UNIVERSITY OF FRIBOURG

Detecting Earnings Management

Doctoral Thesis
Presented to the Faculty of Management, Economics, and Social Sciences
at the University of Fribourg (Switzerland)

by
Patrick Chardonnens
from Belmont-Broye (FR), Switzerland,
in fulfillment of the requirements for the degree of
Doctor of Economics and Social Sciences (Dr. rer. pol.)

Accepted by the Faculty of Management, Economics and Social Sciences
on February 27, 2023 at the proposal of

Prof. Dr. Martin Wallmeier (First Advisor) and
Prof. Dr. Franck Missonier-Piera (Second Advisor)

Fribourg (Switzerland), 2023

https://doi.org/10.51363/unifr.eth.2022.003

"If we knew what it was we were doing, it would not be called research, would it?"

Albert Einstein

A hidden quote that made me laugh but does *not* apply to me :)

"The reason I talk to myself is because I'm the only one whose answers I accept."

George Carlin

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to Prof. Dr. Martin Wallmeier. The eye-to-eye exchanges, patience, and constantly open door created a very pleasant working environment. His helpful ideas and feedback, as well as the enriching discussions, helped me to get closer to what I would call a researcher. In my life so far, I have not met anyone more capable of understanding complex matters in such a short time. I would also like to thank Prof. Dr. Franck Missonier-Piera for his availability as a second supervisor.

Secondly, I would like to thank my family. Felizia, my wife, has been helpful and understanding in so many different ways. Be it with supportive words when I faced challenges or with her flexibility when it came to freeing up time. I am very grateful for everything she has done for me. I would also like to thank my parents and parents-in-law for their support during my academic journey, whether it was taking care of the kids or other things outside the academic world.

Finally, I would like to thank my colleagues in Fribourg. I really appreciated the exchanges during coffee breaks and lunches. Being in the same situation, sharing challenges, and learning from each other helped me to stay motivated. I am grateful for their advice and wish them all the best in their future journey.

ABSTRACT

This dissertation discusses the challenges of earnings management research and aims to improve the understanding through the following essays. Chapter 2 examines the earnings discontinuity literature, the methods used, and identifies the research questions for Chapters 3 and 4. For the earnings changes and earnings surprises thresholds, we highlight the different design choices in the studies.

Chapter 3 is a joint paper with Martin Wallmeier and Peter Fiechter published in Finance Research Letters. The dissonance in the U.S. zero earnings literature between Dechow et al. (2003), who find a decreasing discontinuity in the late 1990s, and Gilliam et al. (2015), who suggest that the Sarbanes-Oxley Act (SOX) is causally responsible for the disappearance of the kink, asks for further investigation. We identify the dotcom boom at the turn of the millennium as a possible confounding effect that may have led to mechanically more firms being in the small loss interval. After excluding firms with almost no sales, it is unclear whether the zero earnings kink has completely disappeared. More importantly, the decline in the zero earnings kink appears to be gradual rather than a sharp decline caused by SOX.

Chapter 4 is a joint paper with Martin Wallmeier that is under review at the time of writing. It examines the evolution of the zero earnings discontinuity in Europe and finds that, in contrast to the U.S., the kink has remained remarkably stable. At the country level, we find that firms in environments with higher uncertainty avoidance are associated with a higher zero earnings discontinuity. The cultural dimension seems to be more important than the adoption of the International Financial Reporting Standards (IFRS).

Chapter 5 evaluates three different firm-level earnings management models using the Accounting and Auditing Enforcement Releases (AAER) sample, which includes firms known to have manipulated earnings. The first model is the commonly used modified Jones model. The second model extends the first to include earnings information, and the third model is calculated using the earnings histogram. The results show that the discretionary accruals of firms committing fraud are significantly higher in the first and second models. These two models also lead to a slightly better than a random prediction of overstated firm-years. The third model does not show any superiority over random selection.

Chapter 6 predicts next year's accruals. We use a large dataset including 447 explanatory variables with the standard ordinary least squares (OLS) and three supervised machine learning models, the least absolute shrinkage and selection operator (LASSO), random forest, and support vector machines (SVMs). We find that the LASSO model has the highest out-of-sample prediction accuracy. In general, the machine learning models outperform the OLS model. However, even the most accurate models still contain a relatively high prediction error, showing that accruals remain challenging to predict.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AAER	Accounting and Auditing Enforcement Releases
CEO	Chief Executive Officer
CFO	Chief Financial Officer
CFRM	University of California-Berkeley Center for Financial Reporting and Management
CompAcc	Comprehensive Accruals
EPS	Earnings per Share
EU	European Union
FN	False Negatives
FP	False Positives
GAAP	Generally Accepted Accounting Principles
I/B/E/S	Institutional Broker's Estimate System
IFRS	International Financial Reporting Standards
IPO	Initial Public Offering
LASSO	Least Absolute Shrinkage and Selection Operator
MAE	Mean Absolute Error
MCap	Market Capitalization
MSE	Mean Squared Error
NYSE	New York Stock Exchange
OLS	Ordinary Least Squares
PPE	Property, Plant & Equipment

PPS	Price per Share
RSS	Residual sum of Squares
SEC	Securities and Exchange Commission
SIC	Standard Industrial Classification
SOX	Sarbanes-Oxley Act
SVMs	Support Vector Machines
TN	True Negatives
TP	True Positives
UAI	Uncertainty Avoidance Index
WCA	Working Capital Accruals
WLS	Weighted Least Squares
ZE	Zero Earnings

Chapter 1
INTRODUCTION

> "I also chose to fulminate against the earnings management literature, a body of work that in my view is scandalous in more ways than one." (Ball, 2013, p. 848)

This strong statement by Ray Ball appeared as a commentary in the journal *Accounting Horizons*. Ray Ball is no stranger to the accounting literature. He is a professor at the University of Chicago Booth School of Business and one of the most cited accounting researchers. More important than who he is, however, is why he made this statement.

One reason for this statement lies in a questionable implicit assumption that earnings management models make. They treat at the firm level any deviation from estimated accruals as discretionary. However, it is unclear what accruals should be in the absence of earnings management. Ball (2013, p. 851) rhetorically asks why, if researchers genuinely believe that manipulation occurs on such a large scale, they do not send the list of firms most engaged in earnings management to regulators or analysts. While it appears plausible to believe that not all firms manipulate their earnings, it is also undisputed that at least some firms engage in earnings management. According to the survey of Dichev et al. (2013), chief financial officers (CFOs) believe that nearly 20% of earnings reports are misrepresented.

Earnings management models still have difficulties reliably detecting firm-level earnings management, even though accounting data is readily available. The Refinitiv Datastream database contains over 85,000 live and dead firms and up to 8,000 data items for each firm and year (Refinitiv, 2022). The combination of both, the expectation that earnings management will occur and the sheer mass of available accounting and financial data, generates a great deal of interest because there is likely more to be found in the data. To this day, the research area of earnings management continues to exert an uninterrupted fascination on researchers, with over 2.7 million results on Google Scholar. Nevertheless, the understanding of earnings management measures and the earnings management process is not yet as good as it could be. Progress still needs to be made.

This thesis critically examines the literature on earnings management, identifies some weaknesses of previous studies, and attempts to improve the understanding of earnings management. The remainder of the introduction is organized as follows. First, we introduce the definition of earnings management. Then, we discuss the various earnings incentives and objectives, followed by how to manage earnings. Finally, we introduce the research ideas explored in the following chapters.

What is Earnings Management?

There is no uniform definition of earnings management. However, some are more popular than others. The most widely known definition stems from Healy and Wahlen (1999), who reviewed the existing literature on earnings management in the context of standard setters:

> "Earnings management occurs when managers use judgment in financial reporting and in structuring transactions to alter financial reports to either mislead some stakeholders about the underlying economic performance of the company or to influence contractual outcomes that depend on reported accounting numbers." (Healy & Wahlen, 1999, p. 368)

Healy and Wahlen (1999)'s definition is phrased in such a way that earnings management is understood to be negative and harmful. Subsequently, many other slightly different definitions of earnings management have been proposed, some of which add that earnings management does not need to be purely negative.

Ronen and Yaari (2008, p. 25) group the literature into beneficial, neutral, and pernicious earnings management. An example of beneficial earnings management would be managers who signal the firm's private information about future cash flows by using the flexibility of accounting standards because it increases transparency to external stakeholders. Beneish (2001, p. 3) sees earnings management as an opportunity for managers "to reveal to investors their private expectations about the firm's future cash flows." Neutral earnings management is a hybrid combination of beneficial and pernicious effects. In this sense, earnings management could either maximize the firm's value, which would be economically efficient, or serve only the manager's opportunism, depending on the situation (Watts & Zimmerman, 1990). Pernicious earnings management occurs when managers attempt to reduce the transparency of information in financial reports that would otherwise be available to the public. This view has been taken by former Securities and Exchange Commission (SEC) Chairman Arthur Levitt (1998), who criticized the "numbers game" of earnings management in his commentary in the CPA Journal. Schipper (1989, p. 92) understands earnings management as the managers' "intent of obtaining some private gain." In short, Ronen and Yaari (2008)'s differentiation is in the sense that earnings management is beneficial when it serves to release more useful information to external stakeholders. In contrast, it is harmful when earnings management serves to conceal information.

Incentives to Manage Earnings

During a firm's life, there are potentially many different situations in which earnings management could benefit the firm or, at least, its managers. Depending on the individual case, it may seem better to either beat a benchmark, smooth, or overstate earnings, and sometimes even understate earnings. Below we briefly explain and differentiate these earnings targets.

Beating benchmarks is one of the most apparent incentives for managers to engage in. Examples of benchmarks include releasing positive earnings, earnings exceeding those of the previous year, or exceeding analysts' earnings forecasts. The decision of whether to reach one of these benchmarks can be taken even after the end of the financial statement period. The survey of Graham et al. (2005, p. 21) provides compelling evidence that managers are aware of earnings benchmarks and may engage in earnings management to achieve them. They asked CFOs how important it is to meet quarterly earnings benchmarks. Sixty-five percent responded that reporting positive earnings is important or very important. More than 80% of CFOs also think that meeting earnings benchmarks would establish credibility in capital markets and maintain or increase the share price. They further believe that failure to meet benchmarks would foster uncertainty about the firm's prospects and hurt the stock price. Therefore, if unmanaged earnings are just below a benchmark, it might be tempting to engage in earnings management to reach the benchmark.

Earnings smoothing happens when firms reduce earnings in good years and increase them in bad years to minimize earnings volatility (Barth et al., 1999; Bartov, 1993; DeFond & Park, 1997; J. N. Myers et al., 2007; Payne & Robb, 2000). The survey of Graham et al. (2005, p. 44) reports that nearly 97% of CFOs prefer a smooth earnings path over a bumpy one. This preference for a smooth earnings path is because respondents believe that firms that report consistent earnings are "perceived as less risky by investors" and thus convey a reputation as a stable firm (Graham et al., 2005, p. 45).

Earnings maximization means that managers boost their earnings to a maximum extent. Unlike earnings thresholds or smoothing, earnings maximization incentives are often tied to a specific event. It is important to note that earnings management has to be reversed at some point, so managers are aware that earnings maximization "either depletes past reserves of reported outcome or borrows from future reports" (Ronen & Yaari, 2008, p. 342). The academic literature has identified numerous situations where firms might face incentives to maximize earnings. We briefly present an overview of the most important incentives in the following paragraphs.

An upcoming initial public offering (IPO) carries potentially strong incentives for managers to maximize earnings. Higher earnings before the IPO would lead to higher market value, benefiting the sellers of the shares. Consequently, high earnings management at the time of the IPO is associated with the relatively poor performance of earnings and market value in subsequent years (DuCharme et al., 2001; Teoh, Welch, & Wong, 1998).

Firms financing themselves through capital markets face the incentive to maximize prior earnings to increase the share price and thereby lower their relative financing costs. This has been observed for seasoned equity offerings where managers aim to collect more cash for the same number of shares (Cohen & Zarowin, 2010; S. C. Myers & Majluf, 1984;

Rangan, 1998; Shivakumar, 2000; Teoh, Welch, & Wong, 1998). It has also been shown that firms issuing debt increase their earnings to obtain lower borrowing costs through lower interest rates (Guay, 2008; Y. Liu et al., 2010).

Managers who receive stock-based compensation have been found to manage their earnings upward more than their peers without stock-based compensation (Cheng & Warfield, 2005). Chief executive officers (CEOs) under pressure to be dismissed tend to boost earnings in the short term (Guan et al., 2005).

Further, firms aiming to acquire another firm, on average, increase their earnings prior to the takeover, increasing their stock price and thus reducing the financing costs of the transaction (Erickson & Wang, 1999; Louis, 2004). Finally, firms at risk of becoming targets of a hostile takeover tend to manage earnings upward to increase the firm value, making the transaction more costly for the acquirer (Easterwood, 1998; Erickson & Wang, 1999; Guan et al., 2004).

Although less explored, incentives to reduce or minimize earnings are nonetheless important. In the case of unscheduled CEO changes, newly appointed CEOs tend to reduce their earnings through large write-offs to have reserves for subsequent years (Elliott & Shaw, 1988; Pourciau, 1993; Strong & Meyer, 1987; Wells, 2002). Managers seeking a buyout minimize earnings to lower the transaction price (DeAngelo, 1986; Perry & Williams, 1994). Firms seeking to be acquired in friendly takeovers tend to manage their earnings downward (Anagnostopoulou & Tsekrekos, 2015; Ben-Amar & Missonier-Piera, 2008; Missonier-Piera & Spadetti, 2022). A higher earnings reduction of the acquired firm is associated with a higher takeover premium (Missonier-Piera & Spadetti, 2022). In subsequent years, the firm could increase its earnings by reversing the earnings management (S. Chen et al., 2016).

Earnings minimization can also occur without a specific event. If pre-managed earnings are outstanding and key benchmarks are met, managers can build a “cookie jar” by overstating future costs such as warranties or sales returns. These reserves are reversed in less favorable years to increase reported earnings (Levitt, 1998). If pre-managed earnings, in contrast, are particularly poor, and it would be too costly to meet earnings benchmarks, it may be attractive to take a "big bath" to reduce earnings massively. The firm can then benefit in subsequent years by having more room to manage earnings upward (Cameron & Stephens, 1991; Dechow et al., 2003; Ewert & Wagenhofer, 2005; Kirschenheiter & Melumad, 2002; Strong & Meyer, 1987). Kirschenheiter and Melumad (2002) show that earnings smoothing toward the average and big bath accounting can coexist. Finally, when managers want to use insider trading, they reduce earnings to buy stocks at a low price. Once the negative earnings management is reversed, these shares are expected to be sold at a higher price (Benabou & Laroque, 1992).

How to Manage Earnings?

Here we briefly introduce the techniques to manage earnings towards the desired earnings target. Ronen and Yaari (2008) classify the techniques into accrual earnings management, real earnings management, and classification shifting. Accrual-based earnings management is, in simple terms, the management of non-cash items. For example, writing down an asset less than would be required to reflect the asset's true value results in higher earnings. It is considered the "easiest" way to boost earnings because managers can continue to manage accruals after the end of the accounting period (Zang, 2012). Earnings management through accruals is further discussed in Chapter 5.

The second approach, real earnings management, is used when managers take "real" business actions that affect the firm's cash flow. Examples include deferring or reducing discretionary spending on advertising, maintenance, or new hires. Real earnings management is often measured by operating cash flow, cost of sales, production, and discretionary expenditures (Roychowdhury, 2006, pp. 344–345). Earnings management through real activities must occur during the fiscal year and therefore requires more planning. The study by Graham et al. (2005, p. 40) interviewed managers about whether they would sacrifice long-term value in favor of short-term earnings goals. Of the 20 managers interviewed, 15 indicated that they would engage in real earnings management, even if it did limited harm to the long-term firm value. One strategy for managers who want to reach a certain earnings threshold might be to steer earnings towards the desired threshold during the year through real earnings management and then fine-tune through accrual-based earnings management (Graham et al., 2005; Zang, 2012).

In this thesis, we evaluate accrual-based but not real earnings management for the following reasons. We analyze the models and aim to improve their understanding. Real earnings management is considered more complex to detect than accrual-based earnings management (Kothari et al., 2016; Roychowdhury, 2006). Since there are already skeptical views about the "simple way" of measuring and detecting accrual-based earnings management, such as those expressed by Ball (2013), we refrain from dissecting the more complex real earnings management. Nonetheless, we acknowledge that conducting our studies with real earnings management might be interesting.

The third approach to managing earnings is classification shifting, although it is less explored in the literature. The approach is to shift expenses and special items down the income statement to appear after core expenses (Fan & Liu, 2017; Fan et al., 2010; McVay, 2006). As with accrual-based earnings management, classification shifting is usually executed after the fiscal year. It is important to note that the classification shift does not affect the bottom line but only the core earnings. In this context, Burgstahler and Chuk (2017, p. 731) evaluate the discontinuities in the zero earnings benchmark for earnings before and after special items. They find a large discontinuity only for earnings after special items but not for earnings before special items. It appears that earnings after special items are the earnings number that managers and external stakeholders implicitly agree with. Therefore, classification shifting is not further examined in the following chapters.

Structure of the Thesis

The structure of the thesis is as follows. Chapter 2 analyzes the literature on earnings benchmarks. We place particular emphasis on the measurement methods used and the drawbacks of the standardized differences test statistic. For the U.S. zero earnings discontinuity, we find that it is necessary to evaluate the kink over time to determine whether the Sarbanes-Oxley Act has indeed led to the disappearance of the zero earnings discontinuity, as suggested by Gilliam et al. (2015). This results in the research question answered in Chapter 3, a joint paper with Martin Wallmeier and Peter Fiechter published in Finance Research Letters (Chardonnens et al., 2022). The paper also examines the role of the dotcom boom at the turn of the millennium as a possible effect on the zero earnings kink after excluding firms with almost no revenues.

In Europe, the zero earnings discontinuity is less researched than in the U.S., but interesting nonetheless. The mandatory adoption of the International Financial Reporting Standards (IFRS) in 2005 in some European countries may have reduced the discontinuity. We identify the need to analyze the annual discontinuity of European countries around IFRS adoption.

Chapter 4, a joint paper with Martin Wallmeier, addresses this research question and is under review at the time of writing (Chardonnens & Wallmeier, 2022). It examines whether the zero earnings discontinuity in Europe has declined simultaneously with the U.S. zero earnings kink. Arguments to expect a decline in zero earnings discontinuity are that the mandatory adoption of the IFRS may have restricted loss avoidance (Ball, 2016). A spillover effect from the U.S. may also have contributed to a decline in discontinuity. In contrast, one argument for the persistence of the zero earnings discontinuity is that the principles-based IFRS offer more flexibility than the rules-based U.S. Generally Accepted Accounting Principles (GAAP) (Ball, 2016). Further arguments in favor of discontinuity persistence include reporting incentives and the institutional environment (Ball, 2016; Soderstrom & Sun, 2007). We conduct a country-level analysis of the cultural aspect of the Uncertainty Avoidance Index (UAI), originating from the framework of Hofstede et al. (2010). The UAI measures the extent to which members of a society are uncomfortable with uncertainty and ambiguity. Therefore, we expect managers in countries with a high UAI to place higher importance on reaching the zero earnings threshold than their counterparts in countries with lower uncertainty avoidance.

Chapter 5 establishes a link between the distributional earnings histograms and the earnings management models. The rationale behind this idea is that both approaches attempt to measure earnings management. However, they differ in their characteristics. Because the earnings distribution is at the aggregate level, it does not allow the attribution of earnings management at the firm level. For many research questions, however, precisely such firm-level indicators are needed.

Commonly used firm-level measures stem from Jones-type earnings management models that estimate "normal" accruals using a regression approach based on firm-specific financial statement items that are assumed to be unrelated to earnings management. The unexplained portion of total accruals is considered discretionary and interpreted as earnings management, which led to criticism in the literature. We present and discuss this criticism of the widely used modified Jones model. In addition, we use an extension of the modified Jones model in which we add earnings information to account for the undesirable correlation between earnings and discretionary accruals. Finally, we develop a third model derived from the earnings distribution. We estimate an expected earnings distribution in the absence of earnings management and compare it to the observed earnings distribution.

A fundamental difficulty with all earnings management models is that their reliability and validity are unknown. It remains unclear whether the allocated portion of earnings management truly arises from earnings management or, rather, due to a misspecification of the earnings management model. Against this background, we examine the validity of the three firm-level measures above using firms known to have manipulated earnings because they have been the subject of the Accounting and Auditing Enforcement Releases (AAER) by the U.S. SEC.

Chapter 6 aims to improve the understanding of accruals by predicting accruals for the following year. In general, the explanatory power of regressions on accruals is relatively poor, resulting in high levels of allocated discretionary accruals. Previous explanations for the poor performance have been related to correlated omitted variables and the non-linearity of the accruals.

We address these criticisms by including an extensive set of explanatory variables from the Refinitiv Worldscope database to predict next year's accruals using linear models. We use the standard ordinary least squares (OLS) and three supervised machine learning models: the least absolute shrinkage and selection operator (LASSO), random forests, and support vector machines (SVMs). We compare the predictive accuracy of these measurement methods and expect to find relevant variables with predictive information for next year's accruals. Potentially, this could improve our understanding of the accrual formation process. Identifying key indicators of future accruals is essential not only for academia but also for the public. For stakeholders outside academia, such as investors and financial analysts, predicting the level of a firm's accruals may be helpful in light of the upcoming release of financial statements.

Chapter 2
THE ZERO-EARNINGS DISCONTINUITY: LITERATURE REVIEW

2.1 Introduction

The distribution of earnings frequencies scaled by the lagged market value of equity reveals a discontinuity at zero. This discontinuity arises because more firms report a small profit than a small loss, known as the zero earnings discontinuity or kink. Numerous researchers confirm this discontinuity.[1] Others conclude that the discontinuity is due to factors such as scaling issues, tax effects, and sample selection (Beaver et al., 2007; Durtschi & Easton, 2005, 2009). More recent studies report that the zero earnings discontinuity disappeared in the U.S. after 2002 (Gilliam et al., 2015; Lahr, 2014; Makarem et al., 2018).

The fragmented literature on this topic lacks comparability and conciseness. Previous studies use different income and scaling variables, possibly resulting in contradictory statements. Some studies analyze only one threshold: zero earnings, earnings changes, or earnings surprises. More importantly, few studies consider the change in discontinuity over time. Most studies focus on U.S. firms, whereas few report results for European firms or at the international level. The main objective of this paper is to provide an overview of the discontinuity literature for U.S. and European firms. We highlight the differences and contradictory findings from which we derive several research questions.

In the first section of this chapter, we introduce the zero earnings discontinuity and dissect the prevailing measure for quantifying discontinuities: the standardized difference test statistic by Burgstahler and Dichev (1997). In addition to previous research, this paper argues for evaluating the discontinuity over time, allowing to assess whether the Sarbanes-Oxley Act (SOX) effectively led to the disappearance of the zero earnings discontinuity, as suggested by Gilliam et al. (2015), or whether the thesis of earnings management to reach the zero earnings threshold is still the most complete explanation (Burgstahler & Chuk, 2017). Our research question examined in Chapter 3, a joint work with Martin Wallmeier and Peter Fiechter, is, therefore, whether SOX indeed led to the disappearance of the discontinuity or whether it could be explained by alternative factors.

[1] See Burgstahler and Chuk (2017), Burgstahler and Dichev (1997), Burgstahler and Eames (2006), Daske et al. (2006), Dechow et al. (2003), Degeorge et al. (1999), and Gore et al. (2007)

The discontinuity in Europe is less researched but nonetheless interesting. The International Financial Reporting Standards (IFRS) became mandatory in some European countries in 2005. The second section of this chapter examines whether a similar effect of the disappearance of the kink in Europe occurs, as Gilliam et al. (2015) observed with the introduction of the U.S. SOX. The research question discussed in Chapter 4, a joint work with Martin Wallmeier, is, therefore, whether there has been a change in zero earnings discontinuity after the introduction of IFRS.

The third section of this chapter discusses the thresholds for earnings changes and earnings surprises. These alternative thresholds differ from zero earnings in the sense that the earnings target is different for each firm. We highlight that the interpretation for both thresholds is less intuitive than for the zero earnings threshold because it is unclear how the histogram would look in the absence of earnings management. Regarding the threshold for earnings surprises, we also discuss previous studies' different design choices and the associated drawbacks.

2.2 Zero Earnings Discontinuity in the U.S.

This section provides an overview of the essential discontinuity studies. It starts by illustrating the zero earnings discontinuity. Different academic studies reach different conclusions, with some studies arguing that the zero earnings discontinuity would arise from effects other than earnings management. We review the literature, highlight the essence of the main differences, and summarize the various arguments for and against the thesis that earnings management is the reason for the discontinuity.

The term "discontinuity" or "kink" refers to a histogram in which the number of observations in one interval is visually different from the adjacent intervals. The term "benchmark" is used when firms are incentivized to meet or beat a certain earnings target. Prior literature uses the term "small profit firms" or "firms in the interval $i+1$" to describe firms whose earnings are just beating the threshold. "Small loss firms" or "firms in the interval $i-1$" are located to the left of the benchmark they just missed.

2.2.1 Earnings Histograms

Hayn (1995) groups earnings into intervals and plots them in a histogram. She observes that more firms report a small profit than a small loss. At zero, there is a large gap in frequency referred to as discontinuity, kink, or zero earnings benchmark. The most commonly known figure in the field of earnings management discontinuity analysis stems from Burgstahler and Dichev (1997, Figure 3, p. 109). They scale the earnings of each firm-year by the lagged market value of equity and group them in intervals of 0.5%. Figure 2.1 presents this zero earnings discontinuity. While the number of firms reporting a small profit appears unnaturally high, the number of firms reporting a small loss appears unnaturally low.

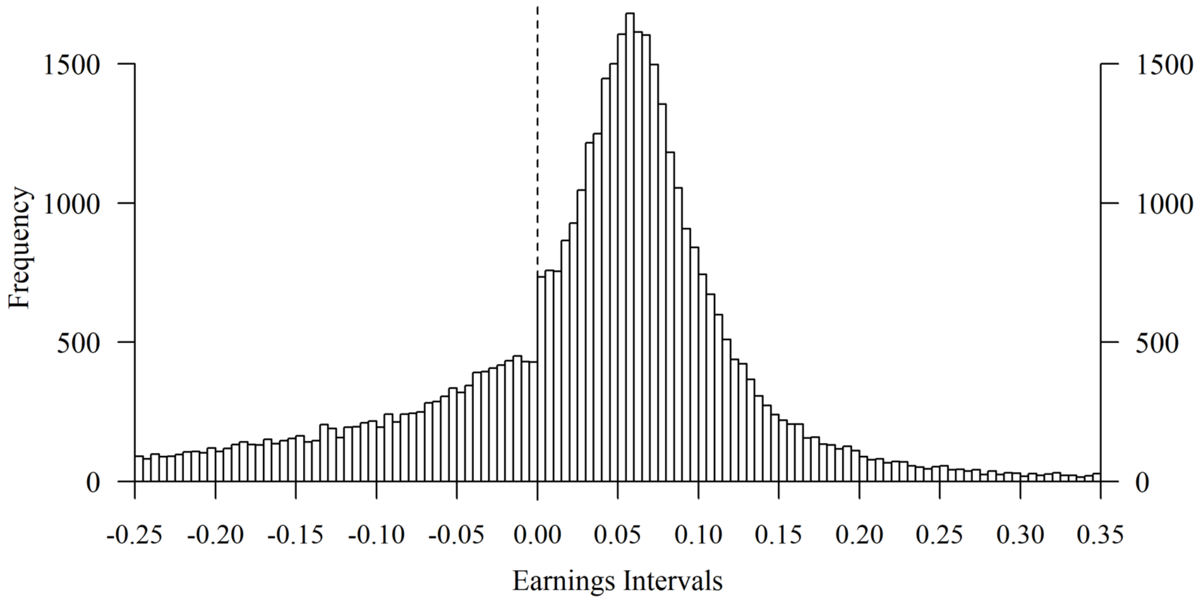

Figure 2.1: Zero Earnings Discontinuity. The figure shows the zero earnings discontinuity, similar to the discontinuity of Burgstahler and Dichev (1997, Figure 3, p. 109). At zero earnings, a relatively large discontinuity appears with more firms reporting small profits than small losses. To plot this figure, we used stock quoted U.S. firms from 1987 to 2001 from the Refinitiv Worldscope universe and scaled the Bottom Line Net Income (WC #1651) by the lagged Market Capitalization (WC #8001) of each year. The size of an earnings interval is 0.005.

In the following, many articles have been written on the zero earnings discontinuity. The various studies differ in data selection and design choices. This discrepancy could potentially lead to different magnitudes of the discontinuity. For this reason, we briefly present the main choices for the firm selection, the earnings, and the scaling variable.

Most studies exclude firm-years in regulated industries with Standard Industrial Classification (SIC) codes from 4400 to 5000 and 6000 to 6500.[2] Some studies make different design choices, but regardless of the industry selection criteria, they tend to observe a discontinuity.[3]

The choice of the earnings variable can have a meaningful impact on the magnitude of the discontinuity. Most studies, but not all, use the bottom-line net income as the earnings variable.[4] Burgstahler and Chuk (2017, p. 730) assess the discontinuities of various accounting income items. They report that "removing virtually any component of earnings" clearly diminishes the discontinuity if the item is not crucial to stakeholders. Therefore, it would be essential to use the most widely regarded earnings variable for the different stakeholders of the firm.[5]

A further important research design choice that affects the discontinuity is whether to remove firms from the sample that reported exactly zero earnings. It is unclear whether exactly zero earnings are due to database errors or whether these firms actually reported a net income of exactly zero. Most studies exclude these observations in order to take a conservative approach that eliminates possible database errors, although this may exclude firms that truly reported zero earnings.[6]

[2] See Beaver et al. (2007), Brown and Caylor (2005), Burgstahler and Dichev (1997), S. K. Chen et al. (2010), Durtschi and Easton (2005, 2009), Gilliam et al. (2015), Haga et al. (2017), Kerstein and Rai (2007), Makarem et al. (2018), and Roychowdhury (2006).

[3] For example, Dechow et al. (2003) and Ipino and Parbonetti (2017) exclude financial firms with SIC codes from 6000 to 6999. Others do not set any exclusion criteria (e.g., Jacob & Jorgensen, 2007).

[4] See Beaver et al. (2007), Brown and Caylor (2005), Burgstahler and Chuk (2017), S. K. Chen et al. (2010), Daske et al. (2006), Dechow et al. (2003), Donelson et al. (2013), Durtschi and Easton (2005, 2009), Enomoto and Yamaguchi (2017), Gilliam et al. (2015), Jacob and Jorgensen (2007), Kerstein and Rai (2007), Leuz et al. (2003), and Yu (2014). Others use either net income before extraordinary items (Cohen et al., 2008; Gore et al., 2007; Roychowdhury, 2006) or earnings per share (EPS) (Degeorge et al., 1999; Hayn, 1995).

[5] In unreported results, we evaluated the magnitudes of the main earnings variable, Bottom Line Net Income (WC #1651), including "income after all operating and non-operating income and expense, reserves, income taxes, minority interest and extraordinary items" (Thomson Financial, 2007, p. 374), and alternative income variables such as Net Income before Extraordinary Items / Preferred Dividends (WC #1551) and Net Income Used to Calculate Earnings per Share (formerly Net Income Available To Common) (WC #1751) (Thomson Financial, 2007, p. 375). Consistent with the results of Burgstahler and Chuk (2017), we observe the largest discontinuity for the bottom-line net income and smaller discontinuities for alternative income variables.

[6] See Beaver et al. (2007), Burgstahler and Dichev (1997), Burgstahler et al. (2006), Dechow et al. (2003), Gilliam et al. (2015), and Lahr (2014).

The choice of the scaling variable is vital because different scalars may lead to different magnitudes of the kink. The most commonly used deflator is lagged market value of equity.[7] The downside of this scalar is that it could push firms with comparatively high growth opportunities and thus high valuations towards the intervals near zero because of its high market valuation and not because of its financials. To account for this drawback, some researchers use lagged total assets either as a primary or alternative scaling variable.[8] However, total assets as a deflator also entails disadvantages. Some asset-intensive industries naturally have a low return on assets, while others tend to have high rates of return on their assets. The asset-intensive industries tend to be gathered in intervals close to zero. In Daske et al. (2006, Figure 1, p. 148), for example, the small loss interval is the interval with the most observations. However, if earnings are scaled by the market value of equity, the interval with the most observations is around 5% of scaled earnings. It remains unclear whether the high concentration of firms around the zero earnings threshold is due to industry effects or effectively due to earnings management to avoid losses.

Another commonly used scalar for robustness tests is the lagged book value of equity. We see several advantages of this scalar. First, it does not push highly valued firms toward the zero threshold since expected future earnings do not affect the book value of equity. Second, the mode peaks at higher profitability levels. This allows a more precise distinction between the discontinuity at zero and the general earnings distribution. A disadvantage of this scalar is its infrequent use in previous studies. In addition, a firm's stakeholders may influence the firm's book equity. Firms may opt to keep their book value of equity relatively low. The higher financial leverage leads to higher profitability and thus a higher return on book equity.[9] Since each scaling variable has its drawbacks,

[7] See Beaver et al. (2007), Burgstahler and Chuk (2017), Burgstahler and Dichev (1997), S. K. Chen et al. (2010), Daske et al. (2006), Dechow et al. (2003), Donelson et al. (2013), Durtschi and Easton (2005, 2009), Gilliam et al. (2015), Jacob and Jorgensen (2007), Lahr (2014), and Yu (2014).

[8] See Burgstahler et al. (2006), Cohen et al. (2008), Daske et al. (2006), Enomoto and Yamaguchi (2017), Gore et al. (2007), Leuz et al. (2003), Makarem et al. (2018), and Roychowdhury (2006).

[9] Total sales has been used by Beaver et al. (2007, p. 546), who report in the interval $i + 1$ slightly higher discontinuities than for other scalars. Finally, some researchers report unscaled EPS (Degeorge et al., 1999; Durtschi & Easton, 2005). The problem with unscaled EPS is that a firm's EPS highly depends on the number of shares outstanding and thus on the share price.

most studies present their results with more than one scaling variable to show that the results are robust to scaling effects.

After identifying and discussing the most common research design choices, we follow most previous research in calculating zero earnings and use each firm's bottom-line net income divided by their lagged market value of equity:

$$Zero\ earnings\ measure_{i,t} = \frac{Net\ Income_{i,t}}{Market\ Capitalization_{i,t-1}}. \tag{2.1}$$

The final decision is the interval width grouping the net income in the earnings histograms. Most prior literature opts for an interval width of 0.005 of scaled earnings.[10]

2.2.2 Standardized Difference Test Statistic

Following the data selection above, the earnings distribution mainly results in a discontinuity at the zero earnings threshold. However, the visual effect alone does not provide a quantifying measure. Burgstahler and Dichev (1997) develop a significance test that allows measuring whether unexpectedly many observations occur in the interval $i+1$ and unexpectedly few observations occur in the interval $i-1$. The standardized differences test statistic (hereafter standardized difference) divides the difference in observations by the square root of the observed and expected observations variance.

The difference in observations is the actual number of observations in interval i minus the number of expected observations calculated from the neighboring intervals immediately to the left and right of interval i. Thus, a positive (negative) difference in observations indicates that there are more (fewer) firm-years in that interval than would be expected from the average of the neighboring intervals, as shown in the following equation:

$$Difference\ of\ observations_i = N_i - \frac{N_{i-1} + N_{i+1}}{2}. \tag{2.2}$$

[10] See Beaver et al. (2007), Burgstahler and Chuk (2017), Burgstahler and Dichev (1997), S. K. Chen et al. (2010), Cohen et al. (2008), Daske et al. (2006), Dechow et al. (2003), Donelson et al. (2013), Durtschi and Easton (2005, 2009), Enomoto and Yamaguchi (2017), Gilliam et al. (2015), Jacob and Jorgensen (2007), Kerstein and Rai (2007), Makarem et al. (2018), Roychowdhury (2006), and Yu (2014).

Figure 2.2 illustrates the calculation of the difference of observations. The estimated number of observations in interval $i-1$ is the average of the neighboring intervals $i-2$ and $i+1$. The black boxes illustrate the observed number of firms in each interval. The green line indicates the average of the two neighboring intervals, which is the expected number of observations. The red box shows the difference between the observed and expected number of observations. In this example, there are fewer firms than expected in the interval $i-1$, resulting in a negative difference of observations.

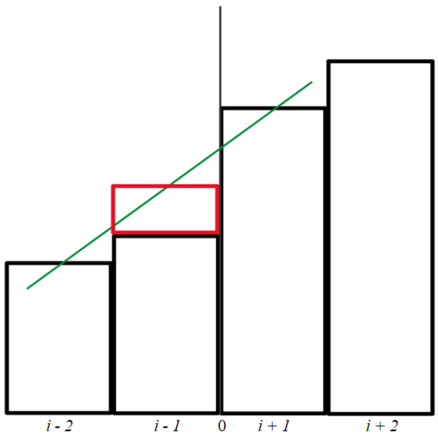

Figure 2.2: Test Statistic Visualization. The figure illustrates the calculation of the difference of observations. The estimated number of observations in interval $i-1$ is the average of the neighboring intervals $i-2$ and $i+1$. The black boxes illustrate the observed number of firms in each interval. The green line indicates the average of the two adjacent intervals, which is the expected number of observations. The red box shows the difference between the observed and expected number of observations. In this example, there are fewer firms than expected in interval $i-1$, resulting in a negative difference of observations.

The variance of the observed and expected observations (hereafter variance) scales the difference of the observations, as presented in the following formula:

$$Variance_i = N\ p_i\ (1-p_i) + 0.25\ N\ (p_{i-1}+p_{i+1})\ (1-p_{i-1}-p_{i+1}), \tag{2.3}$$

where N denotes the total number of observations and p denotes the probability that an observation falls within the interval i. Finally, the standardized difference is the difference of the observations in interval i divided by the square root of the variance of the same interval:

$$Standardized\ Difference_i = \frac{Difference\ of\ observations_i}{\sqrt{Variance_i}}. \tag{2.4}$$

Surprisingly, the standardized difference measure is exposed to little criticism. We briefly present three notable weaknesses of this measure. First, the measure of Burgstahler and Dichev (1997, p. 103) assumes that if firms do not manage earnings, the "standardized differences will be distributed approximately Normal with mean 0 and standard deviation 1." However, this ignores that some firms may be in the small profit or loss intervals because they are extremely highly valued in the capital markets. Since earnings are scaled by the lagged market value of equity, a comparatively small firm with low earnings but a high valuation could be pushed into the small profit interval that could accentuate the zero earnings discontinuity.

