company into bank accounts in the name of shell companies in 2011 (Financial Times, August 8, 2021).

We thus conclude that the financial accounting misconduct likely spanned the years 2011 to 2019 and may have become most significant in the period from 2014 (i.e., one year before separately reporting the third-party business for the first time) to 2019). Our sample only contains the years until 2017.

Worldspreads Group PLC - IE00B2357Y89

Worldspreads Group PLC was a financial spread-betting firm, incorporated in Dublin, Ireland. The company was listed on the alternative investment market of the London Stock Exchange. In December 2009, the company sold Worldspreads (Ireland) Ltd, the Irish arm of the business, through a management buyout. The transaction was paid in two tranches in December 2010 and December 2011. Following the management buyout, a new company called Marketspreads was established out of Worldspreads (Ireland) Ltd. At the time of the buyout, retained earnings as well as receivables from clients of Worldspreads (Ireland) Ltd had been inflated by EUR 6.7m and EUR 3m, respectively. At the level of the group, the accounting misstatements at Worldspreads (Ireland) Ltd caused the restatement of fiscal year 2009 (ending on March 31, 2010) results. Other payables had been understated by EUR 610k, while gains of the disposal reported in the income statement had been overstated by EUR 610k. The company filed for administration in March 2012 (Worldspreads Group PLC, September 26, 2011; FT, March 19, 2012; FT, May 15, 2012).

We view this case to entail accounting misconduct in fiscal year 2009.

We further looked into several cases that could not be included in the final sample for different reasons.

- Cases of firms that are not listed in Compustat Global: Gowex SA - ES0158252017
- Cases of firms where Compustat Global does not provide essential data such as the daily file: OW Bunker A/S - DK0060548386, Hess AG - DE000A0N3EJ6
- Cases of firms where no fiscal year end results are affected by accounting misconduct: M&C Saatchi PLC - GB00B01F7T14, Tesco PLC - GB00BLGZ9862, EMI Group Limited - GB0000444736
- Cases of firms where – at this point – no definite misconduct is identified: Stora Enso Oyj - FI0009005961, Europacorp SA - ES0169350016, Thomas Cook PLC - GB00B1VYCH82, Panaxia Security AB - SE0001718396
- Cases of firms where the misconduct scheme was implemented after 2017: Grenke AG - DE000A161N30, Finablr PLC - GB00BJ7HMW26

Appendix C.5 - Support Vector Machine with Financial Kernel prediction results

	Performance metrics with cutoff at top 10%				
	Sensitivity	Precision	# Firms	NDCG@k	AUC
U.S. Training sample 1991-2003					
U.S. Test sample - 2005	14.28%	1.19%	6/42	0.077	0.553
European sample (2005-2017)					
Relative to U.S. Test sample	6.78%	0.70%	3/21	0.079	0.597
Relative to itself	13.56%	1.10%	6/21	0.079	0.597
Relative to itself per year	15.25%	1.23%	7/21	0.079	0.605
U.S. Training sample 1991-2008					
U.S. Test sample - 2010	30.43%	1.54%	7/23	0.145	0.708
European sample (2005-2017)					
Relative to U.S. Test sample	13.56%	1.56%	6/21	0.086	0.619
Relative to itself	15.25%	1.24%	6/21	0.086	0.619
Relative to itself per year	16.95%	1.37%	7/21	0.087	0.627

This table presents the classification performance evaluation for a Support Vector Machine model using a Financial Kernel with the main and the alternative training samples and the three alternative ways to rank the European percentiles of the distribution of the misconduct prediction. As this approach requires lagged variables, the total number of misconduct firms (firm-years) slightly changes compared with all other results. For instance, the total number of misconduct firms in the 2010 U.S. Test sample is 23 instead of 24.