Second, Burgstahler and Dichev (1997, p. 103) focus on the standardized difference of the small loss interval and report the standardized difference of the small profit interval in parentheses. In their scenario, where a firm manages its earnings upward from the interval $i-1$ to $i+1$, both intervals would be affected. To account for this effect, Burgstahler and Dichev (1997, p. 103) are "arbitrarily selecting the standardized difference left of zero, and report the corresponding standardized difference right of zero in parentheses." However, an unbiased measure would require that the adjacent intervals are unaffected by earnings management. In the Burgstahler and Dichev (1997) scenario, the interval $i+1$ is inflated by firms managing their earnings. Averaging the intervals $i-2$ and $i+1$ leads to a spuriously inflated number of expected firms and standardized difference. Moreover, it is important to note that firms can manage their earnings upward not only from the interval $i-1$ to $i+1$. Firms can also manage their earnings downward, as observed by Mindak et al. (2016, p. 216), who found that more than 14% of firms manage their earnings downward to meet or exceed their earnings target. In the case where firms steer their earnings downwards into the interval $i+1$, the standardized difference of the interval $i-1$ would be affected, even though no earnings management occurred in this interval.

Finally, the standardized difference measures the significance level of a discontinuity. However, the significance rises as the number of observations increases. The reason is that the square root disproportionately reduces the variance of the deflator. Therefore, a larger number of observations with the same discontinuity magnitude leads to a higher standardized difference. We illustrate this with an example. Consider a sample of 2,500 observations where the interval $i-2$ contains 30 observations, the interval $i-1$ contains 25 observations, and the interval $i+1$ contains 50 observations. This results in a difference in observations of -15, a variance of 44.1, and thus a standardized difference of -2.26. Simply doubling all numbers in the above example results in a standardized difference of -3.19. Therefore, one should be careful when comparing the standardized difference of samples of different sizes. In Chapter 3, a joint work with Martin Wallmeier and Peter Fiechter, and Chapter 4, a joint work with Martin Wallmeier, we develop and use a discontinuity measure independent of the number of observations.

2.2.3 Profit-Loss Ratio

Dechow et al. (2003, p. 382) assess the discontinuity and, in particular, attempt to answer whether the kink is solely due to earnings management or whether there could be alternative explanations. They developed an alternative way to the standardized difference to measure the zero earnings discontinuity that has since been used in other studies (Burgstahler et al., 2006; Durtschi & Easton, 2005, 2009; Leuz et al., 2003). The measure consists of the number of observations in the small profit interval $i+1$ divided by the number of observations in the small loss interval $i-1$, as shown in the following equation:

$$Profit - Loss\ ratio = \frac{N_{i+1}}{N_{i-1}}. \tag{2.5}$$

A value of one indicates there are as many firms in the interval to the left of zero as in the interval to the right of zero. The simplicity is a notable advantage, as it combines both intervals in one metric. However, the disadvantage of the profit-loss ratio is that it can result in high values when only a few firms report a small loss. Averaging these values or using annual profit-loss ratios in a regression could lead to high values having a disproportionate impact. One solution would be to use the natural logarithm of the profit-loss ratio. Another drawback is that the ratio does not allow for significance testing.

In their study, Dechow et al. (2003, p. 380) report annual profit-loss ratios from 1989 to 2001 that decrease with fluctuations from 4.26 in 1991 to 1.31 in 2000, indicating that in 1991 more than four times as many firms are in the small profit interval as in the small loss interval. They run a regression to test if a time trend emerges. The intercept of 3.25 in 1991 decreases at a significant annual rate of -0.14, indicating a decreasing zero earnings discontinuity. They provide a possible explanation for this decline, which may be related to the exchange's listing requirements. In the years before 1995, the New York Stock Exchange (NYSE) placed great emphasis on positive earnings as a requirement for listing firms. After 1995, the requirements changed to a broader range of financials, such as market capitalization, operating cash flow, and revenue. In 1999, the listing criteria were reduced to revenue and market capitalization.

2.2.4 Doubts Whether the Kink Arises From Earnings Management

Not all studies in the academic literature agree that the zero earnings discontinuity is due to managers actively engaging in upwards earnings management to avoid reporting losses. Dechow et al. (2003) discuss several alternative factors potentially contributing to the zero earnings kink. These factors are (i) that instead of accruals accounting, managers could take real actions to reach the desired threshold, which could mean, for example, that employees work harder; (ii) exchange listing requirements and (iii) scaling effects, as discussed above; (iv) accounting rules that require losses to be recognized more timely than gains as described by Basu (1997); and (v) financial assets in the sense that they "earn dividends or interest, neither of which can be negative" (Dechow et al., 2003, p. 379). These purely positive returns could be responsible for the discontinuity.

A fraction of the academic literature even doubts that the zero earnings discontinuity truly arises from earnings management. In particular, alternative explanations are presented by Beaver et al. (2007) and Durtschi and Easton (2005, 2009). In the following, we discuss these alternative explanations and the corresponding counterarguments.

The studies of Durtschi and Easton (2005, 2009) contain three arguments why the kink might arise from sources other than active loss avoidance by managers. First, Durtschi and Easton (2005) question whether it is appropriate to scale earnings. They show that deflating earnings with the lagged market value of equity accentuates the earnings distribution, whereas no discontinuity occurs with unscaled earnings. Previous studies using scaled earnings would implicitly assume that scaling does not distort the distribution (Durtschi & Easton, 2005, p. 559).

Burgstahler and Chuk (2015, p. 171) respond to the argument of Durtschi and Easton (2005) by showing that unscaled earnings lead to an overrepresentation of small firms around zero. Moreover, unscaled earnings would be a poor measure because large firms can manage large amounts of earnings more easily than small firms. Therefore, scaling the earnings of individual firms based on their lagged market value of equity would produce more comparable results and be superior to unscaled earnings. Finally, grouping firms into quartiles by size would still lead to significant discontinuities in unscaled earnings and EPS (Burgstahler & Chuk, 2015).

The second argument of Durtschi and Easton (2005, 2009) against earnings management as an explanation for the zero earnings kink is similar to the concerns raised by Dechow et al. (2003, p. 358). They find that the characteristics of small profit firms differ from the rest of the sample in terms of earnings, age, and leverage. Durtschi and Easton (2005) scale reported EPS by the lagged share price and find that the lagged share price is higher for firms in the interval $i+1$ than for firms in the interval $i-1$. This effect would pull profitable firms into the small profit interval. Moreover, since the lagged share price is lower for firms with small losses, it would push them out of the interval $i-1$ into more negative intervals, accentuating the discontinuity (Durtschi & Easton, 2005, p. 573).

Burgstahler and Chuk (2015, p. 181) confirm the effect of a higher lagged share price for firms with small profits but put it into perspective. If firms are grouped into quartiles according to the lagged share price, the effect is observed only for the quartile with the lowest share price. However, if this quartile is omitted, the overall discontinuity would be only slightly reduced.

Finally, Durtschi and Easton (2005, 2009) argue that the discontinuity might arise from sample selection. Scaling earnings by the lagged market value of equity requires data from the previous year. They support their argument by showing that firms with a missing lagged market value of equity are proportionally more likely to report losses in the Compustat database. Including these observations with a missing lagged market value of equity would reduce the discontinuity at zero.

Burgstahler and Chuk (2015) react to this argument by scaling all firms by sales. They find that firms with missing share prices result in a lower discontinuity, but including all firms, with and without share price, still results in a highly significant discontinuity. Moreover, Beaver et al. (2007) point out that the inclusion or exclusion of firms with a missing lagged market value of equity should depend on the researcher's focus. If one wants to assess publicly traded firms, then firms with missing share prices should be excluded.

Another attempt to explain the zero earnings discontinuity other than through earnings management originates from Beaver et al. (2007). They mention that the tax rate reduces the profits of firms reporting positive earnings. This would pull profitable firms toward zero. In addition, negative special items are high for firms reporting losses, pulling these firms away from the zero threshold. The combination of these two effects partially explains the zero earnings discontinuity.

Burgstahler and Chuk (2017, p. 737) mention two limitations of the tax explanation. First, the argument of Beaver et al. (2007) would only be valid if the tax rate above the zero threshold is higher than the tax benefits below the zero threshold. Burgstahler and Chuk (2017) agree that this argument is potentially valid because the average tax rate per interval is effectively higher for firms reporting gains. However, this would only apply to the zero earnings benchmark but not to other thresholds such as earnings changes and earnings surprises. Second, Burgstahler and Chuk (2017, p. 737) point out that the effective tax rate cannot be regarded as a function of pre-tax earnings. In addition to actual pre-tax earnings, the effective tax rate also depends on expected creation and book-tax differences. Burgstahler and Chuk (2017) segment the sample into four subsamples with effective tax rates of up to 15 percent, above 15 to 30 percent, above 30 to 45 percent, and above 45 percent. They find a significant discontinuity for each subsample for firms with a positive tax rate, leading them to state that the tax rate explanation cannot explain the zero earnings discontinuity. Burgstahler and Chuk (2017, p. 726) conclude that "the theory that earnings are managed to meet benchmarks provides the simplest and most complete explanation".

There is evidence from surveys in which chief financial officers (CFOs) were questioned about their earnings management practices (Dichev et al., 2013; Graham et al., 2005). These questionnaires provide compelling evidence that at least some managers engage in earnings management to reach the zero earnings threshold. Graham et al. (2005, p. 21) asked CFOs how important it is to reach quarterly earnings thresholds. Sixty-five percent responded that reporting positive earnings it is important or very important. CFOs believe that if they fail to meet the thresholds, the share price will be affected. Further, over 80% of CFOs agree that meeting earnings benchmarks would establish credibility with the capital markets and maintain or increase the stock price. Failure to meet the benchmarks would foster uncertainty about the firm's prospects. Dichev et al. (2013, p. 24) surveyed CFOs on their general earnings management behavior and how prevalent they believe earnings management is. CFOs responded that they expect nearly 20% of all earnings releases to contain misrepresented earnings. 99.4% of CFOs believe that some reported earnings are managed.

Indeed, it is difficult, if not impossible, to distinguish which part of the zero earnings discontinuity is due exclusively to earnings management and which part is due to alternative effects. There may be factors accentuating the zero earnings discontinuity arising for reasons other than avoiding reporting losses. Nonetheless, most studies conclude that at least part of the zero earnings discontinuity is due to firms managing earnings upwards to avoid reporting a loss.

2.2.5 The Effect of the Sarbanes-Oxley Act

In the late 1990s and early 2000s, several firms, including Enron and WorldCom, misstated their earnings, leading to the period being referred to as the "scandal period". U.S. politicians and regulators recognized the need for public company accounting reform aimed at protecting investors. The result of this was the SOX. Key elements of the Act include increased penalties for fraudulent actions and additional responsibilities for corporate management. It was enacted on July 30, 2002, by former President George W. Bush, who stated that the act contains "the most far-reaching reforms of American business practices since the time of Franklin Delano Roosevelt" (The White House, 2004).

It should be noted, however, that only some sections of the Act became effective immediately in 2002. Section 404 demands an audit of internal control reporting and disclosure requirements. This section is considered the most consequential and controversial, not least because of its high cost. Section 404 became effective as Auditing Standard No. 2 for firms whose fiscal year ended after November 15, 2004, and whose market capitalization exceeded 75 million USD (Iliev, 2010).

In subsequent years, academic literature has examined whether the SOX benefits various stakeholder groups, but the findings do not lead to a consensus. In SOX's favor, fewer fraud cases were recorded after the introduction of the SOX (Dyck et al., 2010). In addition, firms with good internal controls have lower borrowing costs than before the introduction of the SOX (Ashbaugh-Skaife et al., 2009). Finally, the regulation would be beneficial for the market value of larger and more mature firms (Wintoki, 2007).

Criticism of the SOX is primarily concerned with the additional costs. There is no dispute that costs to firms are increasing due to SOX, even though they decrease over time. Coates IV (2007) reports that the costs arising from the SOX are a relatively higher burden for small firms. The standardization of the accounting regulation would be disproportionally harmful to young and high-growth firms (Wintoki, 2007). Smaller firms have been observed to go private to avoid the increased compliance costs of Securities and Exchange Commission (SEC) regulations (Engel et al., 2007; Kamar & Talley, 2008; Leuz et al., 2008). Alternative ways to circumvent the regulations included SEC withdrawal of registration, M&A activities, or liquidation.

Some studies use Jones-type earnings management models to assess the impact of the SOX on earnings management. Cohen et al. (2008) report an increase in accruals prior to the introduction of the SOX and a reversal afterward. In addition, some studies report a shift from accrual-based to real earnings management (Bartov & Cohen, 2009; Cohen et al., 2008; Ipino & Parbonetti, 2017).

Fewer studies assess the effect of the SOX on earnings thresholds. Gilliam et al. (2015) evaluate whether the introduction of the SOX may have affected the zero earnings discontinuity. They group their sample into a pre- and post-SOX period and find a strong discontinuity in the pre- SOX period but not thereafter, concluding the decline may be causally related to the mandatory application of the SOX starting in 2003.

They also test for alternative explanations. Gilliam et al. (2015, p. 125) address the critics of Durtschi and Easton (2005), who argue that the discontinuity could arise due to scaling. They form quartiles based on price per share (PPS) and market value of equity. For both scalings, there is strong evidence of a discontinuity in the pre-SOX era and close to no discontinuity in the subsequent period. Also, scaling by total assets leads to the same findings. In addition, Gilliam et al. (2015, p. 128) evaluate the tax explanation of Beaver et al. (2007) and its effect on the disappearance of the discontinuity. They conclude that taxes and special items strengthen the pre-2002 discontinuity but not the post-SOX period.

Further, Gilliam et al. (2015) report the yearly zero earnings standardized differences test statistics. Prior to the introduction of the SOX, almost all small loss standardized differences are significant. In the years following the SOX introduction, Gilliam et al. (2015) no longer find significant values. However, it is important to note that the annual standardized differences are affected by the number of observations in each year. The sample of Gilliam et al. (2015, p. 124) contains about 3,000 firm-year observations in the years up to 1980. This number increases to nearly 7,000 observations around the turn of the millennium. The standardized differences of all years are comparable. Because of the increasing number of observations between 1980 and 2000, the discontinuity must be stronger for earlier years. The annual standardized differences in Lahr (2014, p. 577) are consistent with that of Gilliam et al. (2015), with similarly high standardized differences for both periods but with more than twice as many observations around the turn of the millennium. Makarem et al. (2018, p. 409) evaluate more recent U.S. data from 2002 to 2011 and find “no obvious discontinuity around zero”, adding to the results of Gilliam et al. (2015) demonstrating the discontinuity has disappeared.

2.2.6 Research Question for Chapter 3

The literature presented above seems to be partly contradictory. The study by Dechow et al. (2003) finds a gradual decline in the zero earnings discontinuity. However, the study only evaluates data up to the year 2000, before the introduction of the SOX. Gilliam et al. (2015) show that there was a kink before the introduction of the SOX that disappeared afterward. They conclude that the disappearance of the discontinuity is causally related to the introduction of the SOX. If the SOX were truly responsible for the disappearance of the kink, one would expect a constant and strong discontinuity every year until 2001. However, this stands in contrast to the results of Dechow et al. (2003). Moreover, the similar standardized differences for earlier years with fewer observations compared with the standardized differences in later years with more observations in the papers of Gilliam et al. (2015) and Lahr (2014) indicate that the magnitude of the discontinuity has not been constant over time in the pre-SOX period.

Given the inconsistency in the literature, we identify the need for a study that revisits the annual discontinuity with a measure independent of sample size. Therefore, we formulate the following research question:

> *Is the disappearance of the zero-earnings discontinuity due to SOX, a gradual decline, or any other effect?*

2.3 International Evidence

Several studies examine the zero earnings discontinuity using international data. In a general context, the discontinuity tends to be larger in magnitude for long-term-oriented countries, while the discontinuity in short-term-oriented countries, such as the U.S., tends to be smaller (Haga et al., 2017, 2019; Leuz et al., 2003; Trimble, 2018). Burgstahler et al. (2006) and Daske et al. (2006) report a large zero earnings discontinuity for European firms. Leuz et al. (2003) use a sample from 1990 to 1999 and report the profit-loss ratio for different countries. For the U.S., they report a profit-loss ratio of 1.6, while for European countries, the ratios vary widely between 1.2 and 6.0. Burgstahler et al. (2006, p. 995) report a ratio of 1.98 for a public European sample from 1997 to 2003. In contrast, Glaum et al. (2004) compare the zero earnings discontinuities in the U.S. and Germany and find no significant differences.

2.3.1 IFRS

A substantial magnitude of the zero earnings discontinuity for European countries is undisputed in the literature. However, the effect of regulations on earnings management and the kink is not that clear and has been investigated in several studies. In 2005, IFRS became mandatory for firms in some European countries. Some studies conclude that the introduction of mandatory IFRS had little effect on earnings management (Doukakis, 2014; Jeanjean & Stolowy, 2008; Pereira & Alves, 2017; van Tendeloo & Vanstraelen, 2005), while others report a decrease in earnings management (H. Chen et al., 2010; Zéghal et al., 2011). Some even note an increase in managed earnings (Callao & Jarne, 2010; Capkun et al., 2016). Barth et al. (2008) find that voluntary adopters of IFRS in the European Union (EU) manage their earnings less, possibly related to the willingness for more transparency. Lin et al. (2012) find a higher level of earnings discretion under IFRS than under U.S. Generally Accepted Accounting Principles (GAAP) and conclude that firms reporting under U.S. GAAP have higher accounting quality on average. Trimble

(2018) assesses the zero earnings discontinuity for 46 countries adopting IFRS. She finds that the discontinuity slightly decreases in EU and non-EU countries, although it does not disappear.

Some studies aim to measure the effect of other accounting regulations on the zero earnings kink. Gore et al. (2007) evaluate the impact of U.K. regulations and find evidence that they have changed firms' earnings management strategies. Generally, they find a smaller discontinuity in the U.K. than in other European countries (Gore et al., 2007, p. 130). The lower discontinuities of Gore et al. (2007) could also be because they evaluate earnings before extraordinary items, whereas most other researchers use earnings after extraordinary items.[11] Enomoto and Yamaguchi (2017) report that for Japan, the zero earnings discontinuity did not disappear after the introduction of the Japanese SOX and conclude that the U.S. SOX reduces earnings management more efficiently than the Japanese SOX due to specific characteristics.

2.3.2 Research Questions for Chapter 4

The studies assessing the discontinuity around the adoption of IFRS tend to divide their sample into two groups, those before and after adopting IFRS. However, only an annual assessment allows the examination of the year-to-year changes in the discontinuity in more detail. Therefore, we formulate the following research question:

> *Is the zero earnings discontinuity in Europe around IFRS introduction stable and still substantial, while it has declined and eventually disappeared in the U.S.?*

The literature tends to find a consistently higher zero earnings discontinuity for Europe compared with the U.S., but it is not entirely clear why such differences arise. In Chapter 4, we (Martin Wallmeier and myself) conduct a European country-level analysis and

[11] Burgstahler and Chuk (2017, p. 731) demonstrate that earnings before extraordinary items have lower discontinuities because the stakeholders' primary focus is on net income after extraordinary items.

explore whether the extent of the zero earnings discontinuity could be related to cultural factors such as the Uncertainty Avoidance Index (UAI) of Hofstede et al. (2010), leading to the following research question:

> *Do European countries with a high UAI show a more pronounced zero earnings discontinuity?*

2.4 Alternative Thresholds

The benchmarks of earnings changes and earnings surprises are clearly less prevalent in the academic literature on discontinuities. Nevertheless, we summarize the state of the research and discuss possible confounding effects of the quantifying measures.

2.4.1 Earnings Changes

The earnings changes threshold measures the change in earnings from the previous year to the current year. Similar to the zero earnings threshold, the bottom-line net income is frequently used.[12] The income variable is often scaled by the market value of equity. The formal notation is presented below:

$$Earnings\ changes\ measure_{i,t} = \frac{Net\ Income_{i,t}\ -\ Net\ Income_{i,t-1}}{Market\ Capitalization_{i,t-1}}. \tag{2.6}$$

Most studies measure the standardized difference and find a discontinuity in earnings changes, with more firms than expected reporting higher or the same earnings as the previous year.[13] The studies observe that earnings changes are correlated with zero earnings (Burgstahler & Dichev, 1997; J. C. Hansen, 2010). Potentially this correlation could be

[12] See Brown and Caylor (2005), Burgstahler and Dichev (1997), S. K. Chen et al. (2010), Daske et al. (2006), Donelson et al. (2013), Durtschi and Easton (2005), Enomoto and Yamaguchi (2017), Jacob and Jorgensen (2007), and Yu (2014).

[13] See Burgstahler and Dichev (1997), S. K. Chen et al. (2010), Degeorge et al. (1999), Donelson et al. (2013), and Jacob and Jorgensen (2007).

caused by effects carrying over from the zero earnings threshold. Similar to their results on zero earnings, Durtschi and Easton (2005) find no discontinuity in earnings changes. At the international level, Daske et al. (2006) find a more pronounced discontinuity of earnings changes in Europe compared with the U.S. data.

There is criticism that the standardized difference measure is not appropriate for assessing the significance of the earnings changes discontinuity, which is in addition to the criticism of the zero earnings threshold. This criticism is raised because the earnings changes threshold is constructed such that the intervals around zero will result in the most observations, possibly even in the absence of earnings management (Dechow et al., 2003; Jacob & Jorgensen, 2007). Therefore, the standardized difference of the small profit interval $i + 1$ would naturally return a positive standardized difference. However, it is unclear what standardized difference should result if there is no earnings management. The same reasoning applies to the profit-loss ratio. A value greater than 1 indicates a discontinuity that could arise because of mechanical characteristics rather than earnings management.

Figure 2.3 presents the earnings changes distribution and illustrates the above-mentioned issue. It is unclear whether the highest frequency in the interval $i+1$ arises due to earnings management. Given this drawback, we refrain from formulating a research question on earnings changes.

2.4.2 Earnings Surprises

In 1998, former SEC Chairman Arthur Levitt expressed concerns in the CPA Journal that firms were engaging in earnings management to meet analysts' forecasts (Levitt, 1998). In this "numbers game", increasingly more firms would beat analysts' forecasts, which would positively affect their market value. This pressure would incentivize firms to provide guidance on their future earnings. Analysts would be briefed on recent information in conference calls before it was released to the public. After the conference calls and before the information is released to the public, problematic, unusual trading would occur.

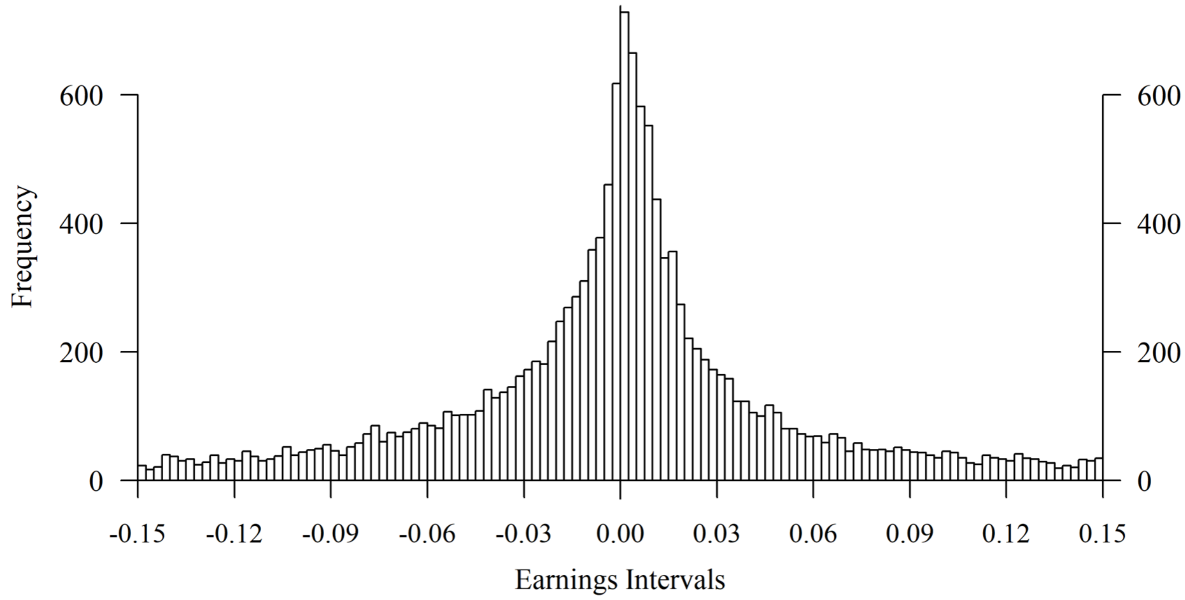

Figure 2.3: Earnings Changes Discontinuity. The figure illustrates the earnings changes discontinuity. It is unclear whether the highest frequency in the interval $i+1$ arises due to earnings management or by construction. To illustrate the effect of potentially many observations in the interval $i+1$, we use stock quoted U.S. firms from 2012 to 2016 from the Refinitiv Worldscope universe and scale the change of Bottom Line Net Income (WC #1651) (i.e., the net income of this year less the net income of last year) by the lagged Market Capitalization (WC #8001) of each year. The size of an earnings interval is 0.0025.

Researchers investigate the "numbers game" and find that significantly more firms meet or exceed analysts' forecasts than firms that fail to reach them.[14] Indeed, meeting or beating the analysts' targets is beneficial to the stock price because it communicates continuity and stability to stakeholders (Burgstahler & Eames, 2006; Payne & Robb, 2000; Xue, 2003). Firms that miss analysts' earnings targets are punished by the market (Xue, 2003).

Some studies observe the earnings surprises discontinuity over different periods and reach different conclusions about whether the earnings surprises discontinuity has increased or not. Brown and Higgins (2001, p. 382) find an increasing discontinuity through the profit-loss ratio between 1994 to 1996 and 1997 to 1999, suggesting that the earnings surprises threshold became more apparent over time. However, the model of S. K. Chen et al. (2010) shows only a marginal increase in the number of firms practicing earnings management in the 1990s. Only in the years after the turn of the millennium does the frequency of firms meeting the earnings surprises thresholds rise massively to over 10%

[14]See Brown and Higgins (2001), Burgstahler and Eames (2006), S. K. Chen et al. (2010), Degeorge et al. (1999), and Payne and Robb (2000).

of all firms. For the years following the SOX introduction, Bartov and Cohen (2009) report, in contrast, that fewer firms meet or beat analysts' forecasts than prior to the SOX introduction. They explain this decline by a lower level of earnings management. European data suggest a much lower earnings surprises discontinuity than in the U.S. (Daske et al., 2006; Gore et al., 2007). On average, the earnings surprise discontinuities in Central European countries and the U.K. increased only slightly between 1988 and 1999 (Brown & Higgins, 2001).

The characteristics of the earnings surprise threshold differ from the zero earnings and earnings changes threshold. It is important to note that the earnings surprise threshold does not refer to an absolute, unchanging number but is defined by analysts. The threshold changes during the financial reporting period. Over the course of the year, analysts often reduce their earnings forecasts so that firms might be able to meet the forecast (Bartov et al., 2002; Richardson et al., 1999). This interaction between management and analysts allows firms to guide analysts close to released earnings in the future.

In addition to the above characteristics, the technical measurement and comparison of earnings surprises are also more complex due to the different choices of periods between the analysts' mean EPS forecast and the earnings announcement date, rounding effects, scaling, and interval size choices. The following list highlights the technical difficulties and differences between the studies.

1. The time difference between the analysts' mean EPS forecast and the earnings announcement date varies. Analysts tend to release relatively high forecasts at the beginning of the year. Over time, they reduce their forecasts which are lowest just before the earnings announcement (Richardson et al., 1999). Different academic studies use different time horizons between the actual earnings forecast and the earnings announcement date. Daske et al. (2006, p. 141) use the last forecast before the earnings announcement month. Lahr (2014, p. 582) uses the forecast in the reporting month as well as one, three, six, and nine months prior to the earnings announcement, similar to Burgstahler and Eames (2006, p. 636). The Institutional Broker's Estimate System (I/B/E/S) provides a self-calculated earnings surprise variable consisting of the latest average EPS forecasts on the day before

the earnings announcement.[15] It appears plausible to expect that the forecast accuracy of I/B/E/S is higher. Therefore the earnings surprise is lower than earnings surprises from analyst forecasts published several months before the earnings announcement, as analysts may include more recent information such as quarterly earnings releases. In general, the results of earnings surprise studies should be compared with caution if these studies use different time horizons between the analysts' mean EPS forecast and the earnings announcement date.

2. EPS data is frequently rounded to whole USD cents. The I/B/E/S provides an item that includes the earnings surprise in USD (EPSSURDIF), consisting of the annual realized EPS less the average EPS forecast prior to the earnings announcement (Thomson-Reuters, 2010, p. 112). However, this item is sometimes rounded to whole USD cents and sometimes to tenths of a USD cent.[16]

Figure 2.4 illustrates the rounding issue with U.S. data from the I/B/E/S computed EPSSURDIF from -0.05 to +0.05 USD. Each interval corresponds to one-tenth of a USD cent, the smallest unit of this variable. In this figure, however, 30.5% of all observations are rounded to whole cent values. The rounding causes a loss of information, potentially resulting in more exact zero forecast errors (Payne and Robb, 2000; Lahr, 2014, p. 561). Even worse, rounding might turn earnings that miss the benchmark into earnings that meet the benchmark. Consider the following scenario, where analysts forecast, on average, EPS of 0.03 USD. If the firm releases EPS of 0.028 USD on the day of the earnings announcement, it will signify that it missed the analysts' earnings target. However, if the negative earnings surprise of 0.002 USD is rounded to whole USD cents, a small miss of the earnings surprise threshold suddenly turns into an exact zero, interpreted as meeting or beating the analysts' earnings target.

[15] Stock splits are treated differently in the I/B/E/S and Worldscope database. I/B/E/S adjusts for stock splits retroactively, while Worldscope does not. Mixing adjusted and unadjusted data would result in misclassified observations (Baber & Kang, 2002; Payne & Thomas, 2003). Since both databases contain realized data, it is recommended to exclusively use I/B/E/S data for the earnings surprises computation.

[16] An alternative method of calculating the earnings surprise in USD cents is to retrieve the I/B/E/S EPS minus the I/B/E/S EPS forecast for the same year. However, this results in an even more frequent rounding to whole USD cents.

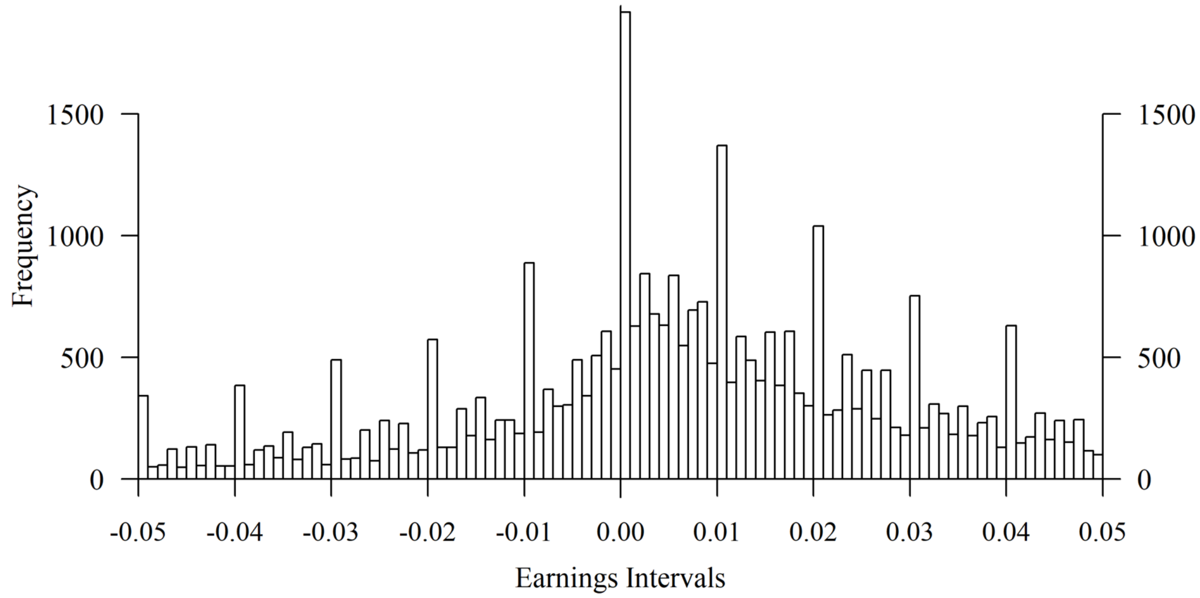

Figure 2.4: Unscaled Earnings Surprises. The figure illustrates the earnings surprise threshold. Frequently unscaled earnings in the I/B/E/S database (item EPSSURDIF) are rounded to whole USD cents. To plot this figure, we used stock quoted U.S. firms from 1993 to 2017, with at least three analysts covering the firm. The size of an earnings interval is 0.001 USD.

This effect has not remained unnoticed by research. Lahr (2014, p. 589) argues that the earnings surprise discontinuity could arise from rounding mechanisms rather than earnings management. The earnings surprise discontinuity in Burgstahler and Eames (2006) would arise due to the high number of exact zeros in the interval $i + 1$ (Lahr, 2014, p. 583). After excluding the exact zeros, he finds fewer observations in the interval $i + 1$ than in the neighboring interval $i + 2$ and about the same number of observations as in the interval $i - 1$, indicating there is absolutely no earnings management to meet the analysts' forecasts. However, we caution from this interpretation. Even though a considerable portion of the exact zeros could be due to rounding effects, some firms might actually report earnings that exactly match the forecasted earnings. In this context, it is challenging to find a consistent method that excludes only rounded exact zeros, but not observations where firms have actually reported earnings that are equal to analysts' forecasts. The proposed solution by Lahr (2014) is that future research should calculate earnings surprises on an earnings basis rather than an EPS basis. The I/B/E/S database contains an unscaled net income forecast (I/B/E/S item INC). However, it has to be noted that it is only available as of December 1994 (Thomson-Reuters, 2010, p. 28) and is also less commonly used in research than the EPS forecast.

3. The literature sometimes scales earnings surprises and sometimes not. Some researchers present earnings surprise histograms of unscaled USD values (Cohen et al., 2008; Degeorge et al., 1999; Dhaliwal et al., 2004; Donelson et al., 2013). Since the PPS varies widely across firms, reporting absolute USD values of earnings surprises presents a distorted picture. Suppose the scenario where a firm beats analysts' forecasts by 1 USD. If that firm's PPS is 10 USD, the earnings surprise would be substantial. However, if the PPS is 100 USD, the earnings surprise would not be as impressive anymore. Therefore, it is essential to set the USD values in context and use a scalar like the lagged PPS.

4. Different interval sizes are used by researchers who scale earnings surprises using the PPS, potentially greatly affecting the discontinuity. Some researchers use an interval size of 0.0025 of scaled earnings (Brown & Caylor, 2005; Daske et al., 2006; Gore et al., 2007; Payne & Robb, 2000), equal to the size of earnings changes used by most researchers. The drawback of this relatively large interval size is that the observations are centered closely around zero. As with earnings changes, it is uncertain whether the spike in the interval $i+1$ is due to distributional properties or actually due to earnings management. However, for earnings surprises, this effect is stronger, leading to even higher positive standardized differences for the interval $i+1$ than in the case of earnings changes.

Some studies use a much smaller interval size of 0.0002 (Burgstahler & Eames, 2006; Lahr, 2014). This small range implies that only firms that exceed analyst forecasts by less than 0.02% of their lagged PPS are classified in the interval $i+1$ and thus tested for earnings management. It has the disadvantage that most firms in the interval $i+1$ have exactly zero earnings surprises. As already mentioned, it remains unclear whether these values arise due to rounding or actually exactly match the analysts' forecasts and complicate the interpretation.

Our literature review noted differences in research design choices and, more importantly, identified effects that might influence commonly used quantifying discontinuity measures. Indeed, the earnings surprises threshold is an exciting phenomenon. However, its measurement and interpretation are difficult. In particular, the inability to separate exact zeros into actual zero earnings surprises and rounding issues, let us refrain from formulating a research hypothesis based on the choice of previous research designs.

2.4.3 Threshold Concurrence

The three earnings thresholds may all appear as attractive earnings targets for managers to achieve. In this context, it is essential to understand that in most cases, managers have to opt for one of the thresholds because the thresholds have different earnings values. Only on rare occasions do two thresholds take the same value. This is only the case for the zero earnings and earnings changes threshold if last year's earnings were zero. Identical zero earnings and earnings surprises thresholds imply that analysts forecast zero earnings. Finally, earnings changes and earnings surprises are identical only if analysts predicted earnings to be equal to last year.

Figure 2.5: Different Earnings Thresholds. The figure visualizes possible different thresholds managers have to opt for. Real, unmanaged earnings are likely to differ from the thresholds. Managers can then decide whether to engage in earnings management, depending on the firm's individual situation.

Figure 2.5 illustrates the trade-off firms often face when choosing the best threshold for the current situation. In the situation shown, the real unmanaged earnings would be 10. Managers can decide whether to engage in earnings management to reach a threshold and, if so, which threshold they want to reach. From the managers' perspective, not all thresholds might be equally important. Graham et al. (2005, p. 22) asked CFOs in their questionnaire about the importance of different quarterly earnings thresholds. The most critical quarterly threshold is to beat last year's EPS, followed by analysts' forecasts, reporting earnings above zero, and beating earnings of the previous quarter.

Mindak et al. (2016) aim to identify the most relevant threshold for each firm-year using a data-driven approach. Depending on which threshold is closest, they assign an achievable threshold to each firm, which can be either zero earnings, earnings changes, or earnings surprises. They observe that most firms aim for analysts' forecasts, followed by earnings changes and zero earnings. The analysts' forecast is often the highest threshold in terms of earnings. The market would classify firms that meet the highest of the three thresholds as "good" and reward them with a positive cumulative average market return in the three days following the earnings release. Mindak et al. (2016, p. 200) report that two-thirds of firms would manage earnings upward. Firms that choose to meet a threshold other than the highest threshold are classified as "bad", resulting in a negative cumulative average market return (Mindak et al., 2016, p. 202).

2.5 Conclusion

This chapter presents the literature on earnings thresholds, critically analyzes the methodological shortcomings, and formulates research questions that form the basis for Chapter 3, joint work with Martin Wallmeier and Peter Fiechter, and Chapter 4, joint work with Martin Wallmeier.

We discuss the different data selections of previous studies and the drawbacks of the standardized difference commonly used to quantify a discontinuity. We argue that the significance of the standardized difference varies with the sample size for discontinuities of the same magnitude. Therefore, we caution comparing standardized differences. Further, we present the profit-loss ratio that divides the number of observations of the small profit interval by the number of observations of the small loss interval.

The studies of Beaver et al. (2007) and Durtschi and Easton (2005, 2009) raised doubts whether the zero earnings discontinuity truly arises from earnings management to avoid losses. However, there are several reasons to believe that alternative explanations cannot fully explain the zero earnings discontinuity. Sixty-five percent of CFOs report that it is important or very important to report positive quarterly earnings (Graham et al., 2005). Likewise, CFOs expect nearly 20% of all earnings releases to contain misrepresented earnings (Dichev et al., 2013).

The findings of the various studies in the literature are not entirely consistent. Dechow et al. (2003) report a decreasing zero earnings discontinuity from 1989 to 2001. In contrast, Gilliam et al. (2015) observe a large discontinuity in the years before the SOX introduction in 2002 and no discontinuity afterward. From this dissonance, we derive our research question that we (Martin Wallmeier, Peter Fiechter, and myself) attempt to answer in Chapter 3 by analyzing whether the disappearance of the zero earnings kink is attributable to the SOX introduction or possibly to other effects.

In the European context, previous literature has identified a zero earnings discontinuity that is even greater than the U.S. discontinuity. IFRS became mandatory for some European countries in 2005. The study of Trimble (2018) grouped earnings similarly to Gilliam et al. (2015) in a period before and after adopting IFRS and found a slight decrease in zero earnings discontinuity for most countries. However, it remains unclear whether this small decline ultimately occurred around the adoption of IFRS. This leads to our first research question in Chapter 4, where we (Martin Wallmeier and myself) revisit how the zero earnings discontinuity evolves from year to year. The second research question is whether the UAI of Hofstede et al. (2010) can explain the differences in discontinuities at the country level.

In the third part of this chapter, we present the earnings changes and earnings surprises threshold. We caution against interpreting that a positive standardized difference in the small profit interval $i+1$ is entirely due to earnings management. In addition, we highlight the different measurement methods for earnings surprises. We agree with Lahr (2014) that some of the discontinuity may arise from rounding to the nearest USD cent. However, excluding observations with exact zeros may also introduce a bias. As long as it remains unknown whether zero earnings surprises result from rounding errors, true exact forecasts, or earnings management it is difficult to quantify the extent of earnings management for the earnings surprise threshold.

Finally, we would like to emphasize that the thresholds compete with each other. In most scenarios, only one threshold can be attained. The management may opt for the best threshold at reasonable costs and risks.

Chapter 3

THE DISAPPEARANCE OF THE ZERO-EARNINGS DISCONTINUITY: SOX, DOTCOM BOOM OR GRADUAL DECLINE?

The Disappearance of the Zero-Earnings Discontinuity: SOX, Dotcom Boom or Gradual Decline?

June 2021, this version May 2022

Patrick Chardonnens

University of Fribourg, Bd. de Pérolles 90, CH-1700 Fribourg, Switzerland,
Email: patrick.chardonnens@unifr.ch.

Peter Fiechter

Université de Neuchâtel, Rue A.-L. Breguet 2, CH-2000 Neuchâtel, Switzerland,
Email: peter.fiechter@unine.ch.

Martin Wallmeier

University of Fribourg, Bd. de Pérolles 90, CH-1700 Fribourg, Switzerland.
Email: martin.wallmeier@unifr.ch (corresponding author).

Abstract

The zero-earnings discontinuity in the US disappeared around the time when the Sarbanes-Oxley Act (SOX) became effective, suggesting that SOX may have reduced the small loss avoidance by firms. In this paper, we examine a potential confounding effect arising from the dotcom boom at the turn of the millennium. Many newly listed dotcom firms had no revenues but high market capitalizations. Therefore, they mechanically fell into the smallest loss interval, artificially reducing the zero-earnings discontinuity. Once this dotcom effect is accounted for, our results no longer suggest a sharp (causal) effect of SOX on the decline in the zero-earnings discontinuity.

Keywords: Earnings management, Zero-earnings discontinuity, SOX, Dotcom boom, Earnings distribution, Loss avoidance

JEL: M48, G38, M41

3.1 Introduction

Burgstahler and Dichev (1997) show that the frequency distribution of earnings scaled by the lagged market value of equity shows a discontinuity at zero, indicating that many firms report small profits, while few firms report small losses. This phenomenon is known as the zero-earnings (ZE) discontinuity or ZE kink. Numerous studies in several countries have confirmed the discontinuity.[1] Although different explanations for the ZE kink have been proposed, "the theory that earnings are managed to meet benchmarks provides the most simple and complete explanation for the body of evidence" (Burgstahler & Chuk, 2017, p. 744).[2]

Interestingly, Gilliam et al. (2015) show that the zero earnings (ZE) kink has disappeared in the United States (US), which is of great interest to regulators and researchers.[3] Gilliam et al. (2015, p. 188) "are able to identify a critical turning point when the zero-earnings discontinuity becomes imperceptible." This turning point is the year 2002, which "is consistent with ... a decline in loss avoidance after SOX" (p. 188). It is also "consistent with prior research suggesting SOX reduced accrual earnings management" (p. 124).[4] It therefore appears that the Sarbanes-Oxley Act (SOX) had a real and lasting impact on earnings management in line with the regulation's general objective to improve the reliability of financial reporting.[5]

While the time of the turning point appears to be narrowly defined, Gilliam et al. (2015, p. 199) "caution that passage of SOX is not the only important event occurring during our time period of interest that may have affected the discontinuity. Examples

[1] See, e.g., Degeorge et al. (1999), Dechow et al. (2003), Burgstahler and Eames (2006), Leuz et al. (2003), Burgstahler et al. (2006), Gore et al. (2007), Daske et al. (2006), and Burgstahler and Chuk (2017).

[2] For the alternative explanations, see Durtschi and Easton (2005, 2009).

[3] The results are different in other countries. See, e.g., Enomoto and Yamaguchi (2017) for Japan.

[4] Gilliam et al. (2015) refer to Bartov and Cohen (2009), Cohen et al. (2008), Lobo and Zhou (2006) and Lobo and Zhou (2010). See Chung et al. (2009) for the impact of Sarbanes-Oxley Act (SOX) on the relation between earnings management and equity liquidity and C. C. S. Chen et al. (2020) for country factors in earnings management of ADR firms.

[5] See Amar et al. (2022) for evidence on the motives of financial misconduct.

include the collapse of Enron in 2001, the registration of US firms with the PCAOB in 2003, and the global investment settlement in 2003".

We identify the dotcom boom as another potential confounder that has not yet been considered in prior literature. In the second half of the 1990s, many firms with sales revenues close to zero went public and were added to financial databases. Without revenues, they inevitably suffered losses that were often small in relation to the high market values of equity at the peak of the dotcom boom. Therefore, small losses occurred much more often than small profits in this group, which apparently reduced the ZE discontinuity in the overall sample. However, this reduction was mechanically driven by the addition of the dotcom firms to the sample, rather than a true decline in earnings management. As these events occurred around the turn of the millennium, their impact might be confounded with the effect of SOX in 2002. Therefore, the objective of this research note is to examine the role of the dotcom effect in explaining the observed decline in the ZE discontinuity over time.

Our empirical findings show that once the dotcom effect is accounted for, there is no sharp decline in the ZE discontinuity when SOX was introduced; rather, the decline is gradual over time. As this evidence does not suggest a causal effect of SOX on the ZE discontinuity, our findings contribute to the literature that investigates the effects of SOX. In addition, our findings have implications for regulators and policymakers in assessing the costs and benefits of SOX, lest they overestimate the disciplinary effect of SOX on earnings management.

3.2 Data and Method

3.2.1 Data

We use merged CRSP and Compustat data of US firms from 1987 to 2019. We exclude firms in regulated industries and financial institutions with Standard Industrial Classification (SIC) codes ranging from 4400 to 5000 and 6000 to 7000.[6] We require net income and total sales to be available. The beginning of the year market value of equity and total assets must be larger than zero. Following previous research, we remove observations with a net income of exactly zero (47 firm-years).[7] Net income is the bottom line position that includes operating and nonoperating income after extraordinary items (Compustat item NI). It is scaled by the beginning of the year market value of equity (Compustat item CSHO*PRCC_F). The final sample contains 161,072 firm-years. Figure 3.1 shows the distribution over the years.

To capture the dotcom effect, we identify firms without substantial revenues by using a threshold for total sales of 2 million USD. We do not set the limit at zero in order to also capture firms with slightly positive sales that are not economically important. Our results are robust with respect to the exact threshold at least in a range of 0 to 10 million USD. Overall, the percentage of firms without substantial sales is 12.7%. Figure 3.1 shows the distribution over time (dark grey). The proportion rises from a low of 6.9% in 1995 to 15.7% in 2003 and a maximum of 18.4% in 2014. Note that not all firms can be classified as "dotcom firms" even if they went public during the dotcom boom. There is, for example, an important group of firms engaged in biotechnology and pharmaceutical research.

3.2.2 Discontinuity Measures

As in most of the literature, our primary earnings measure is net income scaled by the market value of equity at the beginning of the fiscal year. As a robustness check, we use total assets as a scaling variable.

A first commonly used discontinuity measure is the standardized difference between the actual number of observations in the smallest loss interval and the expected number of

[6] See, e.g. Beaver et al. (2007), Brown and Caylor (2005), Burgstahler and Dichev (1997), S. K. Chen et al. (2010), Durtschi and Easton (2005), Durtschi and Easton (2009), Gilliam et al. (2015), Haga et al. (2017), Kerstein and Rai (2007), Roychowdhury (2006), Makarem et al. (2018) and Li (2019).

[7] See Burgstahler and Dichev (1997), Gilliam et al. (2015), Dechow et al. (2003), Beaver et al. (2007), Burgstahler and Eames (2006), and Lahr (2014).

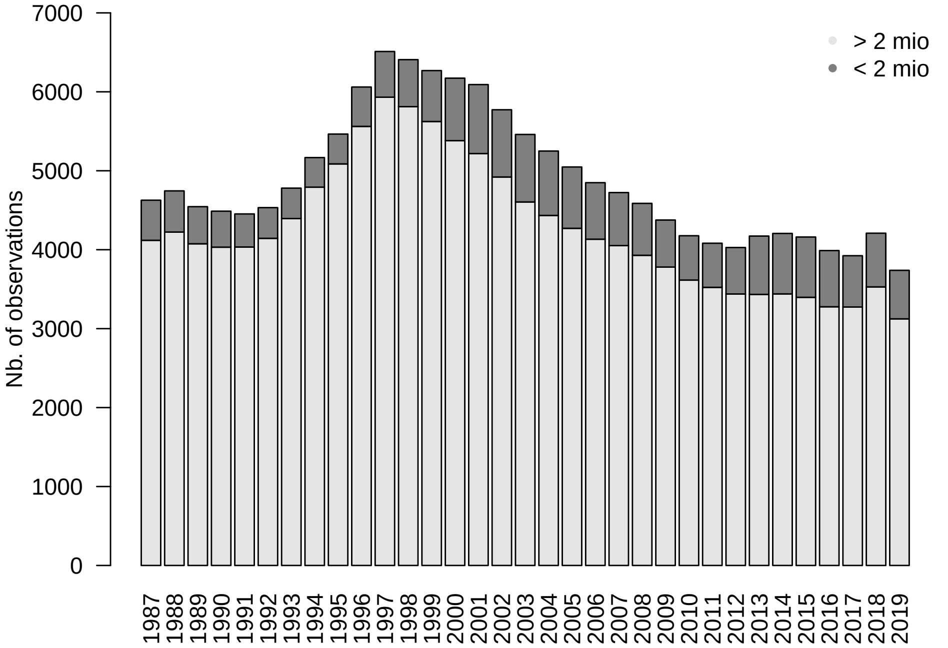

Figure 3.1: Sample Size by Year (1987-2019). The two categories are 1) Total sales greater than 2 million USD (lightgrey), and 2) Total sales less than 2 million USD (darkgrey).

observations assuming no discontinuity. Formally, $SD_{-1} = \left[N_{-1} - 0.5(N_{-2} + N_1)\right]/s_{-1}$, where N_{-1} is the number of observations in the smallest loss interval, N_{-2} and N_1 are the numbers of observations in the neighboring intervals to the left and right, respectively, and s_{-1} is the standard error of the difference.[8] The standardized difference for the first profit interval, SD_1, is defined analogously. A second commonly used measure is the profit-to-loss ratio. It is defined as the ratio of the number of observations in the first profit and loss intervals: $PL = N_1/N_{-1}$ (Dechow et al., 2003). In the presence of a ZE discontinuity, SD_{-1} will be negative, SD_1 will be positive and PL will be larger than 1.

Both measures have certain weaknesses. For a given level of the discontinuity, the SD statistic increases as the sample size increases. This blending of the effect size and test power is undesirable in comparisons over time when the sample size varies. Specifically, when the sample size is small in early years, SD_{-1} and SD_1 will underestimate the potential decline in the ZE discontinuity. Another critical aspect is that the expected frequency is defined as the average frequency of the adjacent intervals. When these in turn

[8] The standard error is computed as $s_{-1} = \sqrt{Np_{-1}\left(1 - p_{-1}\right) + 0.25N\left(p_{-2} + p_1\right)\left(1 - p_{-2} - p_1\right)}$, where p_i is the proportion of observations falling in interval i.

are distorted, the statistic is difficult to interpret. Furthermore, in the case of a positive mean of scaled earnings, the normal profit-to-loss ratio (without a ZE discontinuity) is already greater than 1.

To avoid these weaknesses, we modify the two measures by using a kernel density estimation inspired by Lahr (2014). Our purpose is not to estimate the complete density but to find a smooth representation of the frequency distribution in the relevant area. Therefore, we only include scaled earnings of -0.15 to 0.15. As the time pattern of the estimated discontinuity is almost the same for different versions of kernel estimators, we apply a standard Gaussian estimator.[9] Figure 3.2 shows an example (data for 1997 with a substantial ZE discontinuity). Our measure "small loss deviation" is defined as $SLD = (Actual_{-1} - Expected_{-1})/Expected_{-1}$, where $Actual_{-1}$ is the actual number of observations in the first loss interval and $Expected_{-1}$ is the expected number according to the kernel density estimation (integral over the first loss interval). The "small profit deviation" is analogously defined as $SPD = (Actual_1 - Expected_1)/Expected_1$, where subscript 1 represents the first profit interval. Our modified profit-to-loss ratio MPL is defined as $MPL = ln\,[N_1/N_{-1}] - ln\,(EPLR_1)$, where $EPLR$ is the expected profit-to-loss ratio without discontinuity (integral of the kernel density over the first profit interval divided by the integral over the first loss interval).

To allow comparison with previous results, we use the same interval widths of 0.005 and 0.015 as Gilliam et al. (2015). In the case of the interval width 0.015, the observations of the first three loss and profit intervals of width 0.005 are compared to the expected value according to the kernel density over the same range.

[9] The density is estimated at 512 equally spaced points. We implement the estimation in R ("density" function) with the bandwidth proposed by Scott (1992) (option bw.nrd in R).

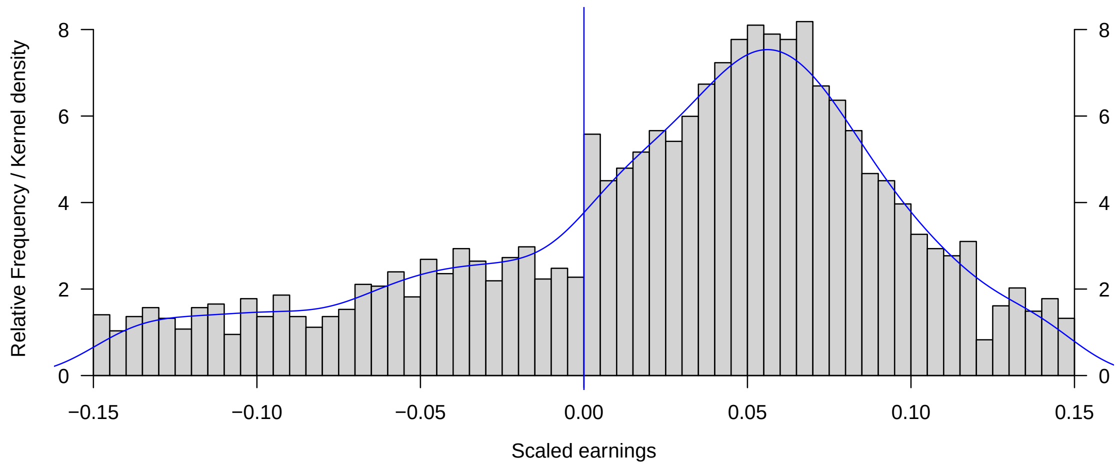

Figure 3.2: Illustration of Kernel Density Estimation (1997, all Firms). For interval width 0.005, we have: $SLD = -0.36$ and $SPD = 0.40$. This means that the number of observations in the first loss (profit) interval is 36% smaller (40% larger) than expected according to the kernel density (in blue). For interval width 0.015, we have: $SLD = -0.28$ and $SPD = 0.13$. This means that the discontinuity is smaller when the three first profit intervals and the three first loss intervals (of width 0.005 each) are taken together.

3.3 Empirical Results: The Zero Earnings Discontinuity Over Time

To track the change in the ZE discontinuity over time, we compute the small loss difference SLD and the small profit difference SPD on a yearly basis and show the results in Figure 3.3. The graphs on the left and right are based on interval widths of 0.005 and 0.015, respectively. For scaling with the market value of equity (MVE), the upper panels (A1 and A2) represent the total sample and the middle panels (B1 and B2) represent the subsample of firms with sales revenues greater than 2 million USD in the respective year. The graphs in the bottom panels (C1 and C2) show results for scaling with total assets (so that scaled earnings correspond to the return on assets, ROA). The year 2001 is marked by a vertical line because it is the year before SOX was enacted (on July 30, 2002).

Before 1995, half of the expected number of cases in the narrow loss interval are “missing”, indicating a pronounced discontinuity ($SLD \approx -0.5$; graph A1), which is consistent with prior studies (e.g., Burgstahler & Eames, 2006; Dechow et al., 2003;

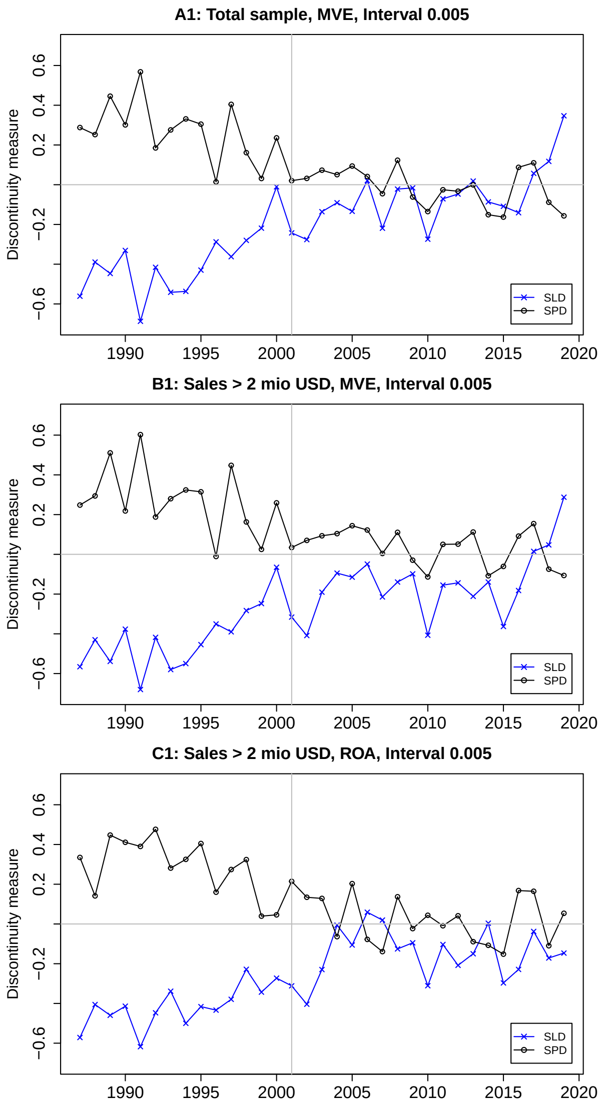

Figure 3.3: Discontinuity Measure Over Time. The discontinuity measure captures the share of excess observations (pos. sign) or missing observations (neg. sign) in the intervals of scaled earnings directly below and above the zero threshold. SLD: small loss difference, and SPD: small profit difference. The year 2001 is highlighted by a vertical line because it is the last year before SOX was enacted. MVE: scaling with the market value of equity, and ROA: scaling with total assets.

Degeorge et al., 1999; Leuz et al., 2003). From 1995 to 2006, SLD shows an upward trend, with a spike in 2000 to a level of zero. After 2006, SLD stays mainly negative without a clear trend. This overall pattern is reflected in the downward trend of the small profit difference SPD. The absolute values of SPD are generally smaller than those of SLD, which means that the missing small losses are not simply turned into small profits. We find the same overall pattern for the wider interval width of 0.015, but with smaller values (graph A2). Despite this similarity, the graph for interval 0.015 supports a different view: it suggests that the discontinuity declined but still remained at a high level until

2002 before it disappeared in 2003 and did not reappear afterward. This pattern replicates the one reported by Gilliam et al. (2015).

In the reduced sample of firms with revenues of at least 2 million USD (panels B1 and B2), the pattern is less similar to the one in Gilliam et al. (2015). Based on the narrow loss interval of 0.005, SLD decreases significantly, especially in the periods from 2001 to 2002 and 2010 to 2015. The decrease is less strong for the wide loss interval of 0.015, but still considerable. Given the negative SLD values from 2008 onwards, the degree to which the kink has disappeared is unclear.

The reason why firms without sales revenues affect the measured discontinuity is apparent from the frequency distribution of scaled earnings shown in the right graph of Figure 3.4. As expected, the distribution is almost fully positioned in the negative range. It tends to rise uniformly almost up to the zero threshold so that the smallest loss interval represents many more cases than the smallest profit interval. Therefore, including these firms mechanically reduces the ZE discontinuity in the overall sample.

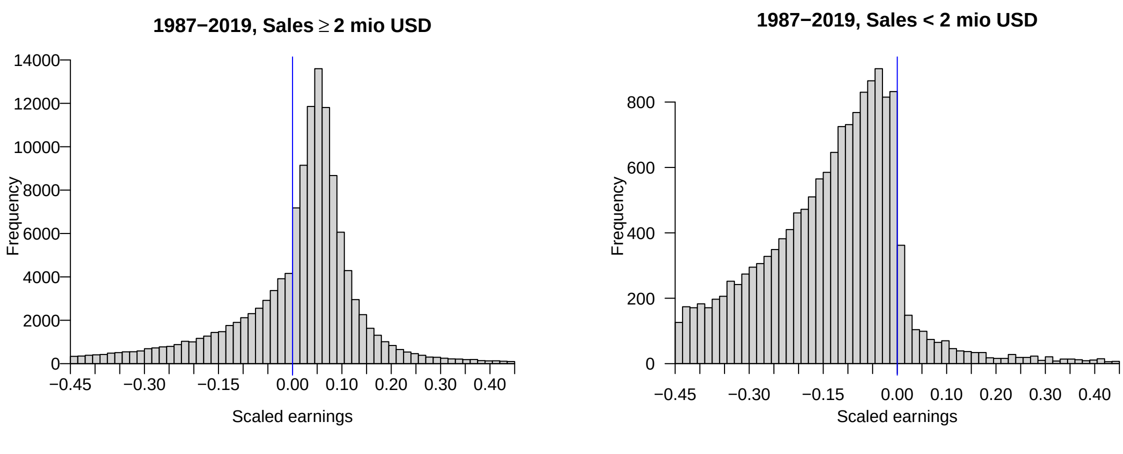

Figure 3.4: Frequency Distribution of Scaled Earnings for Firms With Sales Revenues of at Least 2 Million USD (Left Graph) and Less Than 2 Million USD (Right Graph). The zero threshold is highlighted by a vertical line. The interval width is 0.015. Scaling with the market value of equity.

For scaling with total assets, we only show results for the reduced sample of firms having sales revenues larger than 2 million USD (Fig. 3.3, C1 and C2). The graphs for the total sample are practically identical. The reason is that the losses of firms without sales are mostly substantial in relation to total assets so that few cases fall into the smallest loss interval. Based on the interval of 0.005 (C1), SPD and SLD converge until 2003. The discontinuity varies in the following years, becoming pronounced in some years (in particular 2010) and even inverse in others. The results for the interval of 0.015 (C2) are more stable. The graph suggests that the discontinuity measures followed a declining trend over the sample period and did not experience noticeable shifts.

Since we do not observe a sharp decline of the ZE discontinuity in 2002, our results are not consistent with a causal effect of SOX on the disappearance of the discontinuity (Atanasov & Black, 2016). A possible explanation for the gradual decline starting in the 1990s refers to changes in the listing requirements of the New York Stock Exchange (NYSE). For example, Dechow et al. (2003, p. 379) argue that prior to 1995, positive income was an important requirement of the NYSE's continued listing standards. In 1995, losses became irrelevant for continued listing as long as certain levels of revenues, market capitalization and operating cash flows were met. Since 1999, the rule is that losses do not prevent continued listing if certain (low) thresholds of market capitalization are met (currently, e.g., shareholders' equity of 6 million USD if net losses were reported in the last 5 fiscal years). Dechow et al. (2003, p. 379) state the following: "This shift away from an "income" focus has had a dramatic effect on the distribution of net income for NYSE firms." As a result, investors and analysts also may have adopted a more neutral stance towards small losses, which encouraged even more firms to report small losses as they occurred.

Appendix A.1 reports supplementary statistics defined in Section 3.2, including the standardized differences that can be used to assess the statistical significance of the kink in individual years. We also check whether changes in the sample composition that are not related to firms without sales have an effect. Our results prove to be robust in this respect as we find the same gradual decline for a balanced sample of stocks that were available in 2002.

3.4 Conclusion

The disappearance of the ZE discontinuity is an important discovery in the earnings management literature. Previous findings identify a turning point around 2002, which suggests that SOX might have been successful in decreasing earnings management activities. We argue that the dotcom boom at the turn of the millennium introduces a potential confounding effect, which (mechanically) contributes to the decline in discontinuity measures. When filtering out the dotcom effect, we note two key findings. First, it is unclear whether the ZE discontinuity has disappeared completely, because it flares up again in later years. Second, the decline is gradual over time rather than immediately after the introduction of SOX; yet such a sharp decline would be required to conclude a causal effect of SOX.

A limitation of our study is that we cannot clearly determine what caused the gradual decline of the ZE discontinuity. It is plausible that changes in the NYSE listing requirements played an important role, but more research on the driving forces of the decline would be valuable.

A general implication of our study for regulators and policy makers is that it is important to consider potential confounders in policy evaluation. In our case, without taking the dot-com effect into account, there is a risk of overestimating the disciplinary effect of SOX on earnings management. It would have been a great success of the SOX reform if it had ended the widespread practice of avoiding small losses. However, according to our analysis, the decline of this practice was not caused by SOX. Our results could therefore significantly alter the cost-benefit assessment of the SOX reform.

Chapter 4

KINKED ACCOUNTING? LOSS AVOIDANCE IN EUROPE AND (NOT) THE US

Kinked Accounting?
Small Loss Avoidance in Europe and (Not) the US

August 2022

Patrick Chardonnens

University of Fribourg, Bd. de Pérolles 90, CH-1700 Fribourg, Switzerland,
Email: patrick.chardonnens@unifr.ch

Martin Wallmeier

University of Fribourg, Bd. de Pérolles 90, CH-1700 Fribourg, Switzerland.
Email: martin.wallmeier@unifr.ch (Corresponding author.)

Abstract

For a long time, the most vivid evidence of earnings management has been the discontinuity of earnings distributions at the threshold of gains and losses, indicating loss avoidance. Strikingly, this discontinuity has declined and eventually disappeared in the US. We investigate whether a similar evolution has occurred in Europe and find that this is not the case: the discontinuity in Europe is still substantial and has remained remarkably stable. Our country-level analysis suggests that loss avoidance is rooted in a culture of uncertainty avoidance, which may explain the observed stability in Europe. The cultural factor seems to be stronger than countervailing factors such as international accounting harmonization, global competition and SOX-related regulations, which would rather suggest an alignment of loss-avoidance behavior.

Keywords: Earnings management, Loss avoidance, Earnings quality, Zero earnings discontinuity, IFRS, Earnings distribution

JEL: M48, G38, M41

4.1 Introduction

Probably the most vivid evidence of earnings management is that earnings distributions show a discontinuity at the threshold of gains and losses. This so called zero earnings (ZE) discontinuity (also referred to as the ZE kink) was discovered by Burgstahler and Dichev (1997) and has been confirmed by many other studies.[1] Strikingly, the situation in the US seems to have changed drastically as Gilliam et al. (2015) show that the ZE discontinuity disappeared at the turn of the millennium. The reasons for this are not entirely clear. One hypothesis is that the Sarbanes-Oxley Act (SOX) has successfully pushed back earnings management (Gilliam et al., 2015). Another view is that the discontinuity declined already in the second half of the 1990s owing to changes in the continued listing rules of the New York Stock Exchange (NYSE) (Dechow et al., 2003). There is also evidence that a change in the sample composition during the internet bubble contributed to the observed decline (Chardonnens et al., 2022).

Some arguments make a similar decline in Europe seem logical.[2] A decade after SOX, the European Union (EU) established its own set of rules for audits, public oversight of audits and investor protection (Directive 2014/56/EU; Regulation EU No 537/2014). Since 2005, the mandatory adoption of International Financial Reporting Standards (IFRS) has redefined and partly curtailed managers' discretion to smooth reported earnings or hide losses (Ball, 2016). In addition, accounting practices could spill over directly from the US to Europe because in a globalized world, a firm's profitability is assessed with respect to its international peers.

However, there are also reasons to believe that loss avoidance has been preserved in Europe. It is well known that incentives and the institutional environment are crucially

[1]See, e.g., Burgstahler and Chuk (2017), Burgstahler and Eames (2006), Dechow et al. (2003), Degeorge et al. (1999), and Leuz et al. (2003). Measures of small loss avoidance are among the proxies for earnings quality in the review of Dechow et al. (2010) in their Sections 3.1.4 and 3.1.5.

[2]For evidence in Japan, see Enomoto and Yamaguchi (2017).

important for financial reporting practices (Ball, 2016; Soderstrom & Sun, 2007).[3] As the principles-based IFRS provide greater discretion than the rules-based US GAAP (Ball, 2016, p. 553), it is quite possible that the practice of loss avoidance persists in a stable institutional environment. In addition, a possible SOX effect is likely to be smaller in Europe than the US because the regulatory changes were introduced later and were less far-reaching.

Based on these considerations, our first research question is whether the disappearance of the ZE discontinuity in the US is mirrored by a similar decline in Europe.

Having established the overall development, we look more closely at changes at the country level. We expect that the country differences in reporting small losses are related to more general cultural aspects of society as captured in the well-known framework of Hofstede et al. (2010). Among the dimensions distinguished by Hofstede et al. (2010), we consider the Uncertainty Avoidance Index (UAI) to be particularly relevant for this study, as it measures the extent to which members of a society are uncomfortable with uncertainty and ambiguity. Countries with a high UAI attempt to control the future and maintain a stable, predictable environment by applying rigid codes of belief and behavior. Reported losses could violate the perception of a stable, healthy and predictable firm and create discomfort among stakeholders because the firm seems to be in a delicate position. Perceived uncertainty could be much greater for firms reporting small losses than for firms reporting small profits. Therefore, managers in countries with high UAI have a particularly strong incentive to steer earnings into positive territory. Accordingly, we expect that there is a positive relationship between UAI and the size of the ZE discontinuity. Our second research question is whether this relationship is still intact or whether European countries have become more heterogeneous in terms of avoiding small losses.

Our contribution is, first, to analyze the changes in the ZE discontinuity in Europe over time in more detail than in prior studies, and second, to combine an in-depth analysis

[3] For the role of incentives, see, e.g., Ball et al. (2008), Beyer et al. (2010), Ewert and Wagenhofer (2005, 2012), Gassen et al. (2006), and Watts (2003).

of the temporal dimension with an in-depth country analysis. This type of study faces the problem of a limited number of observations at the zero earnings threshold per country per year. To mitigate this problem, we apply a modified discontinuity measure that is closely related to commonly applied measures but avoids weaknesses that are particularly relevant when the number of observations is small.

Our year-by-year analysis of earnings distributions shows that, unlike in the US, the ZE discontinuity in Europe only shows a slight downward trend. It has not disappeared as in the US, but is still substantial. The result is the same for matched industry samples, so differences in the industry composition between the US and Europe do not play a crucial role. While loss avoidance is pronounced in Europe as a whole, Europe is by no means a homogeneous entity in accounting practice. Our results suggest that small loss avoidance has a cultural dimension that can be captured by the UAI index of Hofstede et al. (2010). The relationship of our discontinuity measure to UAI is almost identical in the first and second halves of our sample period, suggesting that one reason for the stable ZE discontinuity in Europe as a whole is that it is rooted in an uncertainty avoidance culture.

The remainder of the paper proceeds as follows. Section 2 reviews the literature and develops our hypotheses. Sections 3 and 4 describe our discontinuity measures and data, respectively. Sections 5 presents our empirical results. We first compare year-by-year changes of the ZE kink in Europe and the US. We then show how the changes on the country level are related to the cultural dimension of uncertainty avoidance. Section 6 concludes the paper.

4.2 Prior Literature and Hypotheses

We only review the literature on the ZE discontinuity and leave out studies that apply discretionary accrual models. The distributional approach and the accrual approach rest on different assumptions that cannot easily be reconciled.[4] Our study is motivated by the disappearance of the ZE discontinuity in the US as reported by Gilliam et al. (2015), which is why we rely on the same approach.

Daske et al. (2006) provide important early evidence on discontinuities in Europe with respect to different targets, namely, targets to report positive earnings, to achieve positive earnings changes and to outperform analyst consensus forecasts (EU sample for 1986 to 2001). Importantly, "[t]hese discontinuities are much more pronounced in the EU compared to the US" (p. 137). Within Europe, the discontinuity in the UK is similar to the US, while it is more pronounced in code law countries. The differences across Europe prevail "despite the various EU harmonization efforts that have taken place" (p. 137). The results are less clear in Glaum et al. (2004) who find that firms in Germany and the US show a similarly strong tendency to avoid small losses in the period from 1991 to 2000. For an EU sample from 1997 to 2003, Burgstahler et al. (2006) find higher levels of earnings management in weak legal systems and in private firms. For UK firms, Gore et al. (2007) document a link between discontinuities and potentially managed working capital accruals. Haga et al. (2017) study a sample of 47 countries from 2003 to 2015 and show that the extent of the ZE discontinuity is positively associated with the long-term orientation of a country, which is one of the cultural dimensions of Hofstede et al. (2010). Han et al. (2010) and Nabar and Boonlert-U-Thai (2007) show a similar association with the Uncertainty Avoidance dimension of Hofstede et al. (2010), but for the accrual-based approach.

To isolate the IFRS effect, some studies compare earnings distributions for short periods before and after the adoption of IFRS. Barth et al. (2008) find that voluntary IFRS adopters in the EU manage earnings less, which suggests that early adopters are willing to be more transparent. The results of studies on mandatory adoption of IFRS are mixed. H. Chen et al. (2010) confirm the prior results of Barth et al. (2008), while Capkun et al. (2016) conclude that IFRS provide greater flexibility that has led to more income smooth-

[4] A disadvantage of using accrual models for our purpose of examining the avoidance of small losses is that the decomposition of total accruals into normal and discretionary accruals is associated with high standard errors that make it almost impossible to recognize a small step over the ZE threshold.

ing. In the same vein, Jeanjean and Stolowy (2008), Callao and Jarne (2010) and Lin et al. (2012) find that earnings management did not decline and in some countries even increased after the introduction of IFRS.

In a more recent study, Trimble (2018) examines the zero earnings discontinuity in a sample of 46 developed and developing countries from 1997 to 2013. She compares pre and post IFRS distributions for (1) EU and non-EU samples, (2) low and high enforcement samples and (3) common and code law samples. The results show that the discontinuity decreases in both EU and non-EU countries. The decline is more pronounced in countries with strict enforcement and high demands for high quality reporting.

The studies cited above pool data across several years to ensure a sufficient number of observations, or they focus on a short period around IFRS adoption. In comparison, our first objective is to examine the development of the ZE discontinuity over time in a more comprehensive and detailed manner. Specifically, we track the kink in a year-by-year design from 1988 to 2019, contrasting results for the US and Europe. We hypothesize that incentives for avoiding small losses generally persist in Europe:

Hypothesis 1: The ZE discontinuity is stable and still substantial in Europe, while it has declined and eventually disappeared in the US.

Our second objective is to provide results on the individual country level and to examine the association of the kink to the UAI of Hofstede et al. (2010). Callao and Jarne (2010) also analyze the kink for individual countries, but only in a short period around IFRS adoption. To ensure a sufficient number of observations near the ZE threshold, the authors have to choose a wide interval width of 4 percentage points ROA, which means that the definition of *small* losses is different than in previous literature. We provide country-level results for standard interval widths, which is possible because we do not study the impact of a specific event but long-term changes. We postulate that loss avoidance is rooted in the cultural UAI dimension, which may explain the observed stability in Europe as a whole:

Hypothesis 2: European countries with a high Uncertainty Avoidance Index show a more pronounced ZE discontinuity, and the association persists in recent years.

4.3 Discontinuity Measures

The most commonly used discontinuity measure, the standardized differences test statistic, is based on the frequency distribution of scaled earnings for a specific interval width. Let i denote ordered intervals such that $i = 1$ is the interval of the lowest profits, $i = 2$ is the interval for the next range of profits, $i = -1$ is the interval of the lowest losses and $i = -2$ is the next loss interval. Additionally, let N_i denote the number of observations in interval i, let N denote the total number of observations in the sample and let $p_i = N_i/N$ denote the proportion of observations falling in interval i. The standardized differences test statistic is then defined as the standardized difference between the actual number of observations in interval i and the expected number of observations assuming no discontinuity (Burgstahler & Dichev, 1997):

$$SD_i = \frac{p_i - 0.5(p_{i-1} + p_{i+1})}{s_i}, \tag{4.1}$$

where s_i is the standard error of the difference:

$$s_i = \frac{1}{\sqrt{N}}\sqrt{p_i\,(1 - p_i) + 0.25\,(p_{i-1} + p_{i+1})\,(1 - p_{i-1} - p_{i+1})}. \tag{4.2}$$

In the presence of a ZE discontinuity, SD_{-1} will be negative and SD_1 will be positive.

A disadvantage of the SD statistic is that it increases as the sample size increases. This blending of the effect size and test power is undesirable in comparisons over time when the sample size varies. Specifically, when the sample size is small in early years, the SD statistic will underestimate the potential decline in the discontinuity. Another critical aspect is that the expected frequency is defined as the average frequency of the adjacent intervals. When these in turn are distorted, the statistic is difficult to interpret.

For these reasons, we use the modified measure proposed by Chardonnens et al. (2022), which is based on a kernel density estimation inspired by Lahr (2014). We choose a standard Gaussian estimator for scaled earnings ranging from −0.15 to 0.15.[5] The measure "small loss deviation" is defined as:

$$SLD = \frac{Actual_{-1} - Expected_{-1}}{Expected_{-1}}, \tag{4.3}$$

where $Actual_{-1}$ is the actual number of observations in the first loss interval and $Expected_{-1}$ is the expected number according to the kernel density estimation (integral over the first loss interval). The "small profit deviation" is analogously defined as:

$$SPD = \frac{Actual_{1} - Expected_{1}}{Expected_{1}}, \tag{4.4}$$

where subscript 1 represents the first profit interval.

Following previous literature, we consider interval widths of 0.005 and 0.015. In the case of 0.015, the actual observations of the first three loss and profit intervals of width 0.005 are compared to the expected value according to the kernel density over the same range.

Figure 4.1 shows examples of frequency distributions and kernel densities (blue lines) for the US (left column) and Europe (right column) and reports the corresponding small loss deviations (SLD) and small profit deviations (SPD).

[5] The density is estimated at 512 equally spaced points. We implement the estimation in R ("density" function) with the bandwidth proposed by Scott (1992) (option bw.nrd in R).

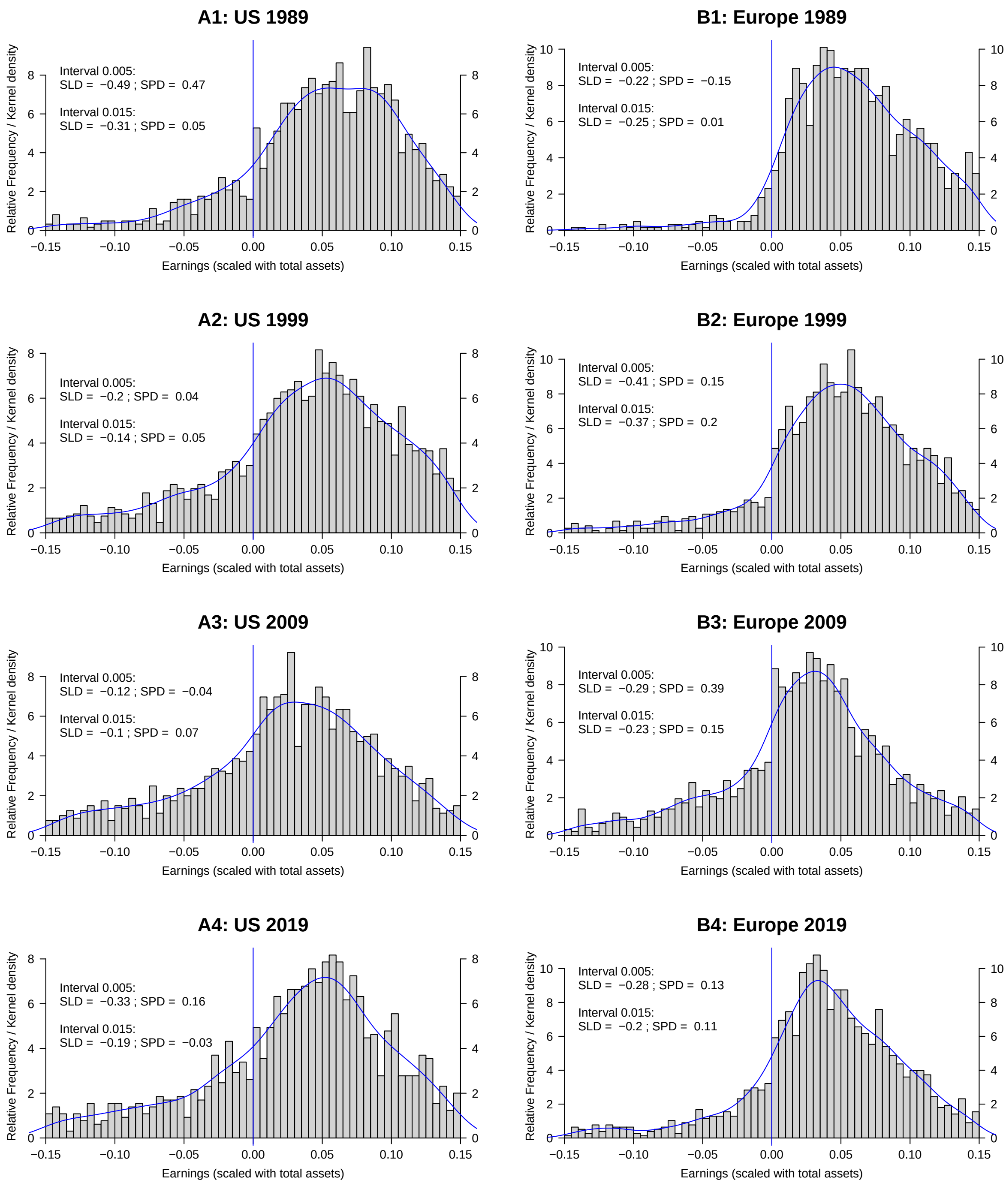

Figure 4.1: Frequency Distributions and Kernel Densities for Scaling with Total Assets. The distributions are shown for four years spread over the sample period. Left graphs: sample of US firms; right graphs: sample of European firms. Vertical line at zero scaled earnings. Earnings are scaled with total assets at the beginning of the year. One bar in the diagram represents a width of 0.005. *SLD*: small loss deviation; *SPD*: small profit deviation.

4.4 Data and Descriptive Statistics

We use Worldscope data from Refinitiv Datastream (formerly Thomson Reuters Datastream) from 1988 to 2019 for the US and the EU extended to the United Kingdom (UK) and the EFTA states Iceland, Liechtenstein, Norway and Switzerland. We exclude firms operating in regulated industries and financial institutions with Standard Industrial Classification (SIC) codes ranging from 4400 to 5000 and from 6000 to 7000.[6] Following previous research, we remove observations with a net income of exactly zero (58 firm-years in the European sample and 13 firm-years in the US sample).[7] We also remove firm-years with insubstantial sales, which we define as total sales of less than 2 million USD. Since these firms report losses by definition, they could distort our loss avoidance measures (see Chardonnens et al., 2022).

We also apply a filter for firms with negligible market capitalization. In finance practice, stocks with a market capitalization value of less than 50 million USD are classified as “nano caps” or “penny stocks” (compared to micro caps, small caps, large caps and blue chips). These stocks are typically traded in OTC markets and are considered highly speculative investments not only in terms of return fluctuations but also in terms of potential market manipulations. Penny stocks are numerous in the Worldscope database because of its broad coverage of OTC markets, but they are not of interest to typical institutional or private investors. Therefore, we require a market capitalization level in 2019 of 50 million USD and deflate this number by 3% per year such that the minimum requirement in 1988 is 20 million USD.[8] To ensure a minimum size regardless of market fluctuations, we apply the same threshold (20 million USD in 1988 to 50 million USD in

[6]See similarly, e.g., Beaver et al. (2007); Brown and Caylor (2005); Burgstahler and Dichev (1997); S. K. Chen et al. (2010); Durtschi and Easton (2005); Durtschi and Easton (2009); Gilliam et al. (2015); Haga et al. (2017); Kerstein and Rai (2007); Roychowdhury (2006); Makarem et al. (2018).

[7]See Burgstahler and Dichev (1997); Gilliam et al. (2015); Dechow et al. (2003); Beaver et al. (2007); Burgstahler and Eames (2006); Lahr (2014).

[8]A value of 50 million USD in 2019 corresponds approximately to the 0.5th percentile of the market capitalization of NYSE stocks. Fama and French (2008) used the 20th NYSE percentile to define micro caps. This shows that we still include very small firms.

2019) to total assets. The remaining firms represent more than 99% of the overall market capitalization and 69.3% (74.6%) of firm-years in the European (US) sample.

Our final filter is to remove firm-years with an absolute value of net income (scaled by total assets) or operating cash flow less depreciation and amortization (scaled by total assets) higher than 50%. This removes 6.9% of firm-years from the European sample and 3.5% of firm-years from the US sample. It is important to note that this outlier correction does not affect our empirical analysis of the ZE discontinuity but only serves to report meaningful descriptive statistics in Table 4.1. Our final sample contains 65,052 firm-years for the US and 56,855 firm-years for Europe. US firms tend to have a larger market capitalization (MCAP) than European firms. The difference is smaller for total assets, and the distributions of earnings are similar in both samples.

Statistic	N	Mean	SD	Min	P25	Median	P75	Max
				US				
MCap	65,052	4,171	22,041	20	141	473	1,818	1,304,756
Total Assets	65,052	3,182	16,924	20	143	427	1,615	797,800
NI	65,052	0.035	0.126	−0.500	−0.001	0.050	0.098	0.499
				Europe				
MCap	56,855	2,550	11,648	20	93	258	988	311,748
Total Assets	56,855	3,243	14,900	20	131	354	1,333	532,474
NI	56,855	0.049	0.091	−0.499	0.015	0.048	0.089	0.500

Table 4.1: Descriptive Statistics for the US and European Samples. N: number of firm-years; $MCap$: market capitalization; NI: net income scaled by total assets at the beginning of the year. $MCap$ and Total Assets in million USD.

4.5 Empirical Results

4.5.1 Year-by-Year Development of the ZE Discontinuity

Figure 4.2 shows the yearly estimates of SLD (blue crosses) and SPD (black circles) for the sample of US (left panels) and European firms (right panels). The upper and lower panels are based on interval widths of 0.005 and 0.015, respectively. Appendix Table B.1 reports the exact values. In the US sample, the ZE discontinuity for an interval width of 0.005 (Panel A1 in Fig. 4.2) is very pronounced at the beginning of the sample period, but it diminishes in the following years. From 2004, the lines for SLD and SPD fluctuate around the same level. For the interval of 0.015, we observe the same decline until the kink disappears in approximately the year 2007.

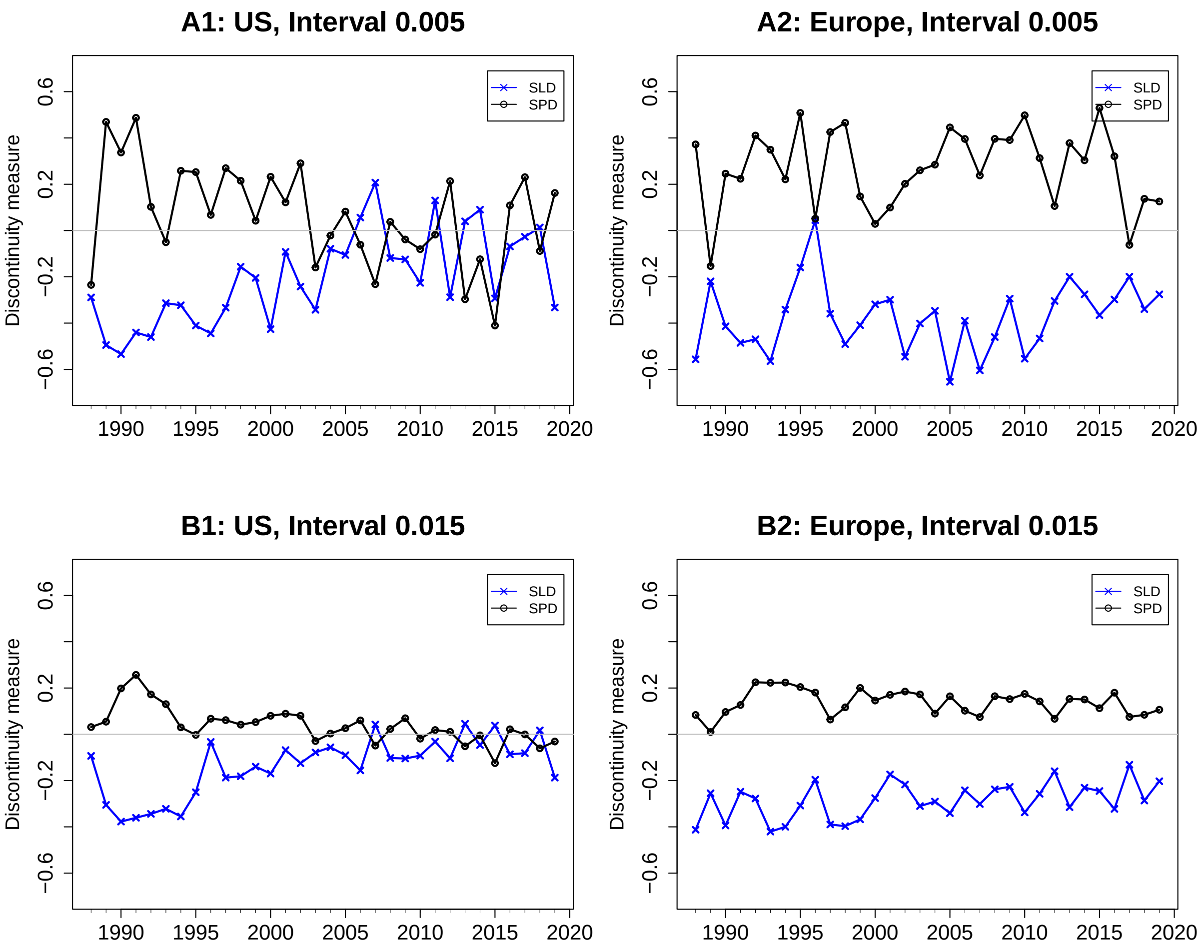

Figure 4.2: Discontinuity Measures for the US and Europe for Scaling With Total Assets. The discontinuity measures capture the share of excess observations (pos. sign) or missing observations (neg. sign) in the intervals of scaled earnings directly below and above the zero threshold. SLD: small loss deviation; SPD: small profit deviation. Earnings are scaled with total assets at the beginning of the year.

The results are strikingly different in the European sample (Panels A2 and B2). The gap between the upper line for SPD and the lower line for SLD signifies a pronounced ZE discontinuity over the whole sample period. On average, 31% of the expected number of small losses for interval 0.015 is missing. The gap tends to be even more pronounced for the interval of 0.005 (Panel A1 compared to Panel B1), where on average, 38.6% of expected small losses are missing. Appendix Table B.1 confirms the statistical significance of these results: The standardized difference SD_{-1} for interval 0.015 is significantly negative in every year and SD_1 is significantly positive in most years. Therefore, compared to the US, the kink in the European sample is remarkably stable over time. We note, however, that the blue line for SLD in Panel B2 (interval 0.015) shows a slightly increasing trend. A linear simple regression of SLD on time yields a slope coefficient of 0.0040 (significant at the 1% level), which means that SLD increases from a fitted value of -0.35 in 1988 to -0.23 in 2019. For the interval 0.005 (Panel A2), the corresponding slope coefficient is 0.0024, which is insignificant.

Our results remain confirmed when earnings are scaled by the market value of equity instead of by total assets. For this alternative scaling, we replicate Figure 4.2 in Appendix Figure B.1. To address the possibility that our results are driven by different industry structures of our US and European samples (e.g., Gaio, 2010), we follow Ball and Shivakumar (2005) and construct a matched sample based on the first two digits of the SIC code and market capitalization.[9] For each year-industry combination, we first identify the subsample with the smaller number of observations (the US or Europe). We match these observations with observations from the other subsample (without replacement) based on market capitalization using the nearest neighbor method. Over the whole sample period, 46,279 observations of both subsamples are matched, which means that 18,773 US firm-years and 10,576 European firm-years are discarded.

Figure 4.3 shows the discontinuity results for the matched sample (interval width 0.015). The results are very similar to the previous results for the full sample: In Europe, the ZE discontinuity is present over the whole sample period, while it has disappeared in the US.

We conduct the following further robustness checks. To test whether our results are driven by stocks with very small market capitalization ("micro caps"), we double the

[9]Matched samples are commonly used in prior literature; see, e.g., Barth et al. (2008).

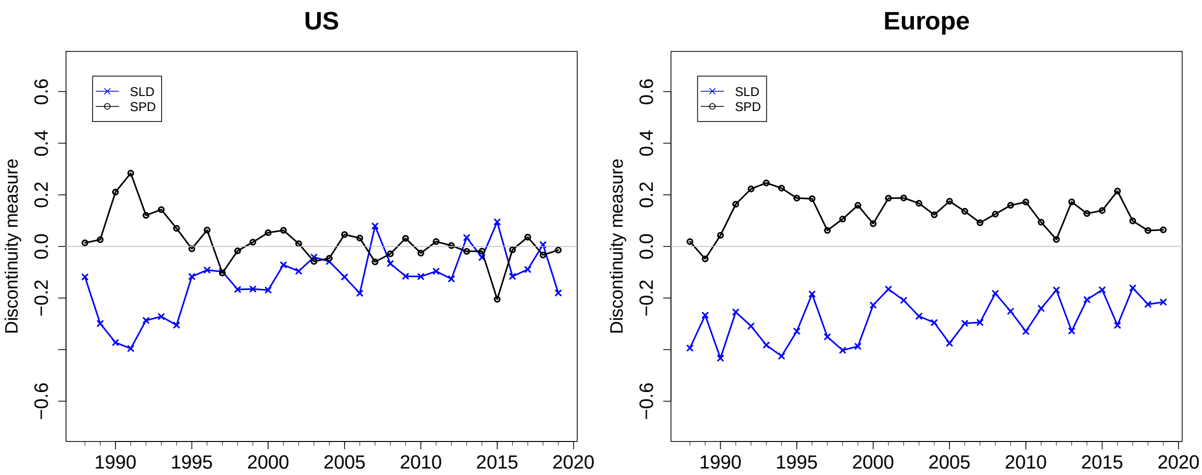

Figure 4.3: Results for Matched Industry Samples. Discontinuity measures analogous to Panels B1 and B2 in Fig. 4.2 (interval width 0.015). SLD: small loss deviation; SPD: small profit deviation.

minimum market capitalizations (20 to 50 million USD; see Section 4.4) and obtain similar results. The same result is found when using net income *before extraordinary items* as our earnings measure.

4.5.2 Cross-Country Analysis of Loss Avoidance and its Relation to the Uncertainty Avoidance Index

For the cross-country analysis, we divide the sample period into two subperiods of equal length (1988-2003; 2004-2019) and pool the country-level observations in each subperiod into one group for which we measure the small loss deviation SLD using an interval width of 0.015. Figure 4.4 compares the country values for SLD in the two subperiods.[10] The positive correlation indicates continued loss-avoidance behavior. The regression of SLD in period 2 on SLD in period 1 provides a slope coefficient of 0.76 (significant at the 1% level) and an R^2 of 0.458 (solid line in Figure 4.4). Consistent with our previous results, Europe as a whole is positioned almost on the 45 degree line, while the US is clearly above, which means that SLD in the US declined in the second period from an already

[10] Appendix Table B.2 reports the exact values.

low level of the first period.[11]

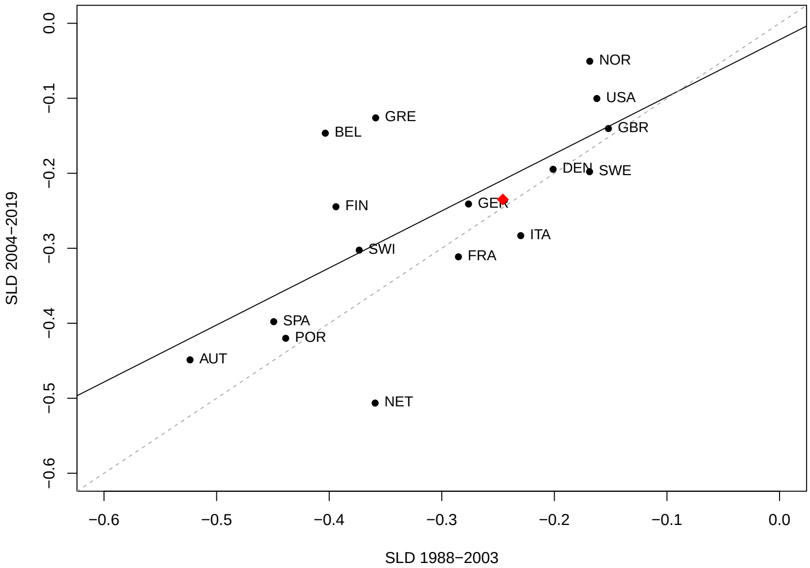

Figure 4.4: Comparison of Small Loss Avoidance in the First and Second Halves of the Sample Period. SLD is the small loss deviation for an interval width of 0.015. The solid line is the regression line, the dashed line is the 45 degree line. The red diamond indicates the weighted average of the European countries (with the number of observations as weights). The country codes are defined in Appendix Table B.2.

A regression analysis of the relationship between UAI and SLD requires a weighted least squares (WLS) estimation because the precision of discontinuity measures depends on the sample size. For the standardized difference measure SD, this can be seen from Eq. (4.2) which shows that the variance s_i^2 is proportional to $1/N$, so the inverse of the variance is proportional to N. It is reasonable to assume that this is also a good approximation for our modified kernel-based measure SLD. Therefore, we choose weights corresponding to the number of available observations in each country. For information, we also show the results of OLS regressions.

In the first period, SLD is negatively related to UAI (left graph in Figure 4.5). This means that small loss avoidance tends to be more pronounced (SLD more negative) where uncertainty avoidance is strong. The WLS and OLS slope coefficients are both significant

[11] Because of the averaging over time, the downward trend in the US is less visible here than in Section 4.5.1. In particular, the average US value of SLD in the first period is already affected by the downward trend.

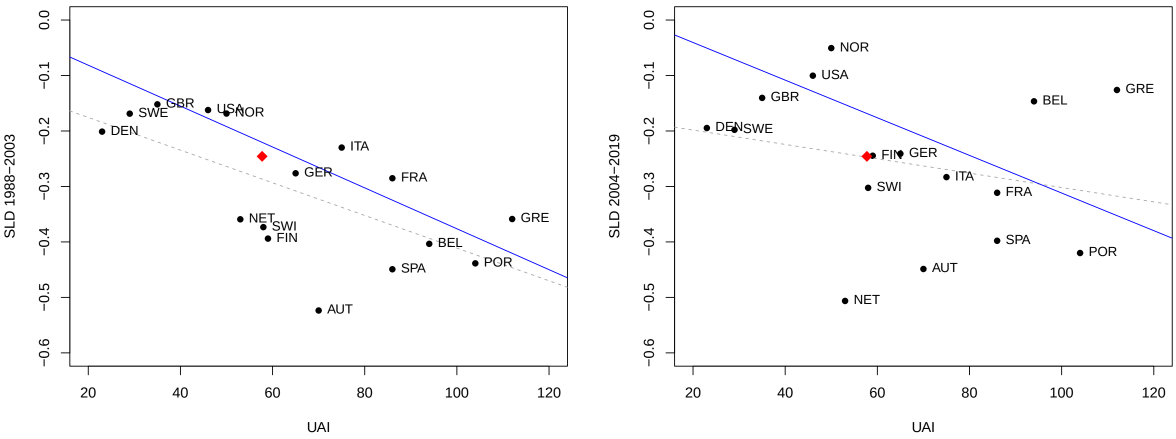

Figure 4.5: Relationship Between Small Loss Avoidance and UAI. The left graph covers the first half (1988–2003), the right graph covers the second half of the sample period (2004–2019). *SLD* is the small loss deviation for an interval width of 0.015 based on pooled data of the respective period. *UAI* is the Uncertainty Avoidance Index. The blue solid line (black dashed line) is the WLS (OLS) regression line. The red diamond indicates the weighted average of the European countries (with the number of observations as weights). The country codes are defined in Appendix Table B.2.

at the 1% level (columns (1) and (2) of Table 4.2). As far as the origin of local accounting is concerned, a general trend can be identified from the left graph in Figure 4.5: Countries with British (UK, US) or Scandinavian accounting origin (DEN, FIN, NOR, SWE) tend to have low *UAI* and low loss avoidance, countries from the German accounting system (AUT, GER, SWI) tend to be found in the middle, and countries associated with the French accounting system tend to have high *UAI* and a strong ZE discontinuity (BEL, FRA, GRE, ITA, NET, POR, SPA). On average, the European countries have a higher *UAI* and show a stronger discontinuity than the US, which is consistent with the results in Section 4.5.1 of this paper (see the red diamond in Figure 4.5 for Europe as a whole).

In the second period, the WLS regression slope is almost the same as in the first period, which suggests that the relationship between *UAI* and *SLD* has remained stable (column (4) of Table 4.2). However, the country-specific *SLD* measure is more dispersed in the second period (right graph in Figure 4.5), and the OLS slope coefficient is negative but insignificant. The difference between OLS and WLS is due to the fact that the measured kink has changed significantly in some small countries, in particular in Belgium and Greece. These two countries had a pronounced kink of -0.4 and -0.36, respectively, in the first period and a much smaller kink in the second period (-0.15 and -0.13, respectively), which are by far the largest changes in the sample. Without Belgium and Greece, the results in the two subperiods would be practically identical also in the OLS regression (slope coefficients of -0.0036 in the first period and -0.0035 in the second period). Because of the small number of observations near the zero earnings threshold, it is difficult to assess whether the changes in Belgium and Greece are real or are partly due to measurement error. This again shows why we put more emphasis on the WLS estimates.

To formally test for differences between the two periods, we estimate the following combined regression model:

$$SLD = \alpha_1 + \alpha_2 \cdot Sub2 + \beta_1 \cdot UAI + \beta_2 \cdot UAI \cdot Sub2 + \varepsilon, \tag{4.5}$$

where α_1, α_2, β_1 and β_2 are regression coefficients and $Sub2$ is a dummy variable equal to 1 for observations from the second period and 0 otherwise. Thus, the interaction-term coefficient β_2 indicates the change in the slope coefficient in the second period compared to the first period.

In the first period, the negative relation to UAI (coefficient β_1) is confirmed by the OLS and WLS estimations (columns (5) and (6) of Table 4.2). Coefficient β_2 is insignificant in both cases, so that we cannot reject the null hypothesis that the relation between SLD and UAI is the same in both periods. In fact, the WLS estimate of β_2 is almost zero. However, the WLS regression suggests that the average kink (intercept at a UAI level of 50) has declined from -0.1920 to -0.1431, which is statistically significant but still implies a substantial average kink in the second period.

	1988–2003 (Sub1)		2004–2019 (Sub2)		1988–2019	
	OLS	WLS	OLS	WLS	OLS	WLS
	(1)	(2)	(3)	(4)	(5)	(6)
Interc	−0.2639***	−0.1920***	−0.2372***	−0.1422***	−0.2639***	−0.1920***
	(−9.64)	(−16.52)	(−6.06)	(−7.75)	(−7.82)	(−12.51)
UAI	−0.0029***	−0.0037***	−0.0012	−0.0034**	−0.0029**	−0.0037***
	(−3.19)	(−4.23)	(−0.98)	(−2.73)	(−2.59)	(−3.37)
Interc : Sub2					0.0267	0.0489**
					(0.56)	(2.25)
UAI : Sub2					0.0017	0.0002
					(1.03)	(0.10)
N	16	16	16	16	32	32
Adj. R^2	0.380	0.529	−0.002	0.300	0.176	0.435
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 4.2: Results of Cross-Country Regressions. The dependent variable is SLD for an interval width of 0.015. UAI is the Uncertainty Avoidance Index less 50 (so that the regression intercept corresponds to the fitted SLD value at UAI of 50). Sub2 is a dummy variable equal to 1 for the second subperiod and 0 otherwise. Weighting in the WLS regression is by the number of observations per country. The regression model for columns (5) and (6) is specified in Eq. (4.5).

4.6 Conclusion

Our year-by-year analysis of the zero earnings (ZE) discontinuity shows a striking difference between the US and Europe: While the discontinuity in the US has steadily declined and finally disappeared, it has only slightly decreased in Europe and is still substantial. Our results suggest that this difference is rooted in cultural underpinnings of loss-avoidance behavior, which we capture using the Uncertainty Avoidance Index (UAI) of Hofstede et al. (2010). Countries with a higher UAI index tend to show a more pronounced ZE discontinuity. This relationship is almost identical in the first and second halves of our sample period. Among small countries, the dispersion of the ZE discontinuity is larger in the second subperiod, but the change in the relationship with UAI is insignificant. Overall, our results support the view that despite harmonization of international accounting, global competition and SOX-related regulations, there is still room for different degrees of loss-avoidance behavior in different European countries, related to cultural aspects of dealing with uncertainty.

We point out three limitations of our study. First, the number of observations near the zero threshold is small for some countries, which adversely affects the precision of the estimates. We address this difficulty by a combination of pooling annual data, running WLS regressions and using a kernel density-based discontinuity measure. Second, changes over time are not available for the UAI index. While cultural dimensions are thought to be relatively stable, changes may have occurred in some countries during the long period of more than 30 years covered in this paper. The third limitation is that the cultural UAI dimension is related to the origin of local accounting (British, French, German or Scandinavian), the legal system (common law versus code law) and the extent of investor protection. For example, the US and the UK have low UAI values, their local accounting is of British origin, and both are common law countries with strong investor protection. Given this overlap, it is impossible to disentangle the effect of purely cultural factors from the effects of other related characteristics of the legal and institutional environment. Therefore, we cannot make a causal statement about the effect of UAI. Nevertheless, we believe that our results contribute to a better understanding of the different evolution of the ZE discontinuity in Europe and the US and the heterogeneity of European countries in avoiding small losses.

Chapter 5

EARNINGS MANAGEMENT MODELS AND ACCOUNTING FRAUD PREDICTION

5.1 Introduction

Aggregate earnings distributions are kinked at the zero earnings threshold because small profits occur much more frequently than small losses. The best explanation for this phenomenon seems to be that firms tend to use their accounting discretion to turn small losses into reported profits. This distribution is very illustrative at the aggregate level, but it has the disadvantage that the overall extent of earnings management cannot be attributed to the firm level. Some firms reporting small positive earnings need to have managed their earnings upward, although not all may have engaged in upwards earnings management. For many research questions, however, precisely such firm-level indicators are needed.

The most commonly used measure at the firm level is the discretionary accrual based on the (modified) Jones model. In Jones-type models, "normal" accruals are estimated using a regression approach based on firm-specific financial statement positions that are assumed to be unrelated to earnings management. The unexplained portion of total accruals is considered discretionary and includes the effects of earnings management. One fundamental difficulty with this and similar approaches is that their reliability and validity are unknown and difficult to assess. It remains unclear whether the allocated discretionary accruals truly arise from earnings management or rather from misspecifications of the earnings management model.

Against this background, this paper aims to investigate the validity of firm-level measures for different earnings management models. Two are typical Jones models, while the third is an attempt to extract information on earnings management from the kinked aggregate earnings distribution. We examine cases in which firms are known to have manipulated earnings. These cases stem from the Accounting and Auditing Enforcement Releases (AAER) by the U.S. Securities and Exchange Commission (SEC) that contains firms found guilty of accounting fraud.

We compare the validated fraud firms with non-fraud firms and evaluate the performance of different earnings management models in detecting these fraud firms. The first is the modified Jones model, which serves as the benchmark model in this study. The second measure comes from an extended version of the Jones model. It is known that the discretionary accruals of the modified Jones model are positively related to the level of earnings, which could be a problem because this relationship could indicate that the discretionary accruals measure is biased by firm performance (Dechow et al., 2003). Therefore, we extend the modified Jones model to include earnings information that removes the earnings-related component from discretionary accruals. The third model attempts to derive a measure of firm-level earnings management from the kinked earnings distribution. In the first step, we estimate the shape of the theoretical earnings distribution that is not affected by earnings management. In the second step, we match the firms to this distribution in the order of reported earnings. This shifts the shape of the earnings distribution from kinked to neutral without changing the order of the firm's earnings ranks. The procedure is similar to standardizing a variable. The firm-level difference between earnings according to the neutral distribution and earnings according to the kinked distribution then serves as our earnings management measure. We define these three measures based on two definitions of accruals: working capital accruals (WCA) and comprehensive accruals (CompAcc). Thus, we consider a total of six measures.

We confront the six measures of earnings management with the fraudulent earnings manipulations in the AAER sample in two different ways. Our first approach is to determine the extent of earnings management for each firm in the AAER fraud sample. Accounting fraud does not necessarily mean that the magnitude of manipulation is large. Therefore, it is also of interest to determine if the values for the fraud firms are at least positive, if not extreme. We test whether the mean of these values is significantly positive, indicating the presence of earnings management. Second, we predict that the 1% highest discretionary accruals within each earnings region of 5% of each method's total assets are included in the AAER fraud sample. This is based on the idea that fraud is an extreme case of earnings management that should most likely correspond to the most extreme observations of a valid measure. We form the earnings regions because abnormal earnings have extreme discretionary accruals. The models assign a high predictive frequency to these high negative or positive earnings. Finally, we count these firm-years and compute the sensitivity and the precision.

The intended contribution of this study is to provide an in-depth analysis of discretionary accruals for the AAER fraud sample. We go beyond the modified Jones model by including promising extensions and alternative models in our comparison. We find that the extended Jones model, which includes earnings, leads to increased explanatory power and, more importantly, prevents earnings from correlating with discretionary accruals. The distribution model, in which the extent of supposed earnings management is inferred from earnings distributions, shows the most prominent indication of upward earnings management for firms reporting a small loss. The modified and extended Jones models provide significantly higher discretionary accruals for WCA and CompAcc for fraudulent firms, which is not the case with the distributional model. Our second test is to predict the highest 1% of discretionary accruals within each earnings region of 5% of total assets to figure in the AAER sample. It shows that the regression-based Jones-type models correctly predict slightly more overstated firm-years than it would be the case if they were randomly selected. However, the differences are minor, so they may arise due to coincidence.

The remainder of this chapter is structured as follows. Section 5.2 reviews the literature. In Section 5.3, we present the different earnings management models included in this study. Section 5.4 presents the evaluation methodology. In the empirical part, we describe the data and the properties of the six models (Section 5.5) and, finally, the results of predicting membership in the AAER sample based on these models (Section 5.6).

5.2 Literature

This section briefly presents the relevant literature on accruals and earnings management models. They estimate each firm's expected level of accruals, which is referred to as normal accruals. This estimate is obtained by regressing reported total accruals on financial accounting items such as sales, receivables, or property, plant & equipment (PPE). Deviations from the total level of accruals are defined as the discretionary portion, which would indicate some upward or downward earnings management of the firm.

5.2.1 Accruals

In short, accruals are required in accounting to record events that affect the income statement results but not the cash of a period. A more formal definition of accruals is provided in the Statement of Financial Accounting Concepts No. 6 by the Financial Accounting Standards Board (Financial Accounting Standards Board, 1985, p. 72).

> "By accounting for noncash assets, liabilities, revenues, expenses, gains, and losses, accrual accounting links an entity's operations and other transactions, events, and circumstances that affect it with its cash receipts and outlays."

In an academic context, the computation of accrual is not uniform. Depending on the research question, different accrual calculations may be appropriate. Larson et al. (2018) provide an overview of previous accrual definitions of academic studies. We briefly introduce two of these definitions. The first, WCA, could be considered the narrowest definition because it considers only short-term accounting items. It represents the change in current assets excluding cash less current liabilities excluding short-term debt:

$$WCA_{i,t} = \Delta CA_{i,t} - \Delta Cash_{i,t} - (\Delta CL_{i,t} - \Delta STD_{i,t}), \tag{5.1}$$

where $WCA_{i,t}$ are the annual changes in WCA for firm i in year t divided by $A_{i,t-1}$;
$\Delta CA_{i,t}$ = Change in current assets for firm i in year t divided by $A_{i,t-1}$;
$\Delta Cash_{i,t}$ = Change in cash for firm i in year t divided by $A_{i,t-1}$;
$\Delta CL_{i,t}$ = Change in current liabilities for firm i in year t divided by $A_{i,t-1}$;
$\Delta STD_{i,t}$ = Change in short-term debt for firm i in year t divided by $A_{i,t-1}$;
$A_{i,t-1}$ = Total assets for firm i in year $t-1$.

Many studies use the balance sheet approach to compute WCA because its items are frequently available for many firms, which avoids a large reduction in the sample (Baber et al., 2011; DeFond & Park, 2001). Some studies extend the WCA definition to include depreciation from the income statement (Beneish & Vargus, 2002; Desai et al., 2004; B. I. Lev & Nissim, 2006). Others calculate accrual using the cash flow approach (Dechow & Dichev, 2002; McNichols, 2002).

Larson et al. (2018) and Richardson and Sloan (2005) define accruals more comprehensively, where accruals consist of the change in common shareholder's equity less the change in cash. Therefore, the second accrual computation method, CompAcc, is:

$$CompAcc_{i,t} = \Delta CE_{i,t} - \Delta Cash_{i,t}, \tag{5.2}$$

where $CompAcc_{i,t}$ are the annual changes in CompAcc for firm i in year t divided by $A_{i,t-1}$;

$\Delta CE_{i,t}$ = Change in shareholder's equity for firm i in year t divided by $A_{i,t-1}$;

$\Delta Cash_{i,t}$ = Change in cash for firm i in year t divided by $A_{i,t-1}$;

$A_{i,t-1}$ = Total assets for firm i in year $t-1$.

Following their idea, non-cash assets and liabilities might contain earnings management. In other terms, only cash is expected to be free of earnings management. Similar to WCA, the definition is data sparse because few accounting items are needed.

5.2.2 Earnings Management Models

The need for earnings management models emerges because accrual bookings sometimes require managerial judgment. Managers may or may not face the incentive to use their discretion to steer earnings towards a more desirable level. To obtain a firm- and year-specific estimate of the amount managed, academic research estimates a firm's total accruals that would be expected if reported earnings were not managed, referred to as normal accruals. The difference between total reported accruals and normal expected accruals is referred to as discretionary accruals.

Figure 5.1 illustrates the concept of discretionary accruals for the case of one explanatory variable. The observed total accruals in blue are regressed on the explanatory variables of the earnings management model. The fitted values represent the normal accruals expected in the absence of earnings management in orange. The residuals (i.e., the unexplained deviation) are then defined as discretionary accruals in green.

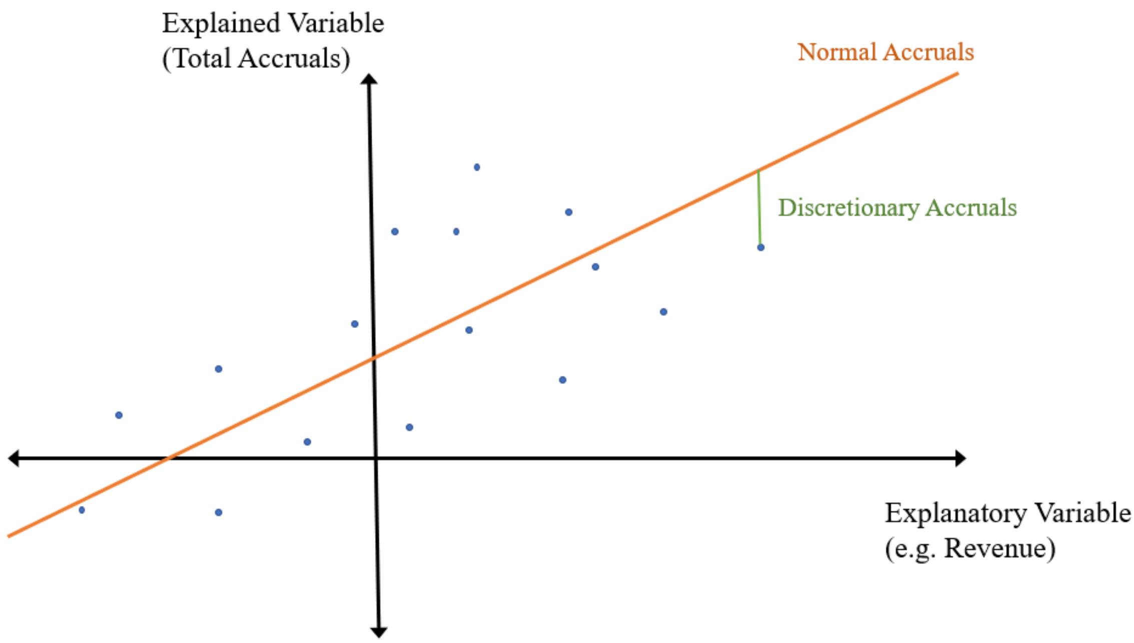

Figure 5.1: Concept of Earnings Management Models. The figure visualizes the concept of earnings management models and discretionary accruals. The x-axis represents independent (explanatory) regression variables such as earnings, revenues, or receivables. The y-axis represents the dependent (explained) variable on which the regression was run. The orange line corresponds to the fitted values of the estimated regression. The discretionary accruals in green correspond to the residuals, measured as the difference between the fitted (expected) and the observed total accruals.

While typical earnings management models follow the above concept of labeling the unexplained variation as discretionary accruals, they use different explanatory variables. In the following, we present the most commonly used models.

DeAngelo (1986) averages accrual variation over several years and defines a firm specific level of normal accruals. The difference between total observed accruals and estimated normal accruals results in discretionary accruals. A large change in discretionary accruals could therefore detect a firm managing earnings. However, the model of DeAngelo (1986) is subject to criticism. Accruals depend on a firm's economic parameters, which may change over time but are unrelated to earnings management. For example, nondiscretionary accruals correlate positively with revenue. Therefore, revenue growth leads to increased accruals and not necessarily to earnings management (Kaplan, 1985). For this reason, subsequent models aim to account for firm growth.

The Jones (1991) model regresses total accruals on the change in revenue and total PPE, both scaled by the last year's total assets:

$$Accruals_{i,t} = \alpha + \beta_1 \Delta Rev_{i,t} + \beta_2 PPE_{i,t} + \epsilon_{i,t}, \tag{5.3}$$

where $Accruals_{i,t}$ = Accruals for firm i in year t divided by $A_{i,t-1}$;
$\Delta Rev_{i,t}$ = Change in revenues for firm i in year t divided by $A_{i,t-1}$;
$PPE_{i,t}$ = Gross property, plant & equipment for firm i in year t divided by $A_{i,t-1}$;
$A_{i,t-1}$ = Total assets for firm i in year $t-1$.

The Jones (1991) model respects the changing firm metrics within certain limits. It includes the change in revenues, which appears to be a reasonable growth indicator since higher revenues are associated with higher WCA such as receivables, inventories, and accounts payable. Jones (1991, p. 212) points out that revenues might partly be subject to earnings management, e.g., by shifting the revenue recognition period. The criticism of the Jones model is that revenue growth is closely related to growth in accounts receivables. An earnings management model that accounts for revenue growth should also directly account for the growth in accounts receivables. Therefore, Dechow et al. (1995) propose a modified version of the Jones model that incorporates the change of receivables to the previous Jones model. Their model (hereafter referred to as the modified Jones model) uses the following regression:

$$Accruals_{i,t} = \alpha + \beta_1 (\Delta Rev_{i,t} - \Delta Rec_{i,t}) + \beta_2 PPE_{i,t} + \epsilon_{i,t}, \tag{5.4}$$

where $Accruals_{i,t}$ = Accruals for firm i in year t divided by $A_{i,t-1}$;
$\Delta Rev_{i,t} - \Delta Rec_{i,t}$ = Change in revenues for firm i in year t minus change in accounts receivables for firm i in year t divided by $A_{i,t-1}$;
$PPE_{i,t}$ = Gross property, plant & equipment for firm i in year t divided by $A_{i,t-1}$;
$A_{i,t-1}$ = Total assets for firm i in year $t-1$.

The Dechow and Dichev (2002) model uses past, present, and future cash flows to explain accruals. The rationale is that current year cash flows are negatively correlated with nondiscretionary accruals, while lagged and forward cash flows are positively associated with nondiscretionary accruals. For example, when a large customer settles the invoice, current year cash flows increase, and total accruals decrease. This leads to the following regression model:

$$Accruals_{i,t} = \alpha + \beta_1 CF_{i,t-1} + \beta_2 CF_{i,t} + \beta_3 CF_{i,t+1}, \tag{5.5}$$

where $Accruals_{i,t}$ = Accruals for firm i in year t divided by $A_{i,t-1}$;
$CF_{i,t}$ = Cash flows for firm i in years $t-1$ to $t+1$ divided by $A_{i,t-1}$;
$A_{i,t-1}$ = Total assets for firm i in year $t-1$.

The McNichols (2002) model combines the explanatory variables of the Jones model and the Dechow and Dichev model and has increased explanatory power. McNichols (2002) argues that the relatively small explanatory power of prior earnings management models leads to an overestimation of the discretionary portion since it is likely that not all reported discretionary accruals are truly discretionary. The regression model is:

$$Accruals_{i,t} = \alpha + \beta_1 \Delta Rev_{i,t} + \beta_2 PPE_{i,t} + \beta_3 CF_{i,t-1} + \beta_4 CF_{i,t} + \beta_5 CF_{i,t+1} + \epsilon_{i,t}, \tag{5.6}$$

where $Accruals_{i,t}$ = Accruals for firm i in year t divided by $A_{i,t-1}$;
$\Delta Rev_{i,t}$ = Change in revenues for firm i in year t divided by $A_{i,t-1}$;
$PPE_{i,t}$ = Gross property, plant & equipment for firm i in year t divided by $A_{i,t-1}$;
$CF_{i,t}$ = Cash flows for firm i in years $t-1$ to $t+1$ divided by $A_{i,t-1}$;
$A_{i,t-1}$ = Total assets for firm i in year $t-1$.

The above earnings management models are still frequently used and form the group of the base models. Since then, studies have evaluated these models, identified some weaknesses, and developed some alternatives. We highlight three notable contributions.[1]

Dechow et al. (2012) show that the inclusion of WCA reversals increases the explanatory power of earnings management models. One criticism pointed out in earlier research is that the models are potentially prone to correlated omitted variables, which may be one of the reasons for the low explanatory power and low success in distinguishing discretionary from nondiscretionary accruals. By accounting for the fact that accruals must be reversed at some point, time series accrual measures are expected to be lower. In particular, WCA often reverse within the following year.

Ball and Shivakumar (2006) extend commonly used earnings management models by considering the asymmetric timeliness of positive and negative earnings. Basu (1997) first described the asymmetric timeliness, as firms tend to recognize losses more quickly than gains. To account for such an effect, Ball and Shivakumar (2006, p. 219) added two variables capturing cash flow information as a proxy for economic gains or losses. The first is a dummy variable that takes the value of 1 when cash flows are negative and 0 otherwise. The second is the interaction term of the dummy variable with the cash flows. This piecewise linear regression can be used to increase the explanatory power of existing earnings management models. Ball and Shivakumar (2006, p. 219) apply the modified Jones model to their data and arrive at an adjusted R^2 of 8.8%. Adding cash flow information increases the adjusted R^2 to 16.8%.

Kothari et al. (2005) develop a performance-matched model for cross-sectional regressions. The concept consists of matching firms based on their return on assets within different groups based on the two-digit Standard Industrial Classification (SIC) code and year. The model relies on the assumption that firms within this group are comparable. Although the reasoning appears convincing and intuitive, the model has some drawbacks for which it has been widely criticized. First, separating the two-digit SIC codes may result in very few observations in some industries, which must then be excluded. Second, the model assumes that accruals are similar within an industry, but firms operating in

[1] We also consider the study by Larson et al. (2018) as a notable contribution, but discuss it in Chapter 6.

multiple industries may not be well represented by one industry (Jackson, 2018; McNichols & Stubben, 2018). Ecker et al. (2013) argue that scaling firm characteristics by each firm's size is at least equivalent to industry-based sampling. The matching procedure would only increase test power if the correlated omitted variables were known and included in the model (Dechow et al., 2012, p. 276). Third, performance matching reduces test power because it leads to an underestimation of discretionary accruals. This is related to failure of correctly detecting earnings management (Keung & Shih, 2014). Banker et al. (2020, p. 848) conclude that "performance matching on ROA should never be used in accrual models because the matching variable (ROA) includes the dependent variable (accruals), which causes false inferences due to self-inflicted endogeneity bias."

5.2.3 Criticism of Earnings Management Models

Even though earnings management models are widely used, they are subject to criticism. This subsection addresses some general criticisms of earnings management models rather than specific models.

Are Discretionary Accruals Truly Discretionary?

We organize this subsection into three distinct critiques of earnings management models. First, the predominant criticism of earnings management models is that the discretionary portion assigned to each firm-year may be overstated. Earnings management models tend to have low explanatory power, resulting in a high unexplained variation, i.e., discretionary accruals. As Jackson (2018, p. 137) states, "the size of the reported effects are simply implausible." Most likely, a substantial portion of the discretionary accruals could be reduced by including the correlated omitted explanatory variables.

Second, the measurement method assigns a discretionary component to nearly every firm-year observation. Only for firm-years in which total accruals are exactly equal to normal accruals, i.e., the fitted value of the earnings management regression, no discretionary portion is assigned. More directly, one could say that earnings management models imply that nearly all managers have managed their earnings, although this is not necessarily true (Ball, 2013). While there is no factual evidence of the percentage of firms managing earnings, the study of Dichev et al. (2013, p. 24) asked chief financial officers (CFOs) the following question: "From your impressions of companies in general, in any given year, what percentage of companies use discretion within the Generally Accepted Accounting Principles (GAAP) to report earnings which misrepresent the economic performance of the business?" CFOs of publicly traded firms reported an average of 18.4%. The CFOs' peer evaluation shows that they expect less frequent earnings management than corresponding models. Finally, earnings management models are based on the assumption that the average of all observations represents unmanaged accruals.

The above arguments indicate that the true accruals, which are in the absence of any discretionary portion, remain unknown. However, they would be necessary to compute pre-managed earnings and the exact amount of earnings management.

Cross-Sectional vs. Time Series

This subsection presents the concept of cross-sectional and time series applications of earnings management models and their advantages and disadvantages. Cross-sectional methods use all firm observations at one point in time. More precisely, for cross-sectional models, the assumption implies that, on average, all firm-years are free of earnings management.

Some cross-sectional models account for fixed effects. A model accounting for year-fixed effects, implicitly assumes that, on average, reported earnings within the same year are unmanaged. For industry fixed effects models, the assumption would be that, on average, no earnings management occurs within this industry. Jeter and Shivakumar (1999) note that the cross-sectional Jones model is well specified for a sample of randomly selected firms, but measurement errors would occur for firms whose cash flows deviate from the industry median.

For time series models, the assumption is that the average earnings of a firm over all its years are unmanaged. Indeed, a firm's accruals must reverse at some point, reducing the problem of omitted correlated variables (Dechow et al., 2012). Empirical research shows that accrual reversal takes place. The top and bottom deciles of WCA in year t reverse almost perfectly in the following year. Long-term accruals also show a reversal effect, which, however, takes up to four years to reverse (Larson et al., 2018, p. 848).

From a conceptual point of view, at two points in time a firm's discretionary accruals can be expected to be zero: ultimately at the firm's inception and liquidation. However, a more detailed investigation of the creation and dissolution of accruals in a firm's lifetime would be difficult because accounting data are not widely available until firms go public, which then might already be contaminated by earnings management.

The time series approach is also criticized. Ronen and Yaari (2008) argue that time series analysis requires a long series of sequential observations. For example, the average observation length in Dechow et al. (1995) is more than 20 years. Requiring an extensive data history not only biases the sample in favor of surviving firms but also drastically reduces the size of the dataset. Moreover, it can be questioned whether the problem of correlated omitted variables is mitigated for long time series. This would assume that the firm's fundamentals are stable over time, which appears rather unrealistic for long periods, because a firm's focus, characteristics, and accrual policy may change over time. Discretionary accruals are also subject to changes in the economic model. G. A. Hansen (1999) provides evidence for this statement by showing that unexpected accruals resulting from structural changes such as acquisitions and divestitures lead to measurement errors in earnings management.

More importantly, Ye (2007) reports that year-to-year fluctuations in accruals are not necessarily due to earnings management. Models that use a firm's past accruals to estimate normal accruals implicitly assume no earnings management in the estimation period. However, prevailing earnings management could contaminate earnings in the estimation period. This can be easily illustrated with an example. Suppose a firm has engaged in earnings management by choosing intentionally too low but constant depreciation values in the past, which serves as our estimation period. In the year the asset is sold, the one-time realization of losses on the sale of assets results in negative accruals because earnings are reduced without affecting cash flows. Since time series models measure the variation in accruals from one year to the next, the large variation and thus the presumed year of earnings management would be the year in which the assets are sold, although earnings were actually managed in all previous years. Jeter and Shivakumar (1999, p. 300) conclude that "cross-sectional models appear to be less imprecise" when compared with time-series models.

Incentives Based on Pre-Managed Earnings

The linear earnings management models cannot capture incentives based on pre-managed earnings. The discontinuity literature expects firms with slightly negative pre-managed earnings to engage in earnings management to achieve slightly positive earnings. While alternative explanations for the kink have been discussed in the discontinuity literature, most authors conclude that the kink most likely results from earnings management.[2]

If firms manage their earnings upwards to zero, one would expect regression-based earnings management models to find higher discretionary accruals just at zero earnings or slightly above. Dechow et al. (2003, p. 371) assess the level of discretionary accruals of the forward-looking model using the reported earnings distribution. However, at the zero threshold, no abrupt change in discretionary accruals is observed between firms reporting small negative and small positive earnings. The lack of change in discretionary accruals seems inconsistent with the discontinuity literature.[3]

[2]For a detailed discussion of earnings threshold management, see Chapter 2.

[3]The results of Ayers et al. (2006), in contrast to Dechow et al. (2003), find significantly increasing discretionary accruals at the zero threshold. However, Ayers et al. (2006) do not consider the positive correlation of earnings and accruals, which could be an omitted correlated variable. Dechow et al. (2003) show that a regression naturally finds higher discretionary accruals for firms with higher earnings. Moreover, Ayers et al. (2006) group the intervals into earnings sizes of 0.01 of total assets, while most other studies use 0.005 as earnings interval size (Burgstahler & Chuk, 2017; Burgstahler & Dichev, 1997; S. K. Chen et al., 2010; Daske et al., 2006; Dechow et al., 2003; Enomoto & Yamaguchi, 2017; Gilliam et al., 2015; Jacob & Jorgensen, 2007). The significance level is determined using a p-test performed at the one-sided 10 % level. Finally, the sample period of Ayers et al. (2006) ranges only from 1994 to 2002 and includes financial institutions but still contains only 22,903 firm-year observations, which is considerably smaller than the sample of Dechow et al. (2003).

We discuss two explanations that could be responsible for this puzzle. First, the magnitude of other earnings management activities might be larger. For example, a firm with pre-managed positive earnings of 5% of total assets might decide to build reserves for future years by lowering its reported earnings to barely above zero. In the small profit interval, the negative discretionary accruals of this firm could offset the positive discretionary accruals of firms managing earnings upwards. Although this could be a possible explanation, it remains unclear whether it would be beneficial for managers to reduce average earnings downwards just barely above zero. The literature on earnings smoothing suggests that firms tend to report average earnings (Fudenberg & Tirole, 1995; Lambert, 1984). If the many firms in the small profit interval could be explained by firms originating from regions other than the small loss interval, the question would arise why fewer firms are reporting a small loss than expected. The argument remains a possible explanation, in our view, but it is unclear to what extent it contributes to the unexpectedly high frequency of firms reporting earnings in the small profit region.

A second possible explanation could be that the relatively high volatility of discretionary accruals covers the relatively small changes in discretionary accruals that turn negative earnings positive.

In general, it may be questioned whether earnings management models are the appropriate tool to measure earnings management around zero earnings. Earnings management models account for the magnitude of presumed earnings management at the firm level. In contrast, the histogram approach appears to capture earnings management at zero earnings well, as firms managing earnings from the small loss to the small profit interval accentuate the kink.

Even further from zero earnings, the earnings variable itself could determine whether managers engage in earnings management or not. Dechow et al. (2003, p. 371) find a "positive relation of earnings and discretionary accruals for extreme loss firms". Firms that report large negative earnings tend to have lower discretionary accruals than firms that report smaller negative earnings. Two possible explanations for this effect are suggested. The first is that firms might take "big baths", which would lead to negative discretionary accruals and thus distort the distribution for loss firms.[4] The second explanation of Dechow et al. (2003, p. 383) is that for loss firms, a part of the normal accruals is included in the discretionary accruals portion, which would be due to a misspecification of the model.

The incentive to manage earnings is not limited to firms that take a big bath. As the literature on earnings smoothing shows, reporting smooth earnings might have beneficial effects on the stock price (Fudenberg & Tirole, 1995; Kirschenheiter & Melumad, 2002; Lambert, 1984). If pre-managed earnings are excellent, managers could build a "cookie jar" and manage earnings downward. A cookie jar is easily built by overstating depreciation, amortization, or warranties. These reserves are reversed in less favorable years to increase reported earnings. Levitt (1998), the former SEC Chairman, called the cookie jar practice "hocus-pocus" because it leads to illusions in accounting and reduces earnings informativeness. It is important to note that incentives based on pre-managed earnings might be nonlinear. For example, Kirschenheiter and Melumad (2002) show that earnings smoothing toward average profit and big bath accounting can coexist.

These arguments suggest that there could be piecewise nonlinear or generally nonlinear relationships not accounted for by linear earnings management models. Firms might have different incentives depending on their earnings, such as taking big baths, building a cookie jar, or simply reaching the zero earnings threshold.[5] Therefore, it is important to aim to include this information in the earnings management models.

5.2.4 Criticism of Earnings Histograms

The histogram distribution of U.S. firms' earnings, scaled by lagged total assets, exhibits a large discontinuity at zero.[6] The vast majority of the literature agrees that the

[4]Other possible motivations for a big bath other than negative pre-managed earnings have been identified, for example, following the appointment of a new chief executive officer (CEO). This CEO is incentivized to build reserves for future years or reverse previously established reserves (Elliott & Shaw, 1988; Strong & Meyer, 1987; Wells, 2002).

[5]For a more detailed discussion of nonlinearities in earnings management models, see Chapter 6.

[6]This chapter consistently uses "earnings" instead of "net income" as in Chapter 2. This is because most of the earnings management literature refers to earnings.

discontinuity is due, at least in part, to firms managing earnings slightly upward to achieve positive earnings.[7] Burgstahler and Chuk (2017, p. 747) conclude that the hypothesis that "earnings are managed to meet benchmarks provides the most simple and complete explanation for the body of evidence." While the zero earnings distribution provides compelling evidence that earnings were managed upward to turn negative pre-management earnings into positive post-management earnings, the approach has some weaknesses. This subsection addresses two related criticisms of the zero earnings histogram that concern the histogram approach and its measurements.[8]

Earnings Discontinuity as an Aggregate Measure

Evidence of earnings management around zero earnings is only observable at the aggregate level of all firms. One cannot conclude that all firms in the small profit interval have managed earnings. Unlike earnings management models, the higher-than-expected number of firms reporting low earnings does not allow identifying the firms that managed earnings upwards.

Additionally, the earnings distribution does estimate the extent of earnings management at the firm level. Suppose a firm only marginally manages earnings upwards to achieve the zero earnings threshold. In that case, the kink increases most because there will be one less than expected firm in the small loss interval and one more than expected firm in the small profit interval, even though the earnings management magnitude is only marginal. In contrast, earnings management models indicate the extent of estimated earnings management, from which the estimated pre-managed earnings level can be calculated.

[7]See, e.g. Burgstahler and Chuk (2017), Burgstahler and Dichev (1997), Burgstahler and Eames (2006), Chardonnens et al. (2022), Daske et al. (2006), Dechow et al. (2003), Degeorge et al. (1999), Gilliam et al. (2015), and Gore et al. (2007). For exceptions, see Beaver et al. (2007) and Durtschi and Easton (2005, 2009).

[8]For a general discussion of earnings histograms, see Chapter 2.

Indication of Earnings Management Only at Zero Earnings

Another weakness of the earnings histograms is that earnings management only becomes visible at zero earnings. The histogram approach fails to capture earnings management at other earnings levels. CFOs also manage earnings towards other important thresholds as last year's quarterly profits or analysts' forecasts (Graham et al., 2005). However, identifying and measuring earnings management at these thresholds through histograms is less clear because the measurement methods are biased to provide too many observations in the small profit interval. Thus, earnings histograms, as presented in prior studies, are not suitable for estimating earnings management and its magnitude at the firm level.

5.3 Models Used: Jones Type and Distributional

This section briefly presents the models we use to measure the levels of accruals: the modified Jones model, the extended Jones model, and the distributional model. For each model, we use both definitions of accruals, introduced in Section 5.2.1.

5.3.1 Jones-Type Models

The modified Jones model was presented in Equation (5.4). The model uses the change in sales less the change in accounts receivable and gross PPE.[9]

In an additional, extended model, we consider the correlation of earnings and discretionary accruals of the modified Jones model. As described in Section 5.2.3, one finding of Dechow et al. (2003) is that negative earnings are positively related to discretionary accruals. It appears that the variables of the modified Jones model fail to account for this

[9]We do not use the earnings management model of Larson et al. (2018) because its explanatory variables, employee growth, the interaction term of net capital intensity, and the lead and lag times of cash flows, would restrict our sample size considerably. To avoid this, we use the explanatory variables of the modified Jones model to explain CompAcc.

relation. Ball and Shivakumar (2006) raise a possible explanation for this asymmetry, which could be the narrative of asymmetric timeliness found in Basu (1997).

To account for the relationship between earnings and accruals, we add two variables to the modified Jones model. First, we include earnings to reflect the generally positive relationship between earnings and accruals. Second, we include the interaction of earnings with a dummy variable for positive earnings to capture the asymmetric slope that could result from asymmetric timeliness.[10] The following equation represents our extended Jones model that is complemented with earnings information:

$$Accruals_{i,t} = \alpha + \beta_1(\Delta Rev_{i,t} - \Delta Rec_{i,t}) + \beta_2\ PPE_{i,t} + \beta_3\ E_{i,t} + \beta_4\ POS * E_{i,t} + \epsilon_{i,t}, \quad (5.7)$$

where $Accruals_{i,t}$ = Accruals for firm i in year t divided by $A_{i,t-1}$;

$\Delta Rev_{i,t} - \Delta Rec_{i,t}$ = Change in revenues for firm i in year t minus change in accounts receivables for firm i in year t divided by $A_{i,t-1}$;

$PPE_{i,t}$ = Gross property, plant & equipment for firm i in year t divided by $A_{i,t-1}$;

$E_{i,t}$ = Earnings for firm i in years $t-1$ to $t+1$ divided by $A_{i,t-1}$;

$POS * E_{i,t}$ = A dummy for which is 1 for positive earnings and 0 otherwise multiplied with earnings for firm i in years $t-1$ to $t+1$ divided by $A_{i,t-1}$;

$A_{i,t-1}$= total assets in year $t-1$ for firm i.

5.3.2 Distributional Model

In this subsection, we develop a firm-level earnings management indicator based on information from the earnings distribution. To assign a discretionary amount to each firm-year, we need to estimate the earnings distribution in the absence of earnings management.

In previous literature, only a few studies aim to derive a firm-level earnings management indicator based on distribution characteristics. The study by Mindak et al. (2016)

[10] Our study differs from that of Ball and Shivakumar (2006), which uses cash flows as an explanatory variable. In addition, Ball and Shivakumar (2006) include a dummy variable for negative cash flows and the interaction between the dummy variable and negative cash flows.

defines an achievable earnings threshold for each firm based on the known targets of zero earnings, last year's earnings, and analysts' forecasts. If the firm has reported earnings are just above the individual threshold, it is expected to have managed its earnings. The study's results suggest that earnings targets may depend on the firms' individual situations.

Some papers estimate the expected earnings management-free distribution. For example, S. K. Chen et al. (2010) construct a model in which the unmanaged earnings distribution is expected to follow a mixed-normal distribution around the three commonly tested thresholds. Similarly, Lahr (2014) and Chardonnens et al. (2022) use a kernel density estimator to estimate the expected number of firms for each earnings interval. Jacob and Jorgensen (2007) construct their expected earnings distribution from the mean of four consecutive quarters. The expected and observed distributions differ mainly near the threshold but remain unaffected for earnings further away from zero (Jacob & Jorgensen, 2007, p. 383). These methods may be promising, but they all focus only on the zero earnings threshold. Our firm-level measure of earnings management aims also to capture earnings management that occurs further from the zero threshold.

To estimate the expected earnings, we follow Dechow et al. (2003, p. 374) and subtract the amount of the firm's discretionary accruals from its reported earnings, as in the following equation:

$$\begin{aligned} Expected\ earnings_{i,t} &= Reported\ earnings_{i,t} - Discretionary\ accruals_{i,t} \\ &= Normal\ accruals_{i,t}. \end{aligned}$$

We can now create the expected distribution from the expected earnings by creating a histogram grouped by expected earnings intervals of 0.5% of lagged total assets. The variation induced by the discretionary accruals smooths the shape of the expected distribution. We see two advantages in using this smoothing method. First, we obtain an estimated expected distribution that extends over broad regions of the earnings distribution and has no kink at zero earnings. Second, this approach ensures that the average earnings of all firms combined remain unchanged because the sum of all discretionary accruals is zero.

In our distributional earnings management model, firms are assigned to the expected distribution in the ordering of their reported earnings.[11] Equation (5.8) formally presents the computation:

$$Distributional\ model_r = Reported\ earnings_r - Expected\ earnings_r, \tag{5.8}$$

where r represents the earnings rank of the firm. For example, Microsoft reported 2014 earnings scaled by lagged total assets of 15.5%, corresponding to earnings rank 88,838 out of 98,491 firm-years, starting with the lowest earnings. Microsoft's 2014 expected earnings would have been 19.9% of lagged total assets, which would correspond to earnings rank 90,332 of our expected earnings of 98,491 firm-years. Reported earnings minus expected earnings results in a difference of -4.4% of Microsoft's lagged total assets. Thus, our distributional model would indicate presumed earnings management of -4.4% of lagged total assets.

5.4 Evaluation Method

We use two different methods to assess the quality of the models. The first method evaluates the amount of discretionary accruals of firms found guilty by the SEC of fraudulently overstating earnings for each model and measures the significance. The second method predicts the highest 1% of discretionary accruals to figure in the AAER sample and then calculates the sensitivity and the precision to assess the accuracy of each model.

[11] Technically, we do this by first ordering and sorting all expected earnings in one data frame. Second, we order and sort the reported earnings in a second data frame. Third, we bind the columns of the first data frame containing the expected earnings to the second data frame containing the reported earnings. Finally, we compute our distributional model indicator from the difference between reported earnings and expected earnings of the same earnings rank.

5.4.1 Discretionary Accrual Level per Model

Our first evaluation method compares the levels of discretionary accruals between overstated firm-years and non-overstated firm-years. Dechow et al. (2012, p. 311) evaluate the WCA of fraudulent and non-fraudulent firms over several years. They find that the mean WCA of fraudulent firms is higher in the years before the fraud is detected and lower in the years afterward compared to non-fraudulent firms. We expect that firm-years identified by the SEC as fraud firm-years have higher accruals, as observed in Dechow et al. (2012).

To assess whether the earnings management models successfully detect earnings management, we compare the discretionary accruals of firms caught by the SEC and found guilty of overstating financial statements to the discretionary accruals of firm-years that are not part of the AAER sample. We measure significance by calculating the z-score as in Devore and Berk (2012, pp. 490–491):

$$z = \frac{DA_O - DA_{NO}}{\sqrt{\frac{s_O^2}{N_O} + \frac{s_{NO}^2}{N_{NO}}}}, \tag{5.9}$$

where DA_O (DA_{NO}) is the mean of the discretionary accruals of the overstated (non-overstated) firm-years, s_O^2 (s_{NO}^2) is the variance of overstated (non-overstated) firm-years, and N_O (N_{NO}) is the number of observations for the overstated (non-overstated) samples.

5.4.2 Predictive Power of Accounting Fraud

Our second evaluation method tests whether the highest 1% of discretionary accruals are associated with a higher incidence of fraud in firm-years. This is similar to the approach developed by Bao et al. (2020).[12] One of the challenges in the fraud prediction literature

[12]Bao et al. (2020, p. 202) also select the highest 1% of their earnings management measure as fraud firms. They use the performance evaluation metric called "Normalized Discounted Cumulative Gain at the position k", which they set to 1% to deal with the comparatively small number of fraud firms. The advantage over conventional performance evaluations is that k can be defined manually.

is that fraud occurs infrequently. The number of firm-years in which fraud is absent far exceeds the number of firm-years in which fraud is present. Therefore, it seems plausible to predict only a small selection of the sample as fraud. Bao et al. (2020) predict the 1% of firms they identify with the highest probability of fraud as fraudulent firms.

Although we use the concept of fraud prediction similar to Bao et al. (2020), our study differs in several ways. First, fraud prediction models typically use more explanatory variables. Cecchini et al. (2010) include 40 raw data variables, while Dechow et al. (2011) identify 11 financial ratios. Bao et al. (2020) run one model with 294 raw data variables of the Compustat database. In contrast, the objective of our study is to evaluate the performance of earnings management models with less demanding data requirements, which does not limit our sample as much. Second, Bao et al. (2020)'s training and testing period differ. They use the years 1991 to 2001 to train and validate their model and predict fraud for 2003 to 2008. Our study does not use an older sample to predict more recent fraud. Instead, our procedure aims to test whether the earnings management models correctly predict fraud in the same year through discretionary accruals. Third, Bao et al. (2020) use flexible and nonlinear machine learning models, more specifically an ensemble model, to predict fraud. Our Jones-type models are linear and piecewise linear models that do not use machine learning. Finally, Bao et al. (2020) consider serial fraud in most of their analyses, while we ignore it.[13]

We evaluate our models' quality consistent with Bao et al. (2020). True positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) allow us to measure the sensitivity and precision of the prediction. TP are firm-years that figure in the top 1% of the model's earnings management measure and are simultaneously identified as fraud firm-years by the SEC. FP are firm-years that gather in the top 1% of the model's earnings management measure but are not identified as fraud firm-years by the SEC. FN are firm-years that do not figure in the top 1% of the model's earnings management measure but are identified as fraud firm-years by the SEC. TN are firm-years that do not figure in

[13]Serial fraud occurs when firms have been identified by the SEC as having fraudulently misrepresented their earnings for several consecutive years.

the top 1% of the model's earnings management measure and are simultaneously not identified as fraud firm-years by the SEC. TP and FN sum up to all firm-years included in the fraud sample, while TP and FP sum up to the predicted fraudulent firm-years. Figure 5.2 illustrates the differences between TP, FN, FP, and TN.

Sensitivity and Precision		**Actual fraud firm-year?** positives	negatives
Predicted fraud firm-year?	positives	TP	FP
	negatives	FN	TN

Figure 5.2: Sensitivity and Precision. The figure shows the sensitivity, precision, and differentiation of TP, FP, FN, and TN. TP are defined as firm-years that figure in the fraud sample and are simultaneously in the top 1% of the model's earnings management measure. FN are defined as firm-years that do not figure in the fraud sample but are in the top percentile of the model's earnings management measure. FP are defined as firm-years that figure in the fraud sample but are not in the top percentile of the model's earnings management measure. TN are defined as firm-years that do not figure in the fraud sample and the top 1% of discretionary accruals. TP and FN sum up all firm-years included in the fraud sample, while TP and FP sum up all predicted firm-years.

The metrics for prediction quality, which are the sensitivity and precision, can be calculated from the true and false classification. The sensitivity ratio consists of the TP divided by the TP and FN:

$$Sensitivity = \frac{TP}{TP + FN}. \tag{5.10}$$

The precision is defined as the ratio of TP divided by TP and FP:

$$Precision = \frac{TP}{TP + FP}. \tag{5.11}$$

5.5 Data and Descriptive Statistics

5.5.1 Accounting Data

We retrieved our data from Refinitiv (formerly Thomson Reuters) Datastream.[14] The sample includes publicly traded U.S. firms from 1987 to 2014. We only use data up to 2014 because the sample of the University of California-Berkeley Center for Financial Reporting and Management (CFRM) includes few fraud cases after 2014.

In our data selection, we closely follow the calculations in Dechow et al. (2012, p. 291).[15] We retain observations with non-negative[16] and non-missing values of Cash & Short Term Investments (WC #02001), Current Assets (WC #02201), Short-Term Debt (WC #03051), Current Liabilities (WC #03101), PPE (WC #02301), Total Assets (WC #02999), and Sales (WC #01001). We also require available values of Earnings Before Extraordinary Items (WC #01551). We exclude financial firms with SIC sectors from 4400 to 4999 and 6000 to 6499. In addition to Dechow et al. (2012), we require non-negative and non-missing Common Equity (WC #03501) to compute CompAcc.

Finally, we impose some additional filters where we might expect some data errors. More specifically, we remove firm-years in which the accrual components as well as gains and losses exceed total assets of the same year.[17] All regression input variables, but not the accrual components, are winsorized at the 1% level.

[14]See GitHub for a detailed description of how the firms were selected, downloaded, and the general sample was constructed: https://github.com/chardonnensp/Datastream-Worldscope-fundamental-dataset

[15]In contrast to our study, Dechow et al. (2012) use the Compustat annual tape. The Compustat sample covers more firms with more accounting items per firm, which explains why our sample is smaller. Nevertheless, both databases, Worldscope and Compustat, should lead to comparable results (Ulbricht & Weiner, 2005).

[16]Dechow et al. (2012, p. 291) "require positive non-missing values" of the balance sheet and income statement items. We assume that they interpret zeros as positive. If zeros were excluded, the sample size would shrink notably, because many firms have zero short term debt. This is the case, for example, for Microsoft from 1993 to 2008 and Apple from 1998 to 2002 and 2004 to 2013.

[17]The motivation for this additional data cleaning is to mitigate possible sample differences and database errors in the Dechow et al. (2012) study that may occur in the Worldscope data but not in the Compustat data.

Descriptive Statistics

Table 5.1 presents the summary statistics for the accrual and the Jones-type model components. WCA in Panel (a), CompAcc in Panel (b), and the Jones-type model components in Panel (c) are winsorized at the 1% level.

(a) WCA Components

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
WCA	98,491	0.014	0.121	-0.384	-0.032	0.050	0.535
Net Working Capital	98,491	0.133	0.244	-0.658	-0.015	0.273	0.845
Current Assets	98,491	0.617	0.416	0.032	0.348	0.786	2.674
Cash	98,491	0.229	0.321	0.000	0.028	0.304	1.915
Current Liabilities	98,491	0.314	0.254	0.026	0.155	0.385	1.570
Short Term Debt	98,491	0.060	0.112	0.000	0.000	0.063	0.643

(b) CompAcc Components

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
CompAcc	98,491	0.047	0.267	-0.709	-0.047	0.104	1.407
Common Equity	98,491	0.532	0.505	-1.375	0.306	0.765	2.603
Cash	98,491	0.229	0.321	0.000	0.028	0.304	1.915

(c) Jones-Type Model Components

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Sales	98,491	1.342	1.046	0.000	0.640	1.738	5.785
Receivables	98,491	0.192	0.171	0.000	0.068	0.262	0.900
PPE	98,491	0.290	0.274	0.001	0.085	0.406	1.320
Earnings	98,491	-0.033	0.228	-0.868	-0.078	0.086	0.432

Table 5.1: Summary Statistics for Accruals and Their Components. The table shows the summary statistics for the variables required to compute accruals.

Panel (a) shows the summary statistics of the change in annual WCA: $WCA_{i,t} = \Delta\ Current\ Assets\ (WC\ 02201)_{i,t} - \Delta\ Cash\ \&\ Short\ Term\ Investments\ (WC\ 02001)_{i,t} - (\Delta\ Current\ Liabilities\ (WC\ 03101)_{i,t} - \Delta\ Short\text{-}Term\ Debt\ (WC\ 03051)_{i,t})$.

Net working capital represents the total share of total assets computed as follows: $Net\ Working\ Capital_{i,t} = Current\ Assets\ (WC\ 02201)_{i,t} - Cash\ \&\ Short\ Term\ Investments\ (WC\ 02001)_{i,t} - (Current\ Liabilities\ (WC\ 03101)_{i,t} - Short\text{-}Term\ Debt\ (WC\ 03051)_{i,t})$.

Panel (b) shows the CompAcc, which are calculated as annual changes: $CompAcc_{i,t} = \Delta\ Common\ Equity\ (WC\ 03501)_{i,t} - \Delta\ Cash\ \&\ Short\ Term\ Investments\ (WC\ 02001)_{i,t}$.

Panel (c) shows the summary statistics for the explanatory variables of the modified and extended Jones model, consisting of $Sales\ (WC\ 01001)_{i,t}$, $Receivables\ (WC\ 02051)_{i,t}$, $PPE\ (WC\ 02301)_{i,t}$, and $Earnings\ (WC\ 01551)_{i,t}$.

All variables are scaled by lagged $Total\ Assets\ (WC\ 02999)$.

Figure 5.3 shows the correlation matrices for the selected variables in Table 5.1 from 1987 to 2001 in Panel (a) and 2002 to 2014 in Panel (b). Pearson correlations are shown in the upper right, and Spearman correlations in the lower left. All variable definitions correspond to those in Table 5.1. Figure 5.4 shows the histograms of accruals, their

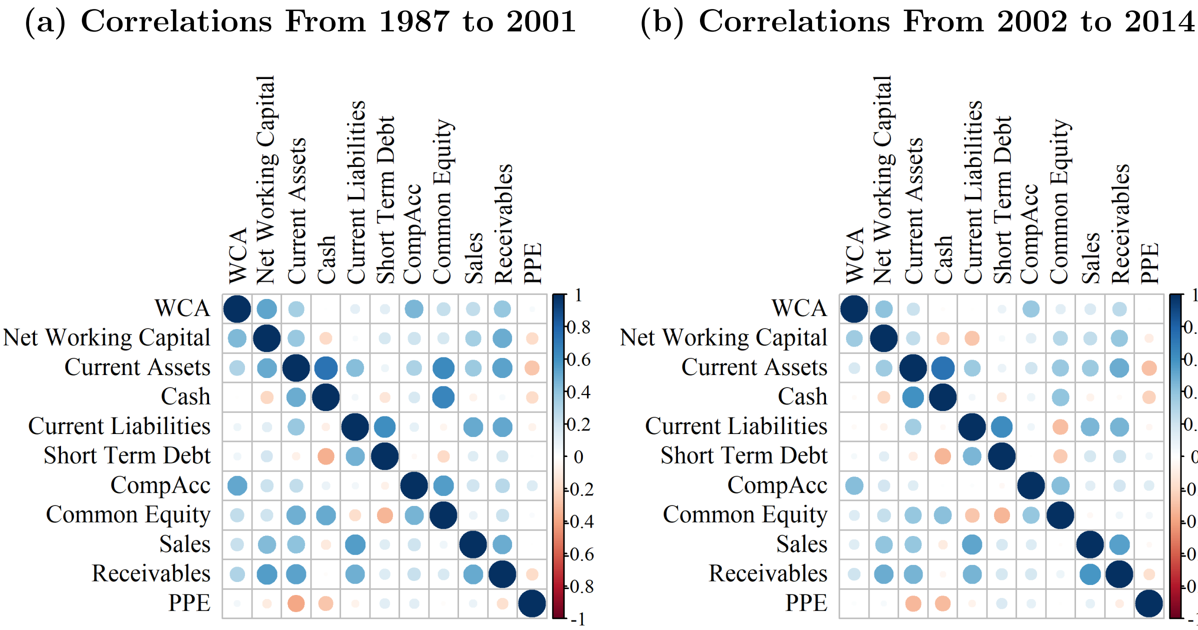

Figure 5.3: Correlation Matrices for Accruals and Their Components. The figure shows the correlation matrices for the selected variables in Table 5.1 from 1987 to 2001 in Panel (a) and 2002 to 2014 in Panel (b). Pearson correlations are shown in the upper right part, and Spearman correlations in the lower left part. The variable definitions correspond to those in Table 5.1.

components, and the variables of the modified and extended Jones model. Figure 5.5 shows the annual distribution of WCA and CompAcc from 1987 to 2014. The distribution of WCA is narrower compared with the distribution of CompAcc.

5.5.2 AAER Sample

Earnings management models estimate the level of discretionary accruals. However, these models do not provide external validity as it is unclear whether the assigned discretionary amount truly arises from earnings management. To verify whether the high levels of discretionary accruals relate to the fraud identification by regulators, the SEC's AAER database is frequently used (Dechow et al., 1995, 2012; Ibrahim, 2009; Jackson, 2018).

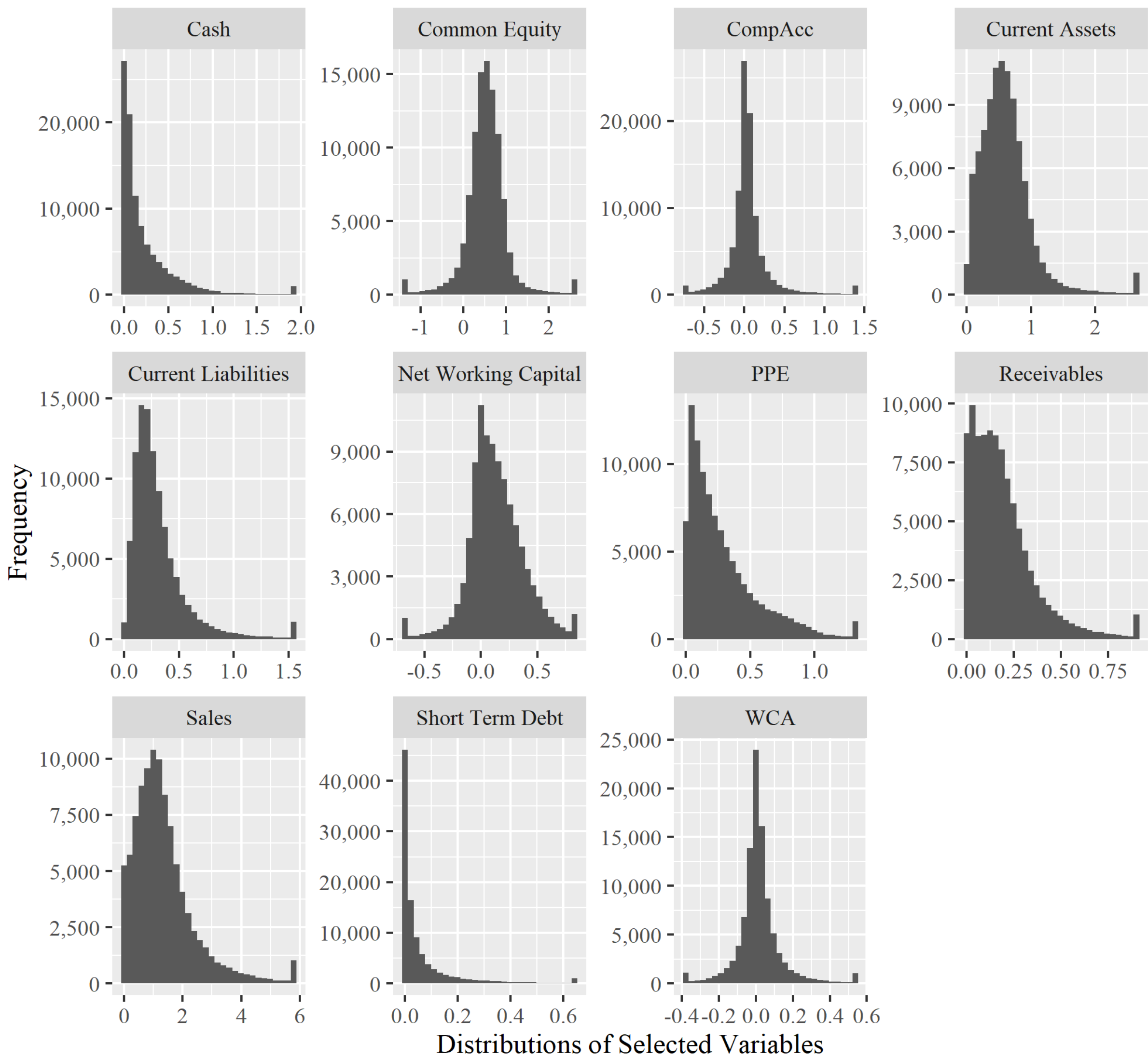

Figure 5.4: Histograms for Accruals and Their Components. The figure shows the histograms of accruals, their components, and the variables of the modified and extended Jones model. The variable definitions correspond to those in Table 5.1.

The purpose of the SEC enforcement program shifted several times. Before the 1990s, the enforcement program was more corrective but became more punitive around the turn of the millennium. A series of bills passed in Congress expanded the SEC's power to punish firms and individuals. A strengthened SEC, coupled with the more frequent use of modern compensation practices, such as stock options, led to higher levels of fraud detection. This period is also referred to as the "scandal period". In the second half of the millennium's first decade, the SEC's primary focus shifted away from accounting fraud. One reason was that Ponzi schemes such as the Madoff fraud were given more weight than accounting fraud (Atkins & Bondi, 2008; Rakoff, 2014).

(a) WCA Distribution by Year **(b) CompAcc Distribution by Year**

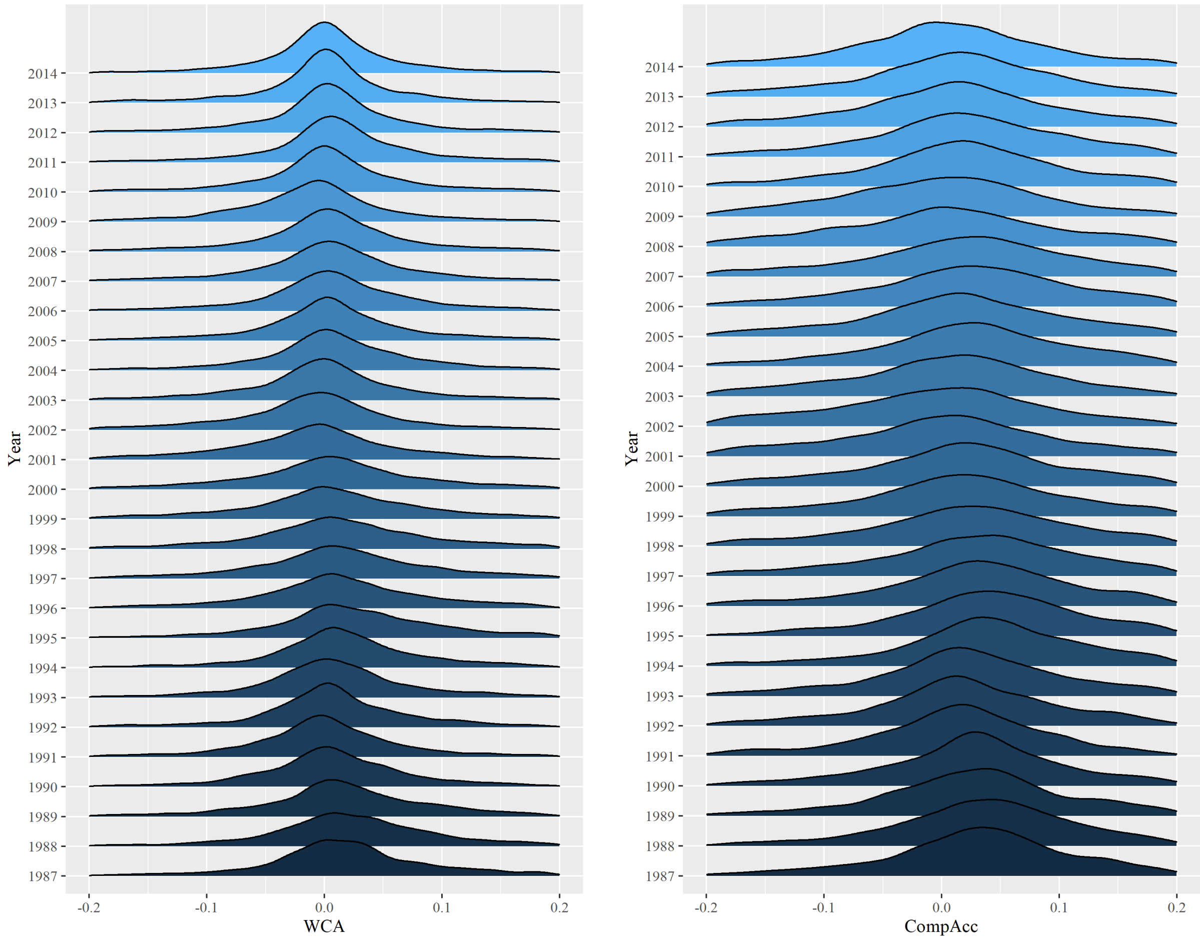

Figure 5.5: Yearly Accruals Distributions. The figure shows the annual distribution of WCA and CompAcc from 1987 to 2014. The distribution of WCA is narrower compared to the distribution of CompAcc. The variable definitions correspond to those in Table 5.1.

Dechow et al. (2011) collected the AAER data from the SEC website and constructed a database called CFRM. The CFRM database is an aggregation “[...] of firms that have been subject to enforcement actions by the U.S. SEC for allegedly misstating their financial statements.” (Dechow et al., 2011, p. 18). It appears sound to assume that the sample includes some of the most severe cases of earnings management. Nevertheless, the SEC has limited resources and may focus on larger firms with large impacts on investors, such as Enron. Not all firms that engage in manipulation will be dismantled or subject to an SEC investigation. This type of selection bias likely occurs in all samples of externally identified manipulation (Dechow et al., 2011).[18]

We received the CFRM Excel file directly from Patricia Dechow.[19] The CFRM Excel file consists of five Excel sheets. The first sheet, "detail", contains each of the 1,657 AAER releases once and the reason for the enforcement. Even though we received the updated version of the December 2018 CFRM file, few fraud cases have been documented after 2014. This is because cases may be open multiple years before a final decision is made (Karpoff et al., 2017). Therefore, analogous to Bao et al. (2020), we only use data up to 2014. The second sheet, "ann", contains all years in which a firm's financial statements were affected by the fraud. The third sheet, "qtr", contains quarterly information. The fourth sheet, "Table 1 - Panel G", contains a matching table of the SEC firm names to the Compustat firm names. Researchers with a Compustat database use this table to match the firms of sheets one through three. The fifth and final sheet, "Table 1-Panel H", lists the old and new AAER sample identification numbers and the SEC-generated Central Index Key.

Since we use Refinitiv Worldscope data, we cannot directly match the CFRM sample with Compustat, as done in Dechow et al. (2011). The CFRM sample contains only three identifiers: the SEC Central Index Key in sheet 5, the firm name used by the SEC, and the Compustat ID in sheet 4. As the Refinitiv universe does not contain any of these firm identifiers, a multi-step matching procedure is required to match the Refinitiv or Worldscope name with the SEC or Compustat name. The following procedure describes the matching of the CFRM database with our accounting data. We start with the firm names in the AAER database in Panel H, which contains 1,547 listed fraud firms, and perform the following procedure:

[18]The accounting literature uses several samples of misconduct. Karpoff et al. (2017) compare the CFRM to three commonly used databases, namely the Government Accountability Office, the Audit Analytics database of restatement announcements, and the Stanford Securities Class Action Clearinghouse database of securities class action lawsuits. While the samples have different strengths and weaknesses, the CFRM appears to be the most appropriate for our study because the CFRM database best captures value-related events. Also, the studies of Bao et al. (2020) and Dechow et al. (2011) use the CFRM database.

[19]We thank Patricia Dechow for providing us with the data free of charge.

1. We exclude duplicate entries of 42 firms.

2. We convert all SEC, Compustat, and Refinitiv Worldscope firm names to uppercase. Additionally, we clean all firm names for punctuation or generic terms such as "corporation", "group", or "ltd".

3. We match 508 firms whose Compustat name in sheet 4 is identical to the Datastream name.

4. We match 67 firms whose SEC name is identical to the Datastream name.

5. For the remaining entries, we use a fuzzy string matching technique called the Jaro Distance, to match Compustat and Datastream names. We set the maximum distance to 1.[20] We manually evaluate the matched names and identify 277 firms for which the Jaro Distance resulted in a correct match.

6. We repeat the previous fuzzy string matching of the SEC name with the Datastream name, manually evaluate the matched names, and identify 76 firms for which the Jaro Distance resulted in a correct match.

This procedure results in 928 unique firms that are part of the CFRM fraud and our Worldscope sample. This is less than the 1,131 matched Compustat firms in "Table 1 - Panel G" in the CFRM data file. The smaller matched Datastream sample size appears to be plausible for two reasons. First, the number of firm-years in the Worldscope sample is generally smaller than the Compustat sample. Second, we only include firms headquartered in the U.S., which means that some foreign firms such as Novartis, SAP, or ABB are not included.

After identifying all matching firms in the CFRM sample and the Worldscope names, we match all firm names in the detail and annual table on sheets 1 and 2 of the Excel

[20]The Jaro Distance is designed for matching human-typed names (van der Loo et al., 2020). A Jaro Distance of 0 would represent a perfect match, and a distance of 1 no similarities. We set the maximum distance to 1, which returns the best fitting match of the remaining Compustat names to the remaining Datastream names.

file. Some firms have committed serial fraud and are therefore listed multiple times in the annual table.

Not all firms are included in the AAER sample because of earnings misstatements. Sheet 1 of the CFRM sample contains firm-years subject to the AAER. More than one-third of the sample is related to other events such as bribes or disclosure. Following Dechow et al. (2011), we remove these firm-years and result with 867 firm-years of earnings misstatements that we could identify in the Refinitiv Worldscope database. We match these firm-years with the detail and annual table of sheets 1 and 2 of the AAER Excel file. We then match this table to our accounting data. Because of our data selection procedure for the accounting data in the previous section, we do not have a firm-year entry for every firm that was found in the Worldscope database. This step reduces the fraud sample to 470 firm-years.

A few firms have understated instead of overstated their earnings. For our purpose of identifying upward earnings management, it is crucial to separate these firms. The sample of Dechow et al. (2011, p. 30) marks these cases with a dummy variable for understatements. We remove these 22 firm-years from our sample to result in our final fraud sample that contains 448 firm-years.[21] Table 5.2 summarizes the selection of the CFRM data.

Firm-years selection	**Dropped FY**	**Remaining FY**
All CFRM firm-years		1,657
Firm-years related to earnings misstatements	(609)	1,048
Firm-years matched to Worldscope firms	(181)	867
Firm-years meeting filter criteria of Section 5.5.1	(397)	470
Firm-years overstated	(22)	448

Table 5.2: Selection of Misstated Earnings Firm-Years. The table shows the selection of firm-years with misstated earnings. We obtain 448 firm-years with misstated earnings for our fraud prediction.

Table 5.3 provides an overview of the yearly observations for the matched Datastream and AAER firms. The ratio column indicates the percentage of firms that committed fraud. Compared with the distribution of Bao et al. (2020), we have about 40% fewer

[21] It would be interesting to assess the discretionary accruals of firms found guilty of underreporting earnings; however, they are very small in number and would potentially lead to unrobust results.

firm-years. One reason is that our sample starts in 1987, while the one of Bao et al. (2020) starts in 1979. Also, the Datastream sample generally contains fewer observations (except for the years around the Internet bubble crisis), and there is less accounting data available per firm-year. The fraction of fraudulent firms in our sample (0.45%) is lower than in the Bao et al. (2020) sample (0.57%). Nevertheless, the distribution of fraudulent firms is quite similar. In both samples, the years from 1999 to 2003 have the highest proportion of fraudulent firms.

Year	**Firm-Years**	**Fraud Firm-Years**	**In %**
1987	1,622	4	0.25
1988	1,727	1	0.06
1989	1,807	1	0.06
1990	1,807	2	0.11
1991	1,842	3	0.16
1992	2,028	3	0.15
1993	2,161	1	0.05
1994	2,282	5	0.22
1995	3,221	11	0.34
1996	3,607	10	0.28
1997	4,126	23	0.56
1998	4,530	31	0.68
1999	5,575	40	0.72
2000	5,305	49	0.92
2001	5,063	58	1.15
2002	4,754	36	0.76
2003	4,486	40	0.89
2004	4,358	32	0.73
2005	4,298	26	0.60
2006	4,307	20	0.46
2007	4,169	14	0.34
2008	4,044	12	0.30
2009	3,777	8	0.21
2010	3,625	8	0.22
2011	3,631	5	0.14
2012	3,600	4	0.11
2013	3,376	1	0.03
2014	3,363	0	0.00
Total	**98,491**	**448**	**0.45**

Table 5.3: Yearly Distribution of Fraud Firm-Years. The table shows the yearly number of total firm-years, fraud firm-years, and the percentage of fraud to total firm-years. A clustering of fraud firm-years occurs around the turn of the millennium, commonly referred to as the “scandal period”.

One firm can be found guilty of misstating more than one accounting item. Table 5.4 lists the fraudulent accounting items identified by the SEC that were the subject of fraud for overstating accounting items and are also included in our Datastream database. The item "Other than income, expense or equity" is the most frequently affected by overstatements. An example of a firm classified in this item is Apple, which was found guilty of reporting "backdated stock options and understated compensation expense" in 2001 and 2002 (Dechow et al., 2011). Revenue was also a frequently manipulated item, as were various balance sheet accounts such as accounts receivable, inventories, and reserve accounts.

Name	**Item**	**Frequency**	**% of Fraud FY**
Other than income, expense or equity	inc ex se	256	57.1
Revenue	rev	227	50.7
Other asset account	asset	85	19.0
Receivable	rec	81	18.1
Inventory	inv	62	13.8
Reserve accounts	res	52	11.6
Cost of goods sold	cogs	49	10.9
Liabilities	liab	32	7.1
Account payable	pay	20	4.5
Bad debts	debt	2	0.4
Marketable securities	mkt sec	1	0.2
Total		**867**	**193.5**

Table 5.4: Fraud per Accounting Item. The table shows the fraud frequency per accounting item for overstated firm-years. Revenue overstatements are the most common, followed by non-specified items and accounts receivable. On average, firms were convicted of fraud for 1.94 accounting items.

Microsoft also figures in the sample for 1995 through 1998. It was found guilty of having "maintained undisclosed reserves, accruals, allowances, and liability accounts (sometimes overstating income and sometimes understating income)" (Dechow et al., 2011). Of the above items, Microsoft falls in the category of "Other asset account".[22] On average, a firm committed fraud on 1.94 accounting items, calculated by dividing 867 affected accounting items by 448 fraudulent firm years.

[22] Although the explanation describes Microsoft as sometimes understating income, it is not flagged as understating firm in the CFRM sample and is therefore included in our analysis.

5.6 Results

The results section is divided into two parts. In the first part, we describe the characteristics of discretionary accruals for the different earnings management models. In the second part, we use the AAER fraud sample to assess whether discretionary accruals are predictive for fraudulent overstatements. We also predict the highest 1% of discretionary accruals as overstated firm-years, which we then test for accuracy and sensitivity.

5.6.1 Discretionary Accruals of the Different Models

This subsection evaluates the characteristics of the discretionary accruals of the modified Jones model, the extended Jones model, and the distributional model. We first present the regressions of the Jones models and evaluate the discretionary accruals graphically. With the extended Jones model, we expect to overcome the problematic correlation between discretionary accruals and earnings. Finally, we present the discretionary accruals computed from the distributional models.

Jones-Type Models

Table 5.5 reports the regression results for our Jones-type models defined in Section 5.3. Regressions (1) and (2) present the results for the WCA definition, while Regression (2) additionally includes earnings variables defined as the extended Jones model. Regressions (3) and (4) show the results for the CompAcc definition, with the explanatory variables of the modified Jones model included in (3) and the explanatory variables of the extended Jones model included in (4).

The results of Regression (1) of the modified Jones model show similar results to

	Dependent variable:			
	WCA		CompAcc	
	(1)	(2)	(3)	(4)
Intercept	0.006***	0.004***	−0.007***	−0.024***
	(10.679)	(6.309)	(−5.650)	(−16.157)
ΔRev - ΔRec	0.077***	0.059***	0.183***	0.134***
	(74.644)	(55.638)	(81.563)	(58.021)
PPE	−0.004***	−0.013***	0.111***	0.095***
	(−2.962)	(−9.199)	(37.257)	(32.404)
Earnings		0.066***		0.105***
		(31.312)		(22.967)
POS*Earnings		0.144***		0.533***
		(23.370)		(40.221)
Observations	98,491	98,491	98,491	98,491
R^2	0.054	0.090	0.082	0.129
Adjusted R^2	0.054	0.090	0.082	0.129
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 5.5: Regression Results of Earnings Management Models. The table shows the regression results of our two different accrual measures, WCA and CompAcc.
Regressions (1) and (2) present the results for the WCA definition for the modified Jones model in Regression (1) and the extended Jones model in Regression (2). Regressions (3) and (4) show the results for the CompAcc definition, with the explanatory variables of the modified Jones model included in (3) and the explanatory variables of the extended Jones model included in (4). The explanatory power R^2 clearly increases if earnings information is included in the regression.
$\Delta REV - \Delta Rec$ stands for the yearly change of a firm's Sales (WC 01001) minus Receivables (WC 02051). PPE stands for balance sheet item gross PPE (WC 02301). Earnings stands for earnings before extraordinary items (WC 01551). POS*EBEI is an interaction term of a dummy for positive earnings. All variables are scaled by lagged total assets (WC 02999).

Dechow et al. (2012, p. 299).[23] All variables are significant at the 1% level. However, the low explanatory power of the adjusted R^2 indicates that the variables cannot capture much of the variation in WCA. In Regression (2), the additional earnings variables, i.e., earnings and a dummy variable for positive earnings multiplied by earnings themselves,

[23] The regression coefficients, as reported in Dechow et al. (2012, p. 299), are 0.091 with a t-stat of 58.7 for the change in sales minus the change in revenue. Gross PPE is -0.007 (-9.6), and the R^2 is 6.8%. Even though the results are similar, the following differences could be responsible for the slightly different statistics. The model in Dechow et al. (2012, p. 299) additionally contains three partitioning variables, all of which are insignificant and of very small magnitude. Other differences are that Dechow et al. (2012) use the Worldscope database and a different time frame.

are both significant, clearly increasing the explanatory power to 9.0%. As expected, higher earnings are associated with higher WCA. This effect is even stronger for positive earnings.

In Regression (3), the adjusted R^2 is 8.2%, which is higher than in Regression (1). All coefficients are significant at the 1% level. The change in sales minus the change in revenue shows a stronger association for the CompAcc model. The extended Jones model in Regression (4) increases the explanatory power to 12.9%. Similar to the WCA definition, the earnings variables appear to capture the correlation between accruals and earnings.

Figure 5.6 illustrates the correlation of accruals and earnings and shows how the extended models correct for it by including earnings information. The black (blue) line represents the mean (median) level of accruals, and the gray area represents the 25th and 75th percentiles for each earnings interval. Panel (a) plots the WCA per earnings interval. A positive correlation of WCA and earnings becomes visible: firms with negative earnings tend to have negative WCA, while firms with positive earnings mostly have positive WCA. Panel (b) plots the CompAcc to earnings. The positive relation between CompAcc and earnings appears to be even stronger. For example, a firm reporting earnings equal to 5% of lagged total assets has, on median, total accruals equal to 5% of lagged total assets.

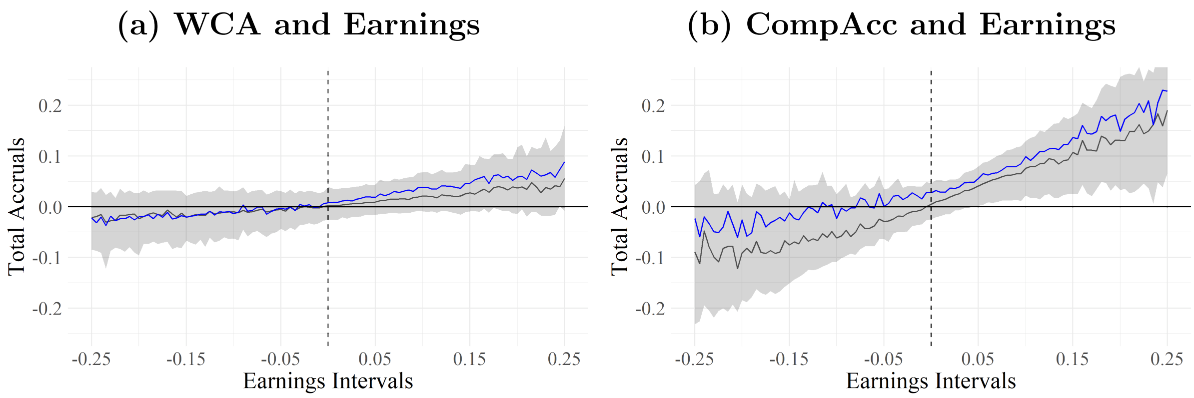

Figure 5.6: Total Accruals per Earnings Interval. The figure presents the accruals (y-axis) and earnings (x-axis) for the modified Jones model as defined in Section 5.3.1. Panel (a) follows the definition of WCA, while Panel (b) presents the CompAcc. The black (blue) line represents the mean (median) of the accruals for each earnings interval. The gray area represents the 25th and 75th percentiles.

Previous research identified and dicussed the positive relationship between earnings and accruals (Dechow et al., 2003). The effect would arise because firms with high growth have disproportionately higher accounts receivables than firms with a lower growth rate (Kothari et al., 2005; Larson et al., 2018).

Wallmeier (2021) makes a different argument as to why accruals are positively associated with earnings and negatively associated with operating cash flows in the cross-section of firms. He considers that earnings and cash flows are linked to true profitability, defined as profitability before discretionary items such as accounting errors or earnings management. While true profitability is unobservable, some knowledge about its cross-sectional distribution generally exists. For example, in a competitive business environment, extremely high and low values are rare, and the true profitability of most companies falls within a narrow range. When earnings and cash flows have a random component, the anchor of true profitability is sufficient to generate the empirically observed correlations between total accruals, earnings, and cash flows. Wallmeier (2021) argues that prior information about the distribution of true profitability is neglected in the (modified) Jones model and can be used to better decompose total accruals into normal and discretionary components.

It is important to note that the correlation of earnings and accruals is not in itself a problem in the definition of accruals. However, it is crucial that the positive correlation of earnings and accruals does not cause a correlation of discretionary accruals and earnings.

Figure 5.7 presents the discretionary accruals (y-axis) and earnings intervals (x-axis) for our accrual definitions WCA and CompAcc, using the variables of the modified Jones model and the extended Jones model. Panel (a) shows a positive relationship between discretionary accruals and earnings for WCA. It appears that this relation is carried over from total accruals. In other words, the modified Jones model does not fully account for the correlation of earnings and accruals with its explanatory variables. This is problematic because firms are assigned a discretionary portion according to their level of reported earnings, which is systematic by the characteristics of the model. However, there is no indication that firms reporting positive (negative) earnings manage, on average, their earnings upward (downward). Firms reporting negative earnings could have managed their earnings upward just as much as a firm reporting positive earnings.

(a) Modified Jones Model on WCA **(b) Modified Jones Model on CompAcc**

(c) Extended Jones Model on WCA **(d) Extended Jones Model on CompAcc**

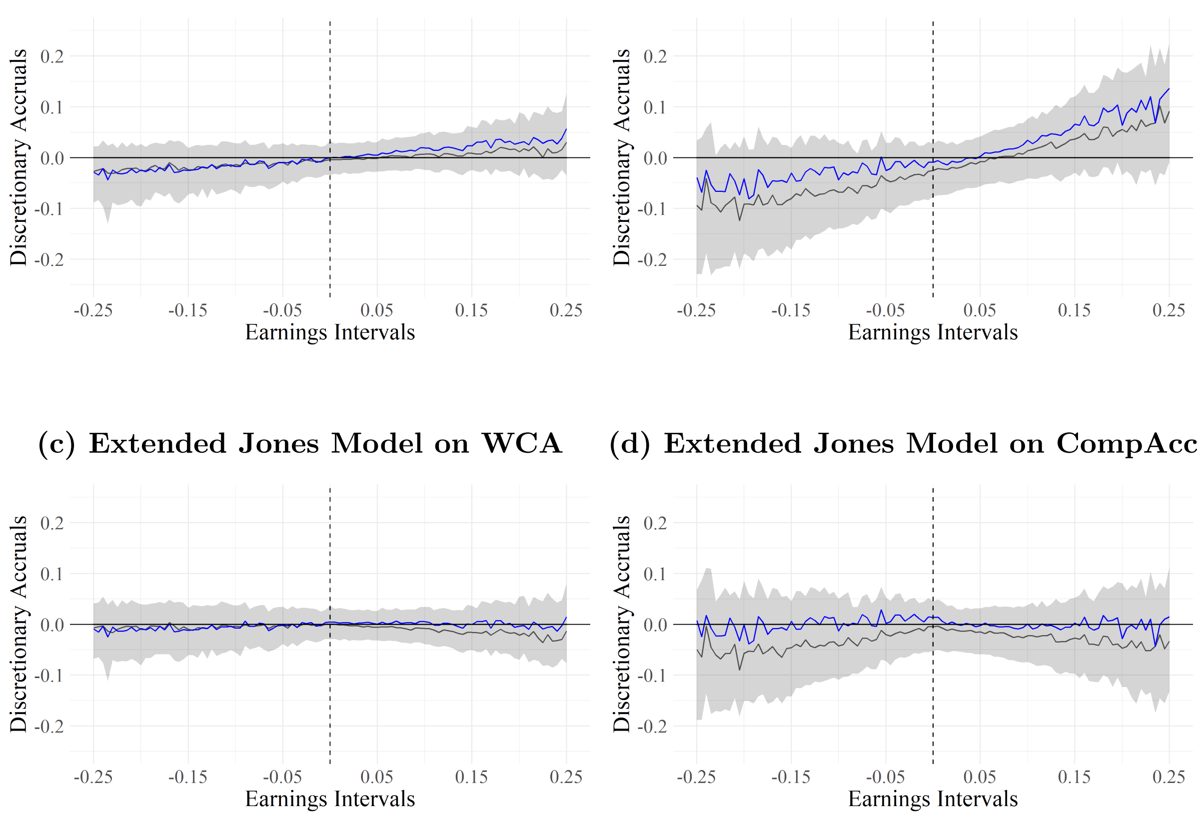

Figure 5.7: Discretionary Accruals per Earnings Interval. Panel (a) shows the discretionary WCA of the modified Jones model as tabulated in Regression (1) of Table 5.5. Discretionary accruals consist of the difference between total accruals (observed values) and normal accruals (fitted values). Since the base model cannot fully account for the correlation of total accruals with earnings, some of the correlation is carried over to discretionary accruals. In other words, since the model does not take into account the impact of earnings on total accruals, firms with high (low) earnings are expected to manage their earnings upward (downward) based on their earnings rather than on individual firm characteristics. Panel (b) presents the discretionary CompAcc of the modified Jones model, as tabulated in Regression (3) of Table 5.5. Panel (c) shows the discretionary WCA of the extended Jones model, as tabulated in Regression (2) of Table 5.5. Unlike Panel (a), no systematic correlation remains, implying that firms do not have discretionary accruals simply because of their earnings level.
Panel (d) presents the discretionary CompAcc of the extended Jones model, as in Regression (4) of Table 5.5.
The black (blue) line represents the mean (median) of the accruals for each earnings interval. The gray area represents the 25th and 75th percentiles.

Panel (b) for CompAcc shows an even higher correlation between discretionary accruals and earnings. Again, the variables of the modified Jones model cannot explain the correlation between accruals and earnings, which is then carried over to discretionary accruals. Depending on the study design, the implications can be important for the results. For example, in further analysis, if we classify firms within the top 1% of discretionary accruals as fraud suspect firm-years, the modified Jones model would assign higher discretionary accruals to firms with positive earnings. This automatically leads to a higher probability of being classified as a fraud-suspect firm-year based on earnings level rather than firm-specific characteristics.

Panel (c) of Figure 5.7 shows the discretionary accruals of the extended Jones model. The added earnings variables account for the positive relationship between WCA and earnings. The mean and median discretionary accruals for each earnings interval are much closer to zero. The extended Jones model does not seem to induce a systematic portion of discretionary accruals depending on the level of reported earnings. Therefore, our selection of firms within the highest 1% of fraud firms is influenced by their relative discretionary accruals rather than the size of their earnings.

Panel (d) shows the extended Jones model for CompAcc. Again, it succeeds in removing the positive correlation between discretionary accruals and earnings of Panel (b). The discretionary accruals of the earnings intervals appear noisier, especially for more extreme positive and negative earnings. The blue line represents the median of the earnings intervals and remains quite close to zero discretionary accruals. The mean tends to be slightly negative. This is because discretionary accruals in this model have a left-skewed distribution with more extreme values for positive discretionary accruals than for negative ones.

Although the extended Jones models successfully reduce the correlation of earnings and discretionary accruals, another problem that all Jones models face becomes apparent. Discretionary accruals appear to be more dispersed in more extreme earnings regions. This can be seen in the gray area delimiting the 25th and 75th percentiles, which increases at more extreme earnings. This occurrence of more extreme earnings can most likely be explained by the characteristics of the ordinary least squares (OLS) regression model. Since there are massively more firms reporting close to average earnings between -5% and +10% of lagged total assets, these earnings regions have a relatively higher weight for which the explanatory variables of the regressions are fitted to maximally reduce the errors (i.e., the discretionary accruals). At the same time, this implies that less importance is attached to the firms with more extreme earnings, which may result in more extreme values of discretionary accruals in the top 1%. An indication of this effect is graphically visible in Panel (d) of Figure 5.7, where the 25% and 75% percentiles have higher values of discretionary accruals for extreme earnings intervals.

Distributional Model

Here we present the results of the distributional model developed in Section 5.3.2. First, we show the histograms of the observed and expected distributions. Then, we compute an indication of firm-level earnings management from the histograms, similar to the discretionary accruals in regression-based earnings management models.

Panel (a) of Figure 5.8 displays the observed distribution of earnings scaled by lagged total assets, Panel (b) presents the expected distribution after subtracting discretionary accruals resulting from the WCA definition from each firm's earnings, and Panel (c) shows the difference in observations between reported and expected earnings.

In Panel (b), there is no or only a negligible kink after subtracting discretionary accruals, contrasting Panel (a). The firm-specific discretionary accruals introduce enough variation so that the discontinuity at the zero threshold disappears. The expected distribution of discretionary accruals resulting from the CompAcc definition can be found in Appendix Table C.1, as both expected distributions are very similar. The only notable difference is that the CompAcc method has a higher frequency for large positive expected earnings. It is important to note that we only use the number of firms in each interval from Panel (b) and not the effective accrual information.

Previous research has estimated the expected distribution using a normal distribution (e.g., S. K. Chen et al., 2010, p. 673). For our research design, we expect to be more

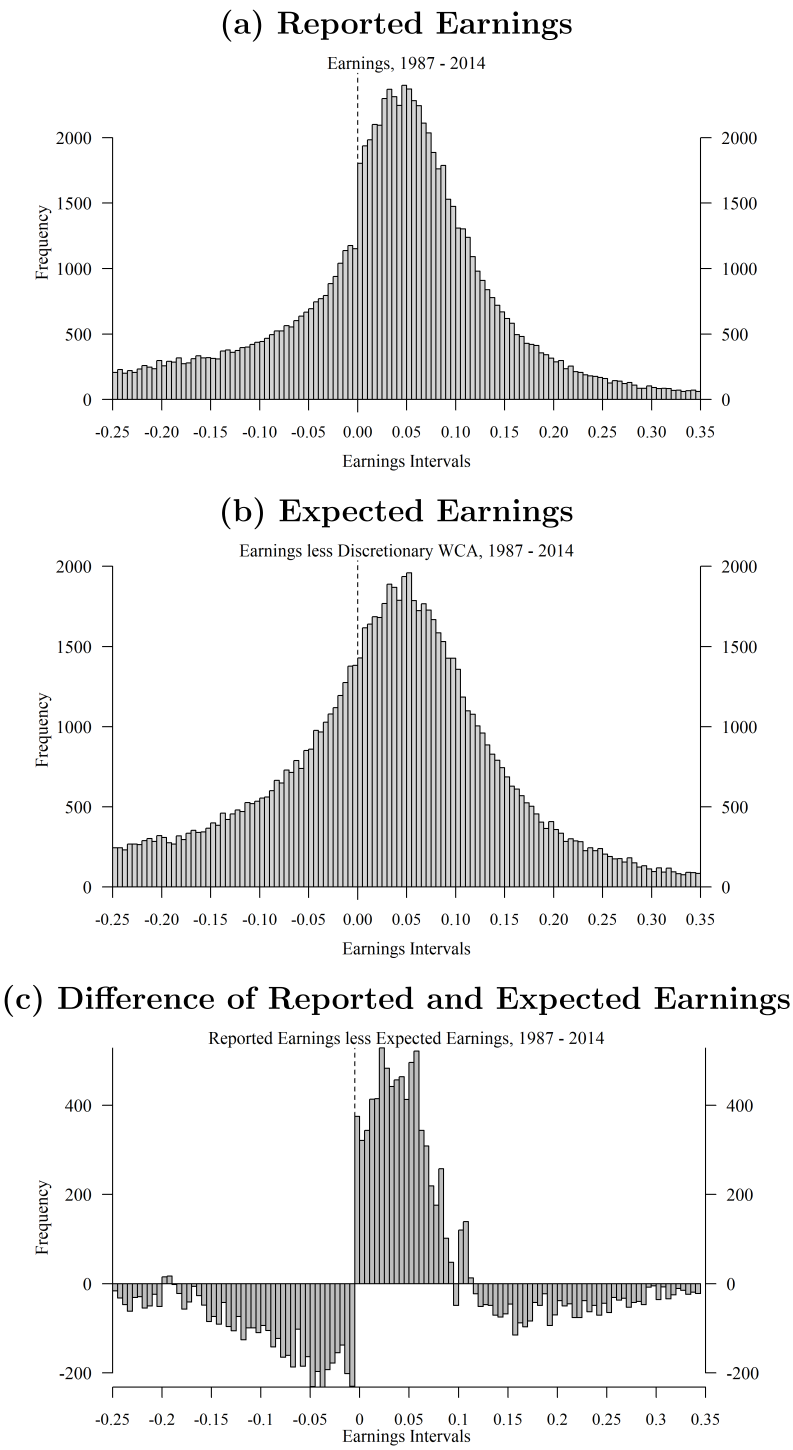

Figure 5.8: Reported and Expected Earnings and the Difference of Both. The figure shows the observed and expected earnings distribution as well as the difference between the two. The observed distribution consists of Earnings before Extraordinary Items (WC #1551) scaled by lagged Total Assets (WC #2999). The expected distribution is the observed distribution less discretionary WCA computed from the modified Jones model. The third plot shows the difference between the observed and expected distribution. The interval width is 0.005 with 120 intervals. The dashed line indicates the zero earnings threshold. The frequency represents the number of observations in each interval.

appropriate to use the distribution of normal accruals as shown in Figure 5.8 Panel (b).[24][25]

From the observed and expected distributions, we now calculate the information on discretionary accruals at the firm level for the distributional model. As described in Section 5.3.2, we subtract the value of the expected returns of an earnings rank from the observed returns of the same rank. Figure 5.9 shows the adjustments of the distributional model derived from the WCA accruals. Positive (negative) values on the y-axis indicate that the reported earnings are higher (lower) than the expected earnings for the same earnings rank. To be consistent with the previous models, we use the discretionary accrual terminology to state that the model indicates that earnings would have been higher or lower in the absence of earnings management.

Discretionary accruals are highest for firms reporting small losses. Here, the estimated extent of earnings management reaches up to 2% of lagged total assets. For example, firms reporting a loss of 5% of their lagged total assets are expected to have increased their accruals by 2% of lagged total assets, so expected earnings without discretionary accruals should have been reported at a value of -7% of lagged total assets. Firms with earnings above 5% of lagged total assets tend to manage their earnings downward. This effect intensifies for firms reporting larger earnings.

In general, the distribution appears to be in line with the earlier literature on earnings smoothing (Fudenberg & Tirole, 1995; Lambert, 1984). Firms that report large positive earnings tend to adjust their earnings downward to produce more average earnings, and may have discretionary reserves for a subsequent year with poorer pre-managed earnings. For firms reporting large losses, the literature has found the effect of big bath accounting, which is to report worse-than-expected earnings. Kirschenheiter and Melumad (2002) show that earnings smoothing towards average earnings and "big bath" accounting can

[24]We use the R library "gamlss" by Rigby et al. (2005) to fit the expected distribution, which is best approximated by a "Generalized beta 2 (i.e., of the second kind)" with the parameters eta mu of 0.16, eta sigma of 2.69, eta nu of -0.11, and eta tau of 1.43.

[25]In unreported results, we estimated the expected earnings by averaging the nearest ten earnings intervals, which also smoothes the kink at zero earnings reasonably well. However, it faces the disadvantage of not showing much difference between the observed and expected distributions in the earnings intervals farther from zero. This would limit our earnings management indicator to the earnings intervals close to zero.

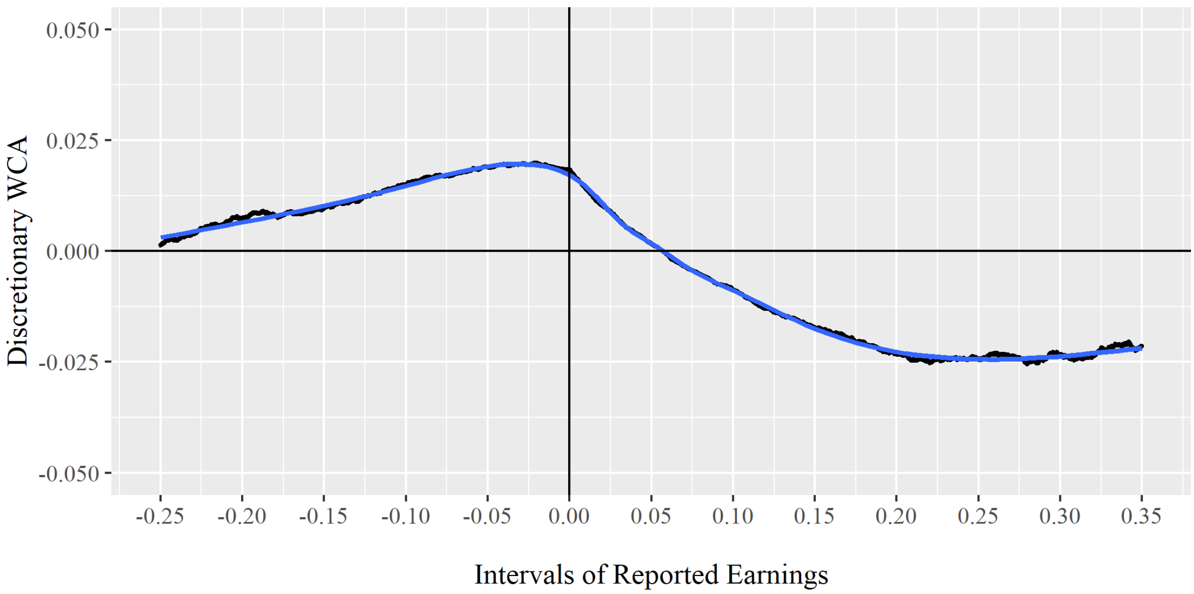

Figure 5.9: Discretionary WCA of the Distributional Model per Earnings Interval. The figure shows the amount of discretionary WCA of the distributional model for each earnings interval of 0.005 of lagged total assets. Positive (negative) values indicate that firms have managed earnings upward (downward) by the indicated amount of lagged total assets on the y-axis. The distribution model indicates positive discretionary accruals for firms reporting losses because of fewer firms than expected.

coexist. However, it is not clear from our figure, which is based solely on earnings information, where the incentive of earnings smoothing ends and the effect of big bath accounting begins, since effects other than earnings might be important factors.

5.6.2 Accounting Fraud Prediction

In this section, we begin by assessing discretionary accruals for firms in the CFRM sample and compare them to non-overstatement firms that are not part of this sample. We test whether there is a significant difference between these two groups and then predict that each earnings group's highest 1% of discretionary accruals overstated their earnings. Finally, we evaluate the precision and sensitivity of this prediction.

Discretionary Accruals of Fraudulent Firms

Panel (a) of Table 5.6 shows the summary statistics for overstated firm-years and Panel (b) for the non-fraudulent firm-years. The Jones-type models, i.e., WCA M for the modified Jones model and WCA E for the extended Jones model, generally indicate that the overvalued firm-years have higher discretionary accruals on average. Discretionary accruals calculated using the modified Jones model are marginally higher than those using the extended Jones model. However, it is unclear whether this effect is due to a better measurement specification of the modified Jones model or the correlation of earnings and discretionary accruals.

The CompAcc calculation method yields much higher average discretionary accruals for overstated firm-years than for WCA. It is not entirely clear why discretionary accruals are higher to such a large extent, but we discuss some potential explanations. Discretionary accruals calculated from CompAcc are more broadly distributed. Since average earnings are higher for overstated firm-years, the stronger correlation of earnings and discretionary accruals for the CompAcc method in the modified Jones model could explain such an effect. However, the extended Jones model includes earnings information and breaks down the correlation for both WCA and CompAcc. Therefore, the above argument does not explain why overstated firm-years computed from the CompAcc of the extended Jones model are higher than in the WCA definition. Potentially, the upward earnings management identified is not limited to WCA but is better captured by a broader definition of accruals, including long-term accounting items. Earnings manipulation related to long-term assets could be of a larger magnitude and therefore have a greater impact on discretionary accruals.

The distributional models show lower discretionary accruals for overstated firm-years. The previous Figure 5.9 indicates that positive earnings are associated mainly with suggested downwards earnings management, while firms with negative earnings are expected to have managed their earnings upwards. Considering that overstated firm-years have above-average earnings could explain the negative level of discretionary accruals.

(a) Discretionary Measure of Overstated Firm-Years

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
WCA M	448	0.016	0.131	-0.425	-0.041	0.051	0.529
WCA E	448	0.014	0.131	-0.475	-0.043	0.051	0.547
WCA D	448	-0.001	0.017	-0.148	-0.010	0.011	0.051
CompAcc M	448	0.050	0.307	-0.815	-0.069	0.104	1.381
CompAcc E	448	0.048	0.307	-0.888	-0.068	0.094	1.480
CompAcc D	448	-0.007	0.076	-0.386	-0.022	0.003	0.705
Earnings	448	-0.003	0.201	-0.868	-0.036	0.088	0.432

(b) Discretionary Measure of non-Overstated Firm-Years

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
WCA M	98,043	0.000	0.118	-0.535	-0.042	0.038	0.608
WCA E	98,043	0.000	0.115	-0.589	-0.045	0.038	0.656
WCA D	98,043	0.000	0.030	-0.534	-0.010	0.011	0.603
CompAcc M	98,043	0.000	0.255	-1.193	-0.091	0.062	1.588
CompAcc E	98,043	0.000	0.249	-1.333	-0.089	0.052	1.649
CompAcc D	98,043	0.000	0.108	-1.188	-0.022	0.006	1.588
Earnings	98,043	-0.033	0.228	-0.868	-0.079	0.086	0.432

(c) Z-Scores

Variable	Z-Score
WCA M	2.58
WCA E	2.26
WCA D	-1.24
CompAcc M	3.44
CompAcc E	3.30
CompAcc D	-1.94
Earnings	3.15

Table 5.6: Summary Statistics of Overstated and non-Overstated Firm-Years. The table shows the summary statistics of the firm-years in which earnings were overstated in Panel (a) compared to the non-overstated firm-years in Panel (b). For overstated firm-years, the average discretionary accruals are higher under the Jones-type models than under the distribution model. Discretionary accruals under the distribution model remain close to zero. For non-overstated firm-years, the average discretionary accruals are zero by construction.
Panel (c) shows the z-scores. The extended Jones-type model for WCA results in a z-score at the 5% level, while the other Jones-type models are significant at the 1% level. For the distributional models, the test score for the CompAcc distribution model is significantly negative at the 5% level.
WCA M represents the WCA from the modified Jones model, as in Regression (1) of Table 5.5.
WCA E represents the WCA from the extended Jones model, as in Regression (2) of Table 5.5.
WCA D represents the WCA using the distribution model.
CompAcc M stands for the CompAcc using the modified Jones model, as in Regression (3) of Table 5.5.
CompAcc E stands for the CompAcc using the extended Jones model, as in Regression (4) of Table 5.5.
CompAcc D stands for the CompAcc using the distribution model.

Panel (b) of Table 5.6 shows the summary statistics for firm-years that were not accused by the SEC of overstating their financial statements. The level of discretionary accruals is zero on average for all firm-years because overstated firm-years account for only 0.45% ($\frac{448}{98{,}491}$) of our total sample. We do not report the table for the entire sample, i.e., for both overstated and non-overstated firm-years, because the summary statistics are nearly identical to those for non-overstated firm-years.

Panel (c) shows the z-scores between the overstated and non-overstated firm-years for each variable, as shown in Equation (5.9). The extended Jones-type model for WCA results in a z-score at the 5% level, while the other Jones-type models are significant at the 1% level. The highest significance arises from the CompAcc definition in combination with the modified Jones models because the discretionary accruals are the highest. For the distributional models, the test score for the CompAcc distribution model is significantly negative at the 5% level.

Highest 1% of Discretionary Accruals as Fraud Prediction

This subsection evaluates the predictive accuracy of the earnings management measures for the firm-years convicted of fraud by the SEC. Table 5.7 presents the frequency of the highest 1% of discretionary accruals of every model grouped by reported earnings of 5% of lagged total assets in the earnings group column "E". The earnings group "-0.90" contains all firm-years with earnings between -0.90 and -0.85 of lagged total assets. Column "N" stands for the number of all reported earnings within this earnings group. A large concentration of reported earnings between -0.15 and +0.15 of lagged total assets appears. The most negative and the most positive earnings groups have more observations because earnings were winsorized at the 1% level. The SEC column contains the percentage frequency of overstated firm-years identified by the SEC relative to all firms for each earnings group. If the overstated firm-years were evenly distributed across earnings, the ratio should be between 0.4% and 0.5% for each earnings group because the frequency of overstatement identified by the SEC is 0.45%. Although the earnings of the fraudulent firm-years are slightly above average earnings, we cannot find a strong clustering of overstated firm-years in some earnings groups. With one exception, all earnings regions contain firms that committed fraud.

			WCA			CompAcc		
E	**N**	**SEC**	**M**	**E**	**D**	**M**	**E**	**D**
-0.90	1,146	0.4	4.1	5.2	65.6	6.9	8.0	86.0
-0.85	506	0.4	2.4	3.2	0.0	5.9	6.7	0.0
-0.80	506	0.0	3.6	4.9	0.0	7.3	8.1	0.0
-0.75	626	0.3	2.4	3.7	0.0	6.2	7.3	0.0
-0.70	621	0.6	2.7	3.7	0.0	5.6	6.6	0.0
-0.65	781	0.1	2.6	3.3	0.0	4.7	5.1	0.0
-0.60	836	0.2	1.9	2.5	0.0	3.3	4.3	0.0
-0.55	942	0.1	1.6	2.3	0.0	3.1	3.6	0.0
-0.50	1,074	0.2	1.7	2.0	0.0	2.7	3.0	0.0
-0.45	1,269	0.4	1.6	2.2	0.0	2.5	3.0	0.0
-0.40	1,512	0.2	1.9	2.6	0.0	2.9	3.3	0.0
-0.35	1,696	0.2	1.2	1.7	0.0	2.1	2.3	0.0
-0.30	1,940	0.5	0.8	1.1	0.0	1.5	1.8	0.0
-0.25	2,337	0.4	1.1	1.3	0.0	1.5	1.7	0.0
-0.20	2,985	0.4	1.0	1.3	0.0	1.1	1.3	0.0
-0.15	3,761	0.5	0.7	1.1	0.0	1.0	1.1	0.0
-0.10	5,481	0.4	0.8	0.9	0.0	0.6	0.7	0.0
-0.05	9,335	0.3	0.5	0.5	2.5	0.5	0.5	0.0
0.00	21,542	0.5	0.4	0.4	0.0	0.3	0.3	0.0
0.05	19,487	0.5	0.5	0.4	0.0	0.2	0.2	0.0
0.10	9,837	0.5	0.7	0.7	0.0	0.4	0.4	0.0
0.15	4,452	0.4	1.5	1.2	0.0	0.9	0.6	0.0
0.20	2,209	0.6	2.0	1.2	0.0	1.1	0.6	0.0
0.25	1,206	0.4	3.2	2.4	0.0	1.7	0.6	0.0
0.30	747	0.8	3.2	1.7	0.0	2.0	0.8	0.0
0.35	447	0.2	5.4	2.7	0.0	4.3	1.8	0.0
0.40	1,210	0.4	8.3	3.4	0.0	5.5	1.3	0.0
	98,491	**0.5**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**

Table 5.7: Prediction of Overstated Firm-Years per Earnings Group. The table shows the frequency of overstated predictions in different earnings groups by 5% of lagged total assets. The -0.90 earnings group includes all firm-years with earnings from -0.90 to -0.85 of lagged total assets. The SEC column represents the frequency of overstated firm-years by the SEC relative to all firms for each earnings group in percent. The columns WCA, WCA E, WCA D, CompAcc, CompAcc E, and CompAcc D represent the percentage frequency of predicted overstatement for each earnings region.

The columns “WCA M”, “WCA E”, “WCA D”, “CompAcc M”, “CompAcc E”, and “CompAcc D” represent the percentage frequency of overstatement prediction for each earnings group, scaled by the number of firms in that earnings group. The Jones-type earnings management models have a relatively large prediction frequency for extreme earnings. For example, in the most negative earnings group of -0.90 to -0.85 of lagged total assets, the Jones-type models show a prediction frequency between 4.1% and 8.0%. This is much more than the overstatement prediction frequency for the earnings group of 0 to 0.05 of lagged total assets, where the prediction frequency is between 0.3% and 0.4%.

The clustering in more extreme earnings regions can most likely be explained by the characteristics of the OLS regression model, as normal earnings regions have more observations and, therefore, a relatively higher weight. In turn, more extreme earnings may result in more extreme values of discretionary accruals located in the top 1%. This effect is also visible in Panel (b) of Figure 5.7, where the 25th and 75th percentiles have higher values of discretionary accruals for extreme earnings intervals.

There are also some differences between the modified and extended Jones models. The figure shows that the modified Jones models have higher discretionary accruals at higher earnings than the extended Jones models for both accrual measures. Therefore, the frequency of predicting overstatements at high earnings is higher in the modified Jones model. Accordingly, if the top 1% of firms with the highest discretionary accruals are predicted with overstatements, firms with higher earnings will naturally be selected more frequently. This would lead to higher fraud prediction accuracy, which is not solely due to the quality of the models, but also to the correlation between discretionary accruals and earnings.

The distributional model shows an even more extreme allocation of predicted overstated firm-years, with a large fraction in the most negative earnings region between -0.90 and -0.85 of lagged total assets. In particular, for CompAcc, all predicted overstated firm-years are assigned to this earnings region. This effect can be explained by some data characteristics. Our regression input variables were all winsorized at the 1%-level. In contrast, we did not winsorize the regression outcomes, namely discretionary accruals. The 1% of most negative earnings are associated with comparatively high negative discretionary accruals leading to high expected earnings. Because we have subtracted expected earnings from reported earnings, this leads to a clustering of observations with a high value of the distributional model.[26] However, the SEC did not identify an abnormally high share of overstated firms in any earnings group. It appears that it is not meaningful to predict a high frequency of firms within an earnings group as overstated earnings.

The bottom line shows values for all earnings groups combined. The prediction frequency for all models corresponds to 1% of all firm-years. In summary, all models have a relatively uneven distribution of the frequency of overstatement prediction. This is because discretionary accruals are not evenly distributed and have extreme values at extreme earnings, as noted in Figure 5.7. To address this issue, we modify our prediction of fraud firm-years. Instead of predicting the highest overall 1% of discretionary accruals as fraud firms, as done in Table 5.7, we predict the highest 1% of discretionary accruals within each earnings group of 5% of lagged total assets as fraud firms. If we reported the prediction frequencies similar to Table 5.7, it would show a fraud prediction frequency of 1% of discretionary accruals for every group.

Panel (a) of Table 5.8 presents the summary statistics of the predicted overstated firm-years, while Panel (b) shows the remaining observations.[27] The Jones-type models result in high discretionary accruals. In particular, the maximum values of discretionary CompAcc exceed lagged total assets. Firms with extremely high CompAcc have in common that they tend to be relatively small firms with a large increase in common equity. Firms with extreme growth in common equity might effectively result in higher accruals than their last year's total assets. Also, less extreme accruals values are predicted as overstatements. The lower end of the highest 1% of discretionary accruals within each earnings group is about 0.5 of lagged total assets for CompAcc and 0.3 of lagged total assets for WCA. For the distributional models, the levels of discretionary accruals for most extreme values are by far lower, with a mean of 0.011 of lagged total assets for WCA and 0.017 for CompAcc. The other 99% of Panel (b) observations are slightly negative.

[26] We do not winsorize discretionary accruals as this would impact them being non-zero on average.

[27] We do not test for significance as in Table 5.6 because discretionary accruals are naturally higher for the top 1% relative to the rest.

(a) Discretionary Accruals of Predicted Overstated Firms

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
WCA M	983	0.467	0.084	0.297	0.393	0.532	0.608
WCA E	983	0.469	0.083	0.298	0.410	0.533	0.656
WCA D	983	0.011	0.062	-0.023	-0.009	0.018	0.603
CompAcc M	983	1.114	0.296	0.551	0.866	1.365	1.588
CompAcc E	983	1.121	0.298	0.540	0.881	1.373	1.649
CompAcc D	983	0.017	0.167	-0.114	-0.011	0.008	1.588

(b) Discretionary Accruals of Predicted non-Overstated Firms

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
WCA M	97,506	-0.005	0.107	-0.535	-0.043	0.036	0.425
WCA E	97,506	-0.005	0.104	-0.589	-0.045	0.036	0.426
WCA D	97,506	-0.001	0.022	-0.534	-0.011	0.011	0.020
CompAcc M	97,506	-0.013	0.219	-1.193	-0.092	0.058	1.143
CompAcc E	97,506	-0.014	0.209	-1.333	-0.090	0.049	1.147
CompAcc D	97,506	-0.009	0.059	-1.188	-0.022	0.006	0.457

Table 5.8: Discretionary Accruals of Predicted Overstated and non-Overstated Firms. The table shows in Panel (a) the summary statistics of the highest 1% of discretionary accruals for each model and in Panel (b) the other 99% of discretionary accruals. In particular, the CompAcc models obtain extremely high values, which could be due to the broad definition of accruals. Model definitions are described in Table 5.6.

After assigning the prediction of the overstated firm-years, we assess the sensitivity and precision of our prediction. Table 5.9 compares the different models in terms of their performance in predicting overstated firm-years. Of the total 983[28] predicted overstated firm-years and 448 effective SEC-identified firms-years, only a few TP arise. If we randomly assigned whether a firm overstated earnings, we would expect a sensitivity of 1% because we selected the top 1% as expected overstated firm-years. In our sample, this would be between four and five TP observations (1% of 448). The precision measures the TP observations divided by the predicted overstated firm-years, of which we would expect a random assignment of 0.46% ($\frac{4.48}{983}$).

The results show that the WCA models predict the sensitivity and precision slightly more accurately than a random assignment. The modified Jones model has six true positives, and the extended Jones model has five correct predictions. The extended Jones

[28]Rounding effects cause the number of predicted overstated firm-years to be slightly less than 1% of all observations, which would be 985.

Model	TP	FP	TN	FN	Sensitivity (%)	Precision (%)
WCA M	6	977	97,066	442	1.34	0.61
WCA E	5	978	97,065	443	1.12	0.51
WCA D	7	976	97,067	441	1.56	0.71
CompAcc M	6	977	97,066	442	1.34	0.61
CompAcc E	6	977	97,066	442	1.34	0.61
CompAcc D	3	980	97,063	445	0.67	0.31

Table 5.9: Sensitivity and Precision of Overstated Firm-Years Prediction. The table shows the number of correctly and incorrectly predicted overstated firm-years. All Jones-type models lead to slightly higher sensitivity and precision than expected from a random assignment, i.e., 1% for sensitivity and 0.46% $\frac{4.48}{983}$ for precision. The distributional model leads to better than random results for WCA and worse results for CompAcc. True positives are overstated firm-years identified by the SEC that are also included in the top 1% of discretionary accruals, false positives are firms included in the top 1% of discretionary accruals but not identified as overstated by the SEC, true negatives are firm-years not identified by the SEC and not included in the top 1% of discretionary accruals, and false negatives are overstated firm-years identified by the SEC that are not included in the top 1% of discretionary accruals. $Sensitivity = \frac{TP}{TP+FN}; Precision = \frac{TP}{TP+FP}$.

model fails to outperform the modified Jones model, although it includes additionally earnings and has a higher explanatory power in the regression model. The distributional model calculated from WCA performs best. Even though it does not contain firm-specific earnings information, it outperforms the other WCA-based models. Nevertheless, it is important to note that the differences between the models and the number of true positives are relatively small, so any of these results could be due to chance. We perform a significance test but find no evidence of significance.[29]

The results for the CompAcc Jones-type models are similar to those for WCA. Regardless of the accrual definition, the models perform similarly, as both regression-based models correctly predict six overstated firm-years. The distributional model computed from CompAcc is the only model that performs worse than a random prediction. Again, no value is statistically significant.

Bao et al. (2020) also select the highest 1% of their earnings management indicator to predict fraudulent firms using a machine learning approach. The results show a higher sensitivity and precision. We identify three possible reasons why Bao et al. (2020) have more precise results than we do. First, their models use more explanatory variables. Bao et al. (2020) use different models with either 24 raw data variables as in the study of

[29]The significance test calculates the ratio of the correctly predicted fraud firm-years of all firm-years to the ratio that would be expected under a random assignment.

Cecchini et al. (2010) or 11 financial ratios as in Dechow et al. (2011). We did not include these ratios because we aim to evaluate the quality of the Jones-type models. Second, the method of Bao et al. (2020) is different. They define a training period that includes information on past cases of fraudulent firms and then predict future fraudulent firms. Our models, in contrast, estimate the overstated firm-years without using information from the AAER database. Moreover, Bao et al. (2020) consider the serial fraud problem in most of their analyses, while we ignore it. Finally, Bao et al. (2020) use flexible and nonlinear machine learning models.

Firm Characteristics of Correctly Predicted Fraud

We manually assess the firms where fraudulent overstatement was correctly predicted. We also summarize the common characteristics of these firms to better understand how they differ from the others. For the above results, all models combined correctly predicted 33 firm-years, as indicated in Table 5.9. Some of these firm-years were predicted by multiple models, resulting in 18 unique firm-years. Table C.1 in the Appendix shows the key accounting variables of the correctly predicted overstated firm-years.

The main points can be summarized as follows. First, the correctly predicted overstated fraud-firms are quite diverse in terms of firm size and market capitalization. The largest correctly predicted firm is Microsoft in 1997 by the Jones-type regression models using the WCA and CompAcc methods. Second, 12 of 18 firms were accused of revenue management. Common examples include improper revenue recognition, fictitious revenue, or inflated revenue through channel stuffing. Third, the Jones-type models predict nearly identical firms. It does not affect the results whether earnings are included in the explanatory variables.

This also applies to the distributional model, regardless of whether the extreme accruals are computed from WCA or CompAcc. All firm-years correctly predicted in the main results by the WCA distributional model are also correctly predicted by the CompAcc distributional model.

Robustness Test

Firms with extreme performance and extreme earnings might have a strong influence on our results. However, most earnings are between -0.15 and 0.15 of lagged total assets. Within this range, discretionary accruals are more stable. Therefore, we make the following changes. First, we exclude all firms that reported earnings outside the range of -0.15 and 0.15 of lagged total assets. Second, we remove all earnings groups and pool all data, independent of their earnings level. Then, we rerun our entire analysis, including the Jones-type regressions and discretionary accruals computations. Finally, we predict each model's top 1% of discretionary accruals as overstated firm-year.

Table 5.10 shows the discretionary accruals for firms that overstated their earnings in Panel (a) and those that did not in Panel (b). Panel (a) shows that the Jones-type models' discretionary accruals for this subsample are lower for firms that overstated their earnings compared with the entire sample. The average discretionary WCA of the modified Jones model is 0.008, compared with 0.016 for the full sample. This effect is even more extreme for discretionary CompAcc. While they reached averages of 0.050 of lagged total assets in the modified Jones model for the full sample, they have now dropped to 0.030 of lagged total assets. Less extreme earnings appear to be associated with less extreme discretionary accruals. For the distributional models, discretionary accruals are exactly zero, regardless of whether they are calculated from WCA or CompAcc. The z-scores are lower compared with the full sample due to the lower average discretionary accruals and the smaller number of observations. Only the Jones-type models for CompAcc result in significant values at the 1% level. For the distributional model, the z-scores are exactly zero because there is no difference in the mean.

Table 5.11 shows the TP, FP, TN, and FN for the cases where the 1% of discretionary accruals are predicted to be overstated firm-years. Because the total sample is reduced to 69,443 firm-years, the absolute numbers of TP, FP, TN, and FN are not directly comparable to the main results. The sensitivity and precision are higher for the Jones-type models than in the main results. The modified and extended Jones models for the

(a) Overstated Firm-Years for Earnings From -0.15 to 0.15

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
WCA M	337	0.008	0.095	-0.308	-0.039	0.044	0.362
WCA E	337	0.008	0.094	-0.299	-0.037	0.041	0.344
WCA D	337	0.000	0.039	-0.239	-0.014	0.011	0.228
CompAcc M	337	0.030	0.183	-0.474	-0.054	0.074	0.744
CompAcc E	337	0.029	0.180	-0.465	-0.052	0.065	0.785
CompAcc D	337	0.000	0.101	-0.396	-0.024	-0.001	0.641
Earnings	337	0.034	0.064	-0.139	0.007	0.078	0.145

(b) Non-Overstated Firm-Years for Earnings From -0.15 to 0.15

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
WCA M	69,106	0.000	0.086	-0.358	-0.037	0.030	0.430
WCA E	69,106	0.000	0.085	-0.360	-0.038	0.030	0.440
WCA D	69,106	0.000	0.040	-0.345	-0.013	0.015	0.418
CompAcc M	69,106	0.000	0.151	-0.634	-0.069	0.043	0.908
CompAcc E	69,106	0.000	0.149	-0.616	-0.065	0.036	0.935
CompAcc D	69,106	0.000	0.098	-0.564	-0.024	0.005	0.895
Earnings	69,106	0.030	0.066	-0.139	-0.005	0.077	0.145

(c) Z-Scores

Variable	Z-Score
WCA M	1.54
WCA E	1.56
WCA D	0.00
CompAcc M	3.00
CompAcc E	2.95
CompAcc D	0.00
Earnings	1.14

Table 5.10: Summary Statistics of Overstated and non-Overstated Firm-Years for Earnings From -0.15 to 0.15. The table shows the summary statistics of the firm-years in which earnings between -0.15 and 0.15 of lagged total assets were overstated in Panel (a) compared with the non-overstated firm-years in Panel (b). For overstated firm-years, the average discretionary accruals are higher under the Jones-type models than under the distributional model. For non-overstated firm-years, the average discretionary accruals are zero by construction. Panel (c) shows the z-scores. The Jones-type models for CompAcc result in significant values at the 1% level. The model definitions are described in Table 5.8.

WCA and CompAcc definitions all produced six correctly predicted overvalued firm-years, except for the extended Jones model under the CompAcc definition, which produced five correct predictions. Again, no model is significantly better than a random assignment. A mixed picture also emerges for the distributional models. Both, the WCA-based calculation with three correct predictions and the CompAcc-based calculation with two correct predictions, perform worse than would be expected from a random assignment.

Model	TP	FP	TN	FN	Sensitivity (%)	Precision (%)
WCA M	6	689	68,417	331	1.78	0.86
WCA E	6	689	68,417	331	1.78	0.86
WCA D	3	692	68,414	334	0.89	0.43
CompAcc M	6	689	68,417	331	1.78	0.86
CompAcc E	5	690	68,416	332	1.48	0.72
CompAcc D	2	693	68,413	335	0.59	0.29

Table 5.11: Sensitivity and Precision of Fraud Firm-Years Prediction for Earnings From -0.15 to 0.15. The table shows the number of correctly and incorrectly predicted fraud firms. All Jones-type models lead to slightly higher sensitivity and precision than expected from a random assignment. TP, FP, TN, FN, sensitivity, and precision are defined in Table 5.9.

5.7 Discussion and Conclusion

The discontinuity literature has identified a kink at zero earnings with more firms reporting slightly positive earnings than slightly negative ones and concluded that firms manage earnings upward to reach the desired earnings threshold. One would expect earnings management models to capture this effect by obtaining higher discretionary accruals at zero earnings. However, earnings management models show no relevant evidence of upward earnings management.

Based on these conflicting findings, we derive a distributional model of firm-level earnings management based on the differences between expected and observed earnings. Additionally, we extend the modified Jones model by adding earnings information to account for the correlation of earnings with discretionary accruals. For our analysis, we evaluate six different models: the modified Jones model, the extended Jones model that considers earnings information, and the distributional model. For each model, we run our calculations with two accrual calculations, WCA and CompAcc.

We assess the quality of the different earnings management models using two separate methods. The first method determines the extent of earnings management for each firm in the AAER fraud sample. We compare the mean discretionary accrual values of overstated to non-overstated firm-years and measure the significance levels. We find that the extended Jones model, which includes earnings, leads to increased explanatory power and, more importantly, prevents discretionary accruals from correlating with earnings. We test the significance of discretionary accruals for overstated firm-years relative to non-overstated firm-years and find significant results for the Jones-type models, while the results for the distributional models are mixed.

The second method predicts the 1% highest discretionary accruals within each earnings group of 5% of lagged total assets of each model to be included in the AAER fraud sample. This is based on the idea that fraud is an extreme case of earnings management that should most likely correspond to the most extreme observations of a valid measure. We count these firm-years and compute the sensitivity and precision. The results show that the prediction precision and sensitivity for Jones-type models are slightly higher than a random selection. It should be noted, however, that the number of TP is generally very small, so caution should be exercised in interpretation. The results of both methods suggest that the Jones models capture some of the earnings management, but certainly not all of it. No model is significantly better than a random assignment. Under random selection, the distributional model does not yield higher discretionary accruals or true positives for overstated firm-years than expected. Possibly, this is best explained by the fact that the distributional model expects that in some earnings regions, firms engage more frequently in earnings management than in others. However, the firms found guilty of fraud by the SEC are not clustered in any particular earnings region.

It remains difficult to draw a pertinent conclusion. However, the few fraudulent firm-years that were correctly predicted involved firms that massively exaggerated their earnings, which was correctly captured by the explanatory variables. The overall sensitivity and precision tend to remain slightly higher than for a random prediction. The Jones-type models may capture a marginal fraction of overstated firm-years for those firms who excessively managed their earnings through revenue since this is one of the explanatory variables. However, this is not sufficient that the TP would deviate significantly from a random prediction.

The results are broadly consistent through our different result sections. Section 5.6.2

compares the levels of discretionary accruals of firms that overstated their earnings to the other firm-years. There, we find that the discretionary accruals of overstated firm-years are positive and significant for all Jones-based accrual models, consistent with a higher TP overstatement prediction for the highest 1% of discretionary accruals of this results section. However, discretionary accruals for the overstated firm-years are particularly high under the CompAcc method but do not lead to a higher overstatement prediction. This is possibly due to the generally broader distribution of CompAcc.

Our extended Jones model mitigates the problematic correlation of earnings with discretionary accruals. For future studies, we recommend including such an earnings variable or another variable that is closely correlated with earnings. In general, it appears that earnings management incentives may not be linear. One avenue of research might be to better understand how managers' incentives depend on pre-managed earnings. Some experimental studies on the "cookie jar" accounting or "big baths" could provide a causal link.

Chapter 6
PREDICTING ACCRUALS

6.1 Introduction

The long-standing and widely used standard discretionary accrual models are easy to apply, but they are also increasingly criticized for their characteristics. According to the standard theory, deviations from the fitted values of earnings management regression models are referred to as discretionary accruals. These discretionary accruals are interpreted to mean that managers actively increase or decrease a firm's earnings. Ball (2013) notes that it is unrealistic to expect that accruals deviations from the fitted values solely originate from earnings management. Due to potentially correlated omitted variables, other factors may be responsible for the unexplained variation unrelated to active earnings management.

Additional variables increase the explanatory power and shrink the unexplained variation. For example, Larson et al. (2018) show that the growth in the number of employees is an important variable in explaining comprehensive accruals (CompAcc). Further previously unconsidered variables could help explain a higher proportion of the variation in accruals. However, even if the models accounted for all relevant effects, accounting relationships are partially nonlinear, which would still expose them to measurement error (Banker et al., 2020).

The objective of this chapter is to improve the understanding of accruals. To achieve this goal, we first incorporate an extensive set of explanatory variables from the Refinitiv Worldscope database to predict next year's accruals using linear models. We expect a broad set of variables helps to reduce the unexplained variation. Second, we use supervised machine learning models: the least absolute shrinkage and selection operator (LASSO), random forest, and support vector machines (SVMs). With these more sophisticated measurement methods, we expect to attain a higher prediction accuracy than the ordinary least squares (OLS) regression.

The reason we predict accruals for one year in advance rather than the same year's accruals is twofold. First, predicting next year's accruals has been little explored in academic research. However, identifying key indicators of future accruals is important not only for academia but also for the public. For stakeholders outside academia, such as investors and financial analysts, predicting a firm's accruals can be useful in light of upcoming financial statement releases. The second reason is that using a large set of explanatory variables for the accruals of the same year would lead to multicollinearity. This is because accruals are calculated from variables in the financial statements themselves, which are already included in the explanatory variables.

The results indicate that a broad set of variables increases the predictive accuracy over the standard OLS models. Further, machine learning models outperform OLS regressions in most, but not all, cases because they avoid overfitting. The most accurate results are predicted with a LASSO model, although a substantial prediction error remains that could be reduced by some tuning of model parameters.

The chapter proceeds as follows. Section 6.2 reviews the literature and highlights the weaknesses of the current models. Section 6.3 presents the machine learning models. In Section 6.4, we select our variables and firms. In Section 6.5, we present the results and robustness tests. Section 6.6 concludes.

6.2 Literature

Discretionary accrual models are widely used in various research areas. They are intuitive to understand, easy to implement, and use only a few commonly available variables without imposing high data requirements. In general, the models provide a simple method for evaluating a wide range of hypotheses. Examples of research areas in which Jones-type earnings management models are used involve topics in corporate governance, including ownership structure, board structure, mergers & acquisitions, diversity measures, chief executive officer (CEO) narcissism, earnings informativeness, investment efficiency, liquidity measures, and more.

However, the simplicity of earnings management models comes at a cost. Often, the explanatory power of the models is low. The most common Jones-type model uses the change in revenue and gross property, plant & equipment (PPE) as explanatory variables, which explain only a small order of magnitude. In earnings management research, the unexplained deviation from the fitted values is defined as discretionary accruals. However, it is unclear whether this deviation truly results from earnings management or is more likely due to the inability of the model to capture other effects that do not result from earnings management.

We believe it is worth exploring whether additional accounting variables could improve existing models. We begin by discussing issues related to classical earnings management models. In particular, we address two main issues. First, we discuss the nonlinearity of accounting items. Such nonlinearities have been found in sales growth (Banker & Chen, 2006; Banker et al., 2020), which is a key explanatory variable in Jones-type earnings management models. Other accounting relationships, such as accruals and cash flow, have also been found to be nonlinear (Ball & Shivakumar, 2006). Second, we discuss the added value of additional variables and possible improvements in accuracy using machine learning methods. Both points lead to the conclusion that newer statistical methods could help find the relevant variables that improve existing earnings management models.

6.2.1 Nonlinearities

Previous literature has mentioned that the process of accrual formation can be nonlinear (Dechow et al., 1995; Jeter & Shivakumar, 1999; Kothari et al., 2005). Recent studies have demonstrated several nonlinear relationships between accounting items and accruals.

One of these is firm growth, which is a well-known source of inaccuracy in linear models. Typical Jones models account for growth by including the change in revenue and PPE. In addition, the modified Jones model, as in Dechow et al. (2012), subtracts the change in accounts receivables from the change in revenues since higher sales lead to higher accounts receivables over the same period. However, Larson et al. (2018) present two arguments why the modified Jones model's explanatory variables are insufficient to account for firm growth. First, not all accounting items grow at the same pace as the firm. For example, PPE is likely to grow more slowly than other accounting items due to economies of scale. Current linear models cannot account for this change in correlation. Second, net capital intensity varies massively in cross-sectional data. In rare cases, operating liabilities exceed operating assets, leading to a negative net capital requirement.

In their study, Larson et al. (2018, p. 840) solve this problem by including employee growth as a proxy for a firm's operational growth. They see the advantage in the fact that it is a non-financial variable. They assume that net capital growth increases in proportion to employee growth. Consequently, they multiply the employee growth rate by the lagged net capital intensity.

Banker et al. (2020) also assess the nonlinear relationship between revenue growth and accruals. Similar to Larson et al. (2018), they find that revenue changes due to operational business decisions are nonlinearly related to accruals. The forced linearity in popular earnings management models overestimates discretionary accruals for moderate revenue changes and underestimates discretionary accruals for extreme revenue changes. Linear control variables for sales growth also fail to account for nonlinearity. They solve the specification error by using a flexible spline correction to capture the operational portion of revenue growth. This method leads to a significant reduction in Type I errors and an improvement in test strength.

Another example of nonlinearities is the relationship between cash flows and accruals. Ball and Shivakumar (2006) explain the nonlinearity by the asymmetrically faster recognition of losses compared with gains, in agreement with the results of Basu (1997). Moreover, they show that separating positive and negative cash flows leads to higher explanatory power. The commonly used Dechow and Dichev (2002) model uses cash flows, one lead, and one lag of cash flows as explanatory variables. In the replication of Ball and Shivakumar (2006), the explanatory power of the Dechow and Dichev (2002) model is 7.9%. Using the same data and additionally partitioning the cash flows by a dummy for negative cash flows and the dummy's interaction term with the cash flows, the explanatory power increases to 11.8%.

W. Chen et al. (2018) highlight a problem in accounting research with discretionary accruals. They provide an extensive list of studies using the two-stage regression procedure to evaluate a hypothesized explanatory variable of earnings management (W. Chen et al., 2018, p. 753). In the first stage, several independent variables are regressed on total accruals. The unexplained variance (i.e., the residual error term of the regression) is then referred to as discretionary accruals. In the second stage, some studies use discretionary accruals as the dependent variable and one or more independent variables to test a hypothesis, including some control variables to possibly establish a causal relationship between discretionary accruals and the hypothesis variable.

In most of these studies, the independent variables in the first regression are not included in the second regression. Researchers may believe that the effect of the independent variables has already been accounted for in the first regression. Therefore, these control variables would not need to be included in the second-stage regression. However, contrary to this belief, if the variables from the first step regression are not included in the second regression, the second stage regression can result in biased estimates and t-statistics towards zero if the hypothesis variable of the second stage correlates with any of the explanatory variables of the first stage. Accordingly, Owens et al. (2017) found that idiosyncratic shocks reduce the accuracy of first-stage linear models for accrual models.

The proposed solution of W. Chen et al. (2018, p. 756) is to avoid using two-stage regression models in earnings management research. Simply including all desired explanatory variables in the model explaining total accruals would not yield econometric issues of over- or underestimated coefficients.[1]

McMullin and Schonberger (2020) evaluate whether accruals follow a linear or non-linear process. They find several deviations from linear relationships, especially for size, performance, and growth. A multivariate matching approach, which is an advanced approach to matching firm performance, offers improvements over the linear regression model because it accounts for nonlinearities. They suggest additional accrual determinants that

[1]In their chapter 5, they propose other more advanced and econometrically correct solutions for unbiased two-stage regression estimates.

have not yet been found and "leave progress on identifying a comprehensive set of accrual determinants for future research" (McMullin & Schonberger, 2020, p. 115).

The above studies show that the relationship between financial statement items may not be linear. Even though the above studies have contributed to the research by finding nonlinear relationships, there may still be numerous undiscovered nonlinear relationships.

6.2.2 Additional Variables

The modified Jones model relies on very few financial accounting variables. Sales, accounts receivable, and gross PPE are relevant and obvious variables, but other accounting items can also increase the explanatory power. In general, leading and lagging cash flows are useful to increase explanatory power since accruals must reverse at some point (Dechow et al., 2012). Commonly used models with leading cash flow data include the model of Dechow and Dichev (2002) and the model of McNichols (2002). The McNichols (2002, p. 66) model adds a cash flow lead and a lag to the modified Jones model, increasing the explanatory power from 7.3% to 30.1% without accounting for nonlinearities.

In addition, less obvious accounting variables can increase explanatory power. For example, Larson et al. (2018, p. 856) use employee growth, net comprehensive assets, and up to eight years of leading and lagging cash flows to explain CompAcc, achieving a remarkably high explanatory power of 50.9%, compared with single-digit explanatory powers in typical Jones models. However, excessively long time series periods limit the dataset severely. In the example of Larson et al. (2018), all firms must have at least 17 years of consecutive financial statement data to be included in the sample, which consists of eight leads and lags plus the current year. In their study, the base model contains 76,541 firm-years, while the model with eight leads and lags contains only 21,084 firm-years, which might introduce a sample bias.

Some studies add control variables to account for potentially correlated effects that would interfere with the measurement effect. For example, the logarithm of a firm's total assets is often used because larger firms have proportionately lower accruals (Watts

& Zimmerman, 1978). The firm's age in years provides information about its stage in the business cycle (Callen et al., 2008). Industry revenue and margin growth control for industry-specific characteristics (Stubben, 2010). In addition to sales growth, asset turnover is also used to control for firm growth (Barth et al., 2008). The ratio of total debt to total assets considers the motivation for entering into debt contracts (DeFond & Jiambalvo, 1994). Longer auditor tenure is negatively associated with discretionary accruals (J. N. Myers et al., 2003). Firms audited by one of the Big 4 auditors exhibit lower levels of earnings management (Becker et al., 1998). The annual change in equity market capitalization divided by a country's gross domestic product should account for country-specific differences in reporting quality in international studies (Haw et al., 2004).

Other control variables include the percentage change in common stock and the percentage change in total liabilities to account for external financing (Ipino & Parbonetti, 2017). Similarly, the logarithmic number of financial analysts following a firm is assumed to reduce earnings management, as firms that engage in earnings management could potentially be detected by analysts (Ipino & Parbonetti, 2017). Finally, more recent studies often control for year, country, and industry fixed effects, where possible.

The fraud prediction literature uses a wide range of variables, including some financial ratios. For example, Bao et al. (2020) predict fraud firms in the Accounting and Auditing Enforcement Releases (AAER) sample. The most accurate model uses 28 raw data items from the Compustat database. The machine learning algorithm also provides the relative importance of variables that increase the predictive power of the test data. The most important variables are common shares outstanding, current assets, and sale of common and preferred stock. These variables may not have been particularly intuitive before, but they have been shown to be most important for predicting firms committing fraud.

The increased explanatory and predictive power of additional variables motivates our study to include a broad dataset. Other previously-used financial statement items may significantly increase the explanatory power of next year's accruals.

6.2.3 Avoiding Multicollinearity

It is impossible to include all financial statement items as explanatory variables in a regression, even though it might appear tempting. The reason is that the explained variable, the change in accruals, is calculated from financial statement items. Thus, if components of the explained variable are used as explanatory variables, the regression will find these components and use only them to explain accruals, assigning a weight of zero to the other variables. This effect can be avoided by using next year's accruals as the explained variable rather than the same year's accruals.

Predicting next year's accruals is important not only to academia but also to private sector stakeholders such as analysts or investors. For these stakeholders, a prediction of the firm's accrual level may be more valuable, given the upcoming publication of the financial statements, than the largest possible explanatory R^2, which requires leading data.

We could not find any study that explains or predicts the accruals of the next year using explanatory accounting variables of the actual year. However, some studies evaluate the predictive power of accruals on other accounting items. For example, Barth et al. (2001) find that current accruals and cash flows have predictive power for future cash flows, which even increases when accruals are disaggregated. Barth et al. (2016) evaluate the information content of accruals for cash flows. They find that accruals can have explanatory power on either past, current, or future cash flows, depending on the characteristics of the accrual components. For example, an increase in accounts receivable, and hence accruals, leads to an increase in cash flows in the following period.

Some studies examine whether cash flows or earnings are more useful for predicting future cash flows or earnings. Subramanyam and Venkatachalam (2007) measure the importance of accrual-based earnings and cash flows for the intrinsic value of equity. Consistent with Barth et al. (2001), they find that operating cash flows are a better predictor of future cash flows than accrual-based earnings. Additionally, earnings have a higher explanatory power for future earnings than cash flows. Ball and Nikolaev (2020) find that earnings are a better predictor of future cash flows than actual cash flows. B. Lev et al. (2010) evaluate out-of-sample cash flows and earnings prediction. They find that accounting estimates do not improve the prediction of cash flows when compared with past cash flow items, but accruals have some predictive power for the following year's earnings.

The study of Callao and Jarne (2010) applies an out-of-sample prediction to discretionary accruals to measure whether there is a change in accrual earnings management from the pre-adoption of the International Financial Reporting Standards (IFRS) to the post-adoption period of the IFRS. In their research design, they use the years 1997 to 2002 and estimate the coefficients of the modified Jones model, including some control variables. Then, they use the coefficients from the first step to predict total accruals for the years 2003 to 2006. Finally, they compare the predicted estimates with the actual total and discretionary accruals for these years. They find that discretionary accruals for firms in various European countries increased after the introduction of IFRS.

6.2.4 Machine Learning Models in the Earnings Management Literature

Given the above literature, using statistically more powerful and flexible models appears intuitive. Some machine learning studies have included such models using supervised and unsupervised machine learning models.

The study by Höglund (2012) evaluates the usefulness of neural networks for the out-of-sample prediction of working capital accruals (WCA). All neural networks outperform the standard OLS regression.[2] However, the study includes only U.S. manufacturing firms with Standard Industrial Classification (SIC) codes from 2000 to 3990 for the years 2006 and 2007, which restricts the sample to 2,032 firm-years. Further, only the explanatory variables of Jones-type models are used. Höglund (2012) concludes that neural networks

[2] Haga et al. (2015) conduct a similar study, with the main difference being that they use real earnings management instead of accruals earnings management and come to similar results.

are more challenging to implement, and interpreting the results is more complicated. He predicted that researchers would continue to use linear models for the sake of simplicity, even if they are inadequate.

Some studies use time series models. The disadvantage is that this severely restricts the sample size because the values must not be missing. For example, Höglund (2013) evaluates the performance of standard time-series OLS and fuzzy linear regression models. He uses the Bureau van Dijks Orbi database, which contains data from 2003 to 2012. Of 8,366 publicly traded U.S. firms in his baseline sample, only 689 firms are included in the final dataset. Most firms dropped out because the data were incomplete within this 10-year period. Similarly, the study by Höglund (2015) evaluates the performance of standard accrual models from self-organizing maps, which are similar to a neural network that aims to reduce dimensionality. The firms in the sample are required to be fully available during the entire 12-year period, restricting the available firms to 934. To avoid such severe limitations in the selection of firms, we use repeated cross-sectional data, requiring only one lag for calculating accruals and one lead for predicting future accruals.

Several other studies apply machine learning methods to earnings management research. However, some studies have limited time horizons or are industry or country specific. Examples include the samples of Taiwanese listed electronic firms reported quarterly in the Taiwan Economic Journal from 2008 to 2012 (F. H. Chen & Howard, 2016), Taiwanese listed biotechnology firms from 2008 to 2012 (F. H. Chen et al., 2015), or data from the Tehran Stock Exchange from 2003 to 2011, yielding 94 firms in the final sample (Namazi & Maharluie, 2015).

The above studies differ from our study in several aspects. We use a larger sample and more financial statement items than previous studies. In addition, we aim to predict accruals for the next year rather than accruals for the same year.

6.3 Methodology

This section introduces the concepts of the "machine learning" algorithms, which have become increasingly important in recent years. Not only the massive increase in available data, but also the simultaneous increase in computer power has led to this development. On the one hand, it seems appealing to extend simple regression models with more explanatory variables and, on the other hand, to use more flexible and accurate but also more resource-intensive statistical models to obtain more precise results.

Machine learning approaches can be used to predict either a continuous variable or a binary variable. Since our goal is to predict accruals more accurately, we use the methods only for continuous numerical variables. We focus on the most common supervised machine learning methods, i.e., the LASSO, tree-based methods, and SVM. Several sources explain the algorithms in detail. We mainly use the book by James et al. (2021), if not stated otherwise, and summarize the main points in this section.

6.3.1 Quality of fit Measurement

It is intuitive that more variables lead to a higher R^2, which is often used as a quality measure for earnings management models. Therefore, it is tempting to increase the number of explanatory variables. However, when too many variables are used, regression models tend to overfit, which is the case when the explanatory variables are overly adapted to the sample data but would not prove to be a good fit for the population data.

This section defines the mean squared error (MSE), which is the most common method for evaluating the performance of models. The MSE refers to the squared difference between the observed and the fitted values. In traditional earnings management models, the regressions are run in-sample, without splitting the train and test sample, referred to as the train MSE.

Because we use a large set of variables, our model would tend to overfit the underlying data, and the coefficients would provide an inaccurate prediction for data that are not

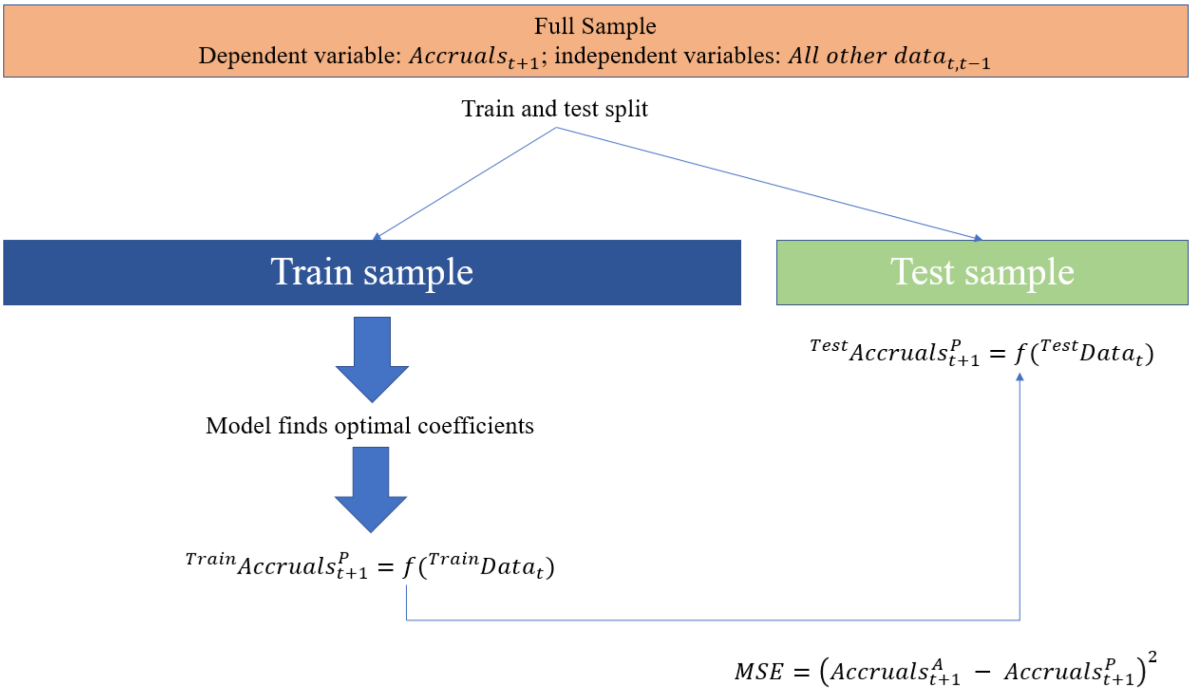

Figure 6.1: Train-Test Split. The figure shows the train-test split of the full sample. The training sample is used to find the coefficients that reduce the in-sample error. Machine learning methods additionally have a mechanism that prevents from overfitting. The coefficients of the training sample are then used to predict the out-of-sample test data. The mean of the squared error of predicted values of the test sample and the actual test values is as the test MSE. Superscript A stands for Actual and superscript P stands for Predicted.

part of the sample. Therefore, out-of-sample research designs are used to ensure that the results are valid for data outside the training sample. Figure 6.1 illustrates the process of the train-test-split, the in-sample validation, and the prediction of test data. The full sample is first split into training and test data. Within the training sample, the model finds the optimal coefficients to reduce the in-sample squared residuals. Unlike the standard OLS, machine learning methods typically account for overfitting with an in-sample validation procedure. The obtained coefficients of the training sample are then used to predict next year's accruals. The squared deviation of the actual and predicted values is referred to as the test MSE, as shown in the following equation:

$$Test\ MSE = \frac{1}{n}\sum(y_0 - \hat{f}(x_0))^2, \tag{6.1}$$

where the y_0 refers to the actual observed next year's accruals of the test sample and x_0 to the predicted accrual values of next year using training sample data. Because all models aim to minimize the MSE, it is our natural evaluation criterion for our models.

Unfortunately, the interpretation of the MSE can be unintuitive because it reports squared values. To facilitate interpretation, we also report the square root of the MSE ($\sqrt{MSE}$) and the mean absolute error (MAE) defined as:

$$Test\ MAE = \frac{1}{n}\sum |y_0 - \hat{f}(x_0)|, \tag{6.2}$$

where the y_0 refers to the actual observed next year's accruals of the test sample and x_0 to the predicted accrual values of next year using training sample data. The main difference between the $\sqrt{MSE}$ and the MAE is that large deviations have a larger effect on the $\sqrt{MSE}$ than on the MAE.[3]

Following Theil (1966, p. 29), the test MSE can be further decomposed into the elements of bias, variance, and correlation of the actual and predicted coefficients, which allows a more accurate assessment of the source of error. Similar to Wallmeier (1997), we use the labels K1, K2, and K3 for the components of the test MSE. We express these components directly as a percentage of the MSE so that the sum of the components equals one.

K1 represents the bias, i.e., the difference between the mean fitted values of the train sample and the mean fitted values of the test sample. A low K1 value indicates that the fitted values of the train sample are similar to the fitted values of the test sample:

$$K1 = \frac{(\frac{1}{n}\sum realized - \frac{1}{n}\sum forecasted)^2}{MSE}. \tag{6.3}$$

[3]Armstrong and Collopy (1992) compare various error measures in economic time series forecasting in terms of reliability, validity, sensitivity, outlier protection, and relationship to decisions. They observe that the $\sqrt{MSE}$ is widely used by practitioners and in academia. However, a measure such as the median absolute error would outperform the $\sqrt{MSE}$ in terms of reliability, construct validity, and outlier protection but not in terms of sensitivity and relationship to decisions. They advise against using the $\sqrt{MSE}$ to evaluate the level of accuracy across time series. In our study, we present both, the $\sqrt{MSE}$ because of its frequent use and the MAE because of its superiority in accuracy assessment. Note that the study by Armstrong and Collopy (1992) is on time series models, while our study design uses a pooled sample.

K2 accounts for the squared differences in standard deviations.[4] A small (large) K2 value indicates that the standard deviations of the train and test samples are similar (different):

$$K2 = \frac{(AF * \sigma_{realized} - AF * \sigma_{forecasted})^2}{MSE}. \tag{6.5}$$

K3 considers the correlation between realized and forecasted values. A correlation of 1 would mean that the model fully accounts for the correlation between realized and predicted values:

$$K3 = \frac{2 * (1 - \rho(realized, forecasted)) * AF * \sigma_{realized} * AF * \sigma_{forecasted}}{MSE}. \tag{6.6}$$

6.3.2 LASSO

The following multivariate equation relates to the standard linear model:

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + \epsilon, \tag{6.7}$$

where Y is the dependent variable and $X_1, ..., X_p$ are the independent variables. If more variables are available, it might be tempting to include them in the model. However, it may not always be beneficial to include them. Some explanatory variables might be highly correlated with other variables, and the model might tend to overfit the in-sample data. The LASSO prevents overfitting by shrinking the weights or completely excluding the least important variables.[5]

Mathematically, the LASSO consists of two parts. The first part is similar to the residual sum of squares (RSS) of the standard OLS regressions, which aims to minimize

[4] Our sample, as part of the total population, requires an adjustment factor (AF) for the standard deviations and variances to account for the potential underestimation of the true population variance and to obtain an unbiased estimate:

$$Adjustment\ factor = \sqrt{\frac{n-1}{n}}. \tag{6.4}$$

[5] A similar but less frequently used method is the ridge regression which also shrinks the variables but does not completely exclude them.

the error for the in-sample data at hand. The second term, beginning in Equation (6.8) with λ, is a zero or positive tuning parameter that shrinks relatively unimportant variables to zero. Thus, a lambda of zero would result in the coefficients of the OLS model. A larger lambda shrinks variables with low predictive power to zero and excludes them from the model. The LASSO aims to minimize the following equation:

$$minimize\ \beta\ \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}, \tag{6.8}$$

where $\lambda \geq 0$ represents an optimal value for the tuning parameter. The model calculates the cross-validation error for different λ values and selects the λ with the lowest cross-validation error.

Figure 6.2 illustrates the in-sample LASSO cross-validation. Cross-validation divides the training data into a predefined number of different folds of training and validation data. All but one fold is used to train a model, which is then evaluated directly against the validation fold. In the example of a 4-fold cross-validation, the sample is randomly divided into four parts. Each fold is used once as the validation sample, while the remaining three folds are used for training. This results in four different coefficients for the explanatory variables. Finally, the coefficients are averaged to obtain the prediction coefficients for the test sample.

The LASSO model has been shown to be an improvement over OLS regression but does not necessarily provide the highest predictive accuracy. Recent research compares the LASSO model with forward stepwise selection[6] and the ridge regression[7]. The results show that a model with stepwise variable selection can outperform the LASSO model (Hastie et al., 2017; Meinshausen, 2007). Iterative adaptive ridge[8] regressions, especially for highly collinear variables, can also produce superior results to the standard LASSO

[6]In a forward stepwise selection, the most significant variable is added first, and other variables are added step by step until a predefined stopping rule is reached.

[7]Ridge regression shrinks variables toward zero but does not eliminate them completely.

[8]An iterative adaptive ridge groups collinear variables and computes a mean effect for the group. In contrast, LASSO randomly selects one of the collinear variables and omits the others, resulting in a slightly worse prediction (Dai et al., 2018).

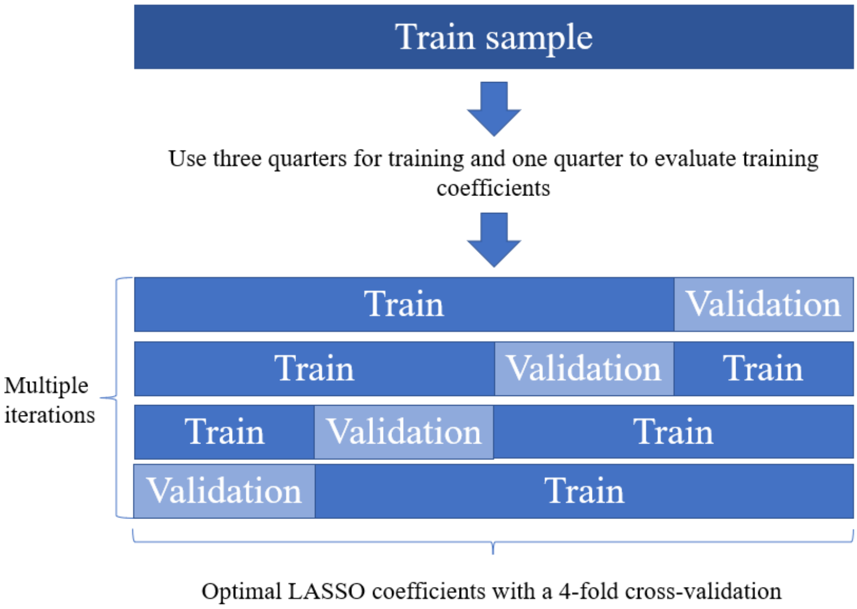

Figure 6.2: LASSO Cross-Validation. The figure illustrates the in-sample LASSO validation, where the training sample is split into multiple folds. Each training sample is optimized for the validation sample. Finally, the average training coefficients are used to predict the out-of-sample test data, as measured by the MSE.

(Z. Liu et al., 2017). The most sophisticated models do not always have to result in the highest prediction accuracy. James et al. (2021, p. 433) achieved the best predictive performance and thus the lowest MSE in the "Hitters" dataset by selecting only four variables with the LASSO, which are then simply used in an OLS model.

To compute the LASSO in R, we use the R package "glmnet" with the function *cv.glmnet*. In the first step, we set the parameters of the function to the Gaussian distribution (family = "gaussian"), choose the LASSO model (alpha = 1), and use ten cross-validations (n-folds = 10), which returns the minimum λ. In the next step, we use the function *glmnet* with the minimum λ as the input value (lambda = "lambda.min"). The other parameters remain unchanged, except that we need to remove the "nfolds" parameter. The model computes the coefficients of the explanatory variables under the optimal λ value for the training sample.

6.3.3 Tree-Based Methods

In short, tree-based methods divide the explanatory variables into two subgroups, also called regions, to obtain two sets of explanatory variables whose subgroups are more homogeneous. Then, for each subgroup, a value is predicted. After multiple splits, the tree reaches a predefined stopping rule and predicts the explanatory variables for each subgroup, forming the tree's final component. Formally, the goal of the model is to find splits, also called leaves, that minimize the RSS as follows:

$$\sum_{j=1}^{J}\sum_{i\in R_j}(y_i - \hat{y}_{R_j})^2, \tag{6.9}$$

where $R_1, ..., R_J$ represent the regions, and $\hat{y}_{Rj}$ represents the mean of the training observations of the j region.

A complete minimization of the RSS would imply that the model looks ahead to all variables. However, accounting for all possible combinations would carry a high computational cost. Therefore, tree-based models use a top-down "greedy" approach, also known as recursive binary splitting. The models decide on the optimal split at the current step of the current branch by selecting the predictor X_j and splitting at point s to minimize the RSS into two regions at $X|X_j < s$ and $X|X_j \geq s$. It is called a top-down approach, as the splitting into branches starts with all observations and then successively splits into smaller branches. The ultimate goal is to choose j and s such that the RSS of the training data of the following equation is minimized:

$$\sum_{i:\ x_i\in R_1(j,s)}(y_i - \hat{y}_{R_1})^2 + \sum_{i:\ x_i\in R_2(j,s)}(y_i - \hat{y}_{R_2})^2, \tag{6.10}$$

where $R_1, ..., R_{j,s}$ represent the regions, $\hat{y}_{R_1}$ stands for the mean response of the training observations in $R_1(j,s)$, and $\hat{y}_{R_2}$ for the mean response of the training observations in $R_2(j,s)$.

Then, the above regions or branches are split again to further reduce the RSS. It is important to note that the further splitting is done only *within* one of the previously created regions, which will add a third region (or branch). This process is repeated until a predefined stopping rule is reached, such as the minimum number of observations in each branch. Finally, when no more branches are created, the model returns the mean estimate of the training observations for each region or leaf.

An illustrative and frequently used example is the prediction of the logarithmic salary of a baseball hitter, as shown in Figure 6.3. The most basic example uses the explanatory variables “years of experience” and “number of hits”. Since the years of experience is the variable that reduces the RSS the most, the first branch divides the data into two separate regions (R_1 and R_2) of hitters with less than 4.5 years of experience and equal to or more. For each of the two regions, the average salary is computed and predicted, referred to as the “leaf.”

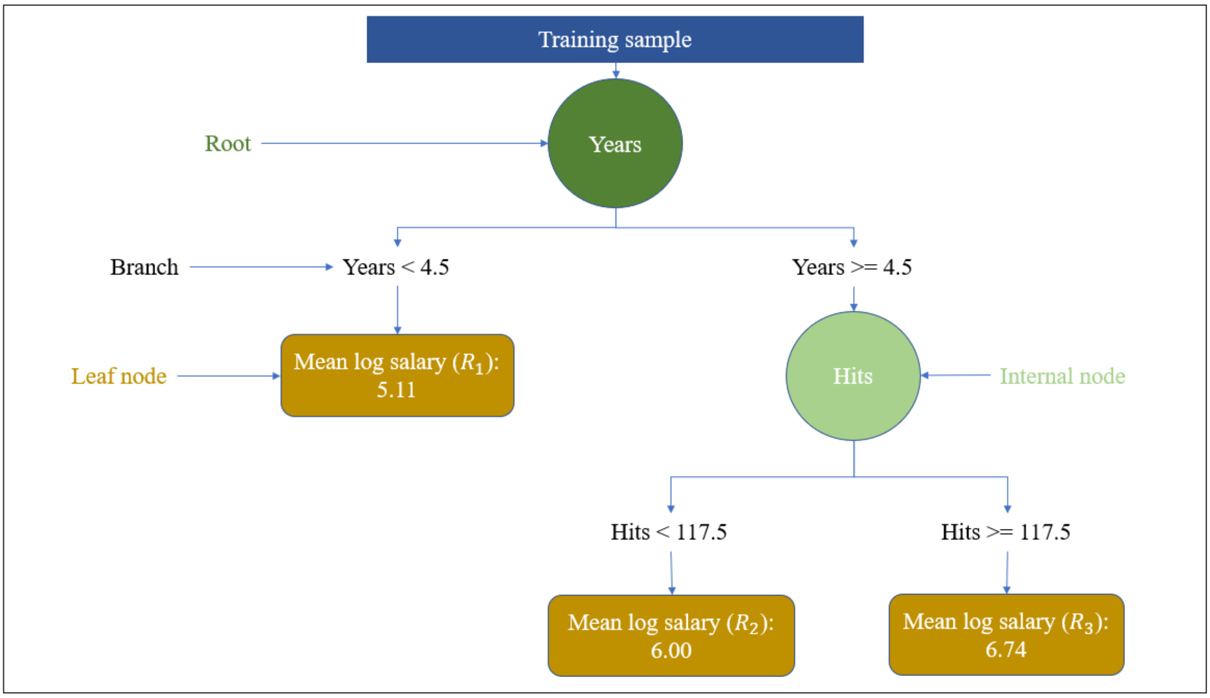

Figure 6.3: Regression Tree From the “Hitters” Dataset. The figure shows the concept of regression trees. A variable is split into two branches at the point that reduces the RSS the most. After reaching a specified stopping rule, the leaf nodes compute the training coefficients for the specific branch. Note: Adapted from *An Introduction to Statistical Learning* (2nd ed., p. 328), by G. James et al., 2021, Springer, New York, NY. Copyright 2021 by Springer Science+Business Media, LLC, part of Springer Nature 2021.

The number of hits can be used as a second variable to further increase the model's predictive accuracy. Since the RSS of the salary is higher for players with an experience equal to or more than 4.5 years of experience, the second branch divides the R_2 subset into two subsets, R_2 and R_3, according to the number of hits. Players with fewer than 117.5 hits are now in R_2, while players with equal or more hits are in R_3. Again, the average salary for each region is predicted and reported. This results in three different "leaves" that predict the log salary. Unsurprisingly, the salary is lowest for players with only a few years of experience and highest for players with more experience and more hits. The "Hitters" dataset contains other not included explanatory variables that could improve the accuracy of the prediction, such as the number of runs, walks, or putouts. These variables would be added as additional nodes and branches, with each additional node generating an additional prediction for the subset.

Any of the above splits will reduce the RSS in the training data. However, a tree grown too large might overfit the training data, resulting in an unnecessarily high MSE on the test data. The solution is to grow a large tree that may overfit the training data but then prune it to keep the MSE of the test data low. The "cost complexity pruning" solves this problem by adding a tuning parameter α, which is similar to the penalty value λ in the LASSO regression. The following equation minimizes the RSS on the test data:

$$\sum_{j=1}^{|T|} \sum_{i \in R_j} (y_i - \hat{y}_{Rj})^2 + \alpha |T|, \tag{6.11}$$

where $|T|$ represents the number of terminal leaves of the tree T, R_j represents the rectangle containing the j-th leaf observations, and $\hat{y}_{Rj}$ is the prediction based on the mean of the training observations in R_j. The optimal value of α is determined by cross-validation by computing the MSE of the test data for different values of α and selecting the α resulting in the lowest cross-validated MSE value. In other words, α ensures that the optimum between the bias-variance trade-off is found.

The properties of regression trees are quite different from linear regressions. Depending on the underlying data, one model may outperform another. If the data follow a linear relationship, it is likely to outperform tree-based methods.[9] However, tree-based methods can better cover complex and nonlinear relationships in the underlying data. They are intuitive to understand, even with little statistical knowledge. They can be represented graphically and are more closely related to human thinking.

Trees are generally not robust to changes in data, which can lead to large variations of the estimators. Already a slightly different selection of training and test samples might result in a different variable selection order. To reduce the unwanted variance in the predicted estimators, the "bagging" procedure is used, which is a combination of the two words "bootstrap aggregation". The bootstrap procedure randomly draws a subset of observations from the training sample. This process is repeated several times so that the training observations occur in some, but not all, bootstrapped training samples. Then, the estimated prediction coefficients of all bootstrapped training samples are averaged. By averaging the prediction coefficients of multiple trees, the variance of the individual trees is reduced, thus increasing the prediction accuracy. Unlike the trees presented above, bagging does not prune the trees.

Random forests are a further improvement over bagged trees. Again, the trees are created based on bootstrapped training samples. However, when splitting into new branches, the model can only choose from a *random* sample of variables instead of all. Often, the number of variables available in the random sample is the square root of all variables. This restriction may seem counterintuitive, however, it is a valuable method for further reducing the variance because it decorrelates the trees. This is because the variables that reduce the RSS the most are often the same for similar training data. If a variable has strong explanatory power, most bagged trees start with the same variable. Therefore, the trees are similar to each other and consequently highly correlated. However, averaging highly correlated trees does not reduce variance as much as averaging trees with low correlation. The restriction to a random selection of a set of variables means that

[9]Under the following conditions, the multiple linear regression is the best linear unbiased estimator (i.e., smallest variance): (1) linearity of parameters, (2) random sampling, (3) conditional mean independence, (4) no linear dependence of regressors and fewer regressors than observations, and (5) homoscedasticity (Bohrnstedt & Carter, 1971).

the variable with the highest explanatory power is often not available to the model. This causes the trees to look different, which can be referred to as *decorrelating the trees*. Random forests that average over low correlated trees tend to result in a lower variance and a higher predictive accuracy than bagged trees.

For the computational part of the study, we use the R library "randomForest" with the same named function. Due to the superiority of random forests over other tree-based methods, we only run computations using them. We set the function parameters (ntree = 200 and importance = TRUE) and run the model. We choose a high number of trees as this leads to more accurate predictions with a decreasing marginal gain in accuracy without overfitting. James et al. (2021, p. 341) consider a tree selection of 100 as sufficient since additional trees would not substantially improve the results.

6.3.4 Support Vector Machines

SVMs follow the concept of maximal margin classification, which consists of hyperplanes that separate binary or multiple classes of observations into two groups. A hyperplane over two dimensions and two classes would simply be a one-dimensional line separating the two discriminant classes. A two-dimensional hyperplane follows the equation:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0, \tag{6.12}$$

where β_0 is a constant and β_1 and β_2 are the parameters representing the properties of a line separating a set of observations.[10] Usually, the separation rule for hyperplanes is to use a *maximal margin hyperplane*. Its goal is to set the line so that a maximum and equal distance is found between the closest observations of each group and the line of the hyperplane expressed by the margin parameter M.

Figure 6.4 presents the concept of SVM. The hyperplane line in black aims to separate the blue and green observations into two groups in such a way that the distance of the

[10] In a multivariate setting, the equation can be extended simply by adding the other variables.

hyperplane to the distance between the two groups of observations is maximized. The orange dashed lines are bounding hyperplanes and run parallel to the hyperplane at the distance of the maximum margin.

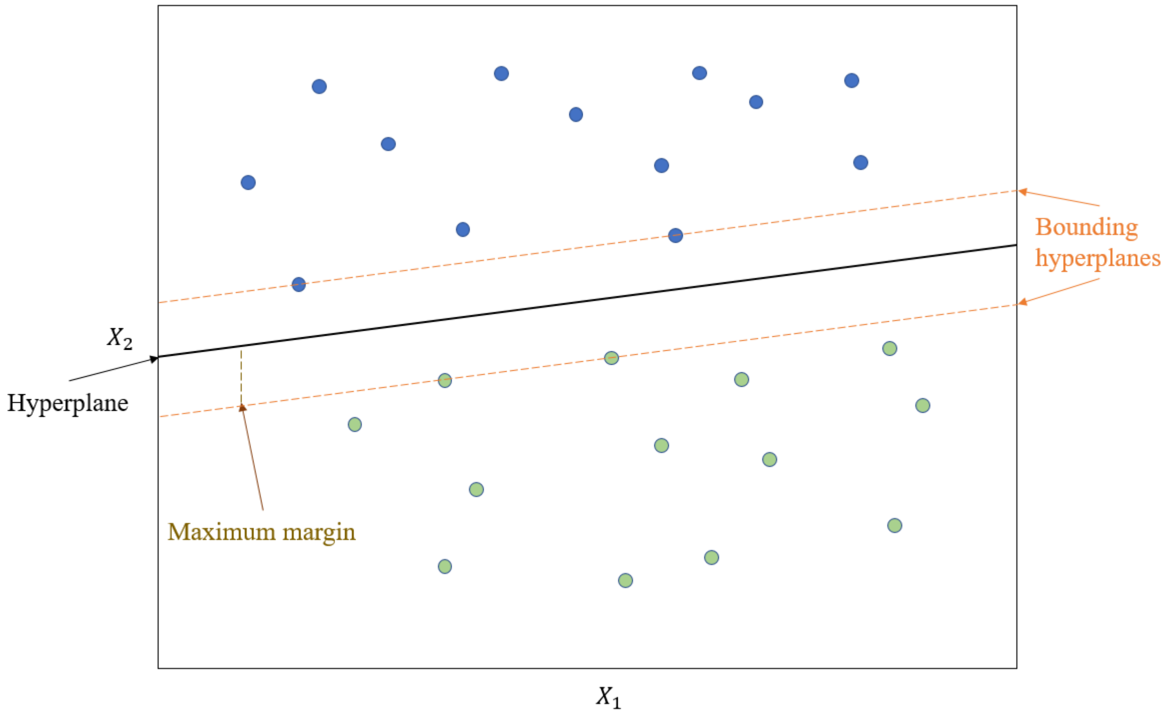

Figure 6.4: Hyperplanes in SVM. The figure shows the concept of SVM. In this example, a black one-dimensional hyperplane line separates the blue and green observations into two groups while maintaining the maximum distance from the bounding hyperplanes.

In cases where the two groups cannot be optimally separated, nonlinear decision boundaries increase the accuracy. Kernels[11] allow greater flexibility to the hyperplane to maximize the margin. Examples of SVMs are often on binary classifications, but regressions on continuous data are also commonly used and have almost identical properties.

To implement this in R, we use the "e1071" library and the *svm* function with the function's default parameters, which include an internal standardization scaling with mean zero and a radial kernel.

[11] Commonly used kernels are polynomial or radial kernels.

6.4 Data and Descriptive Statistics

6.4.1 Sample Construction

One of the peculiarities of this paper is that we use an extensive number of variables. In this section, we describe the data and firm selection process, which forms the basis for data analysis.

Variable Selection

Our data source is the Refinitiv Worldscope universe, which includes over 80,000 dead and live listed firms worldwide and is accessible via the Refinitiv Datastream Excel add-in. Thomson Reuters (the former provider of Datastream and Worldscope) has published a Data Definitions Guide describing the structure of the Worldscope data universe.[12] The guide is divided into different templates: banks, industrial, insurance, and other financial companies, as firms from different sectors have different data items and performance metrics. In our study, we rely on the industrial template, which contains a total of 1,257 individual variables.[13] The template is grouped into several categories, including general information about the firm, financial statement figures, and metrics computed by Worldscope.

Not all variables may be relevant for predicting total accruals. For this reason, we describe our variable selection process in more detail so that others can reproduce it. The general information section contains information about the firm, such as the company name, nation, address, and other identifiers. Since the identifier variables are not useful for predicting accruals, we use only the number of employees and a variable indicating

[12]The most recent version we found was published by Thomson-Reuters (2013).

[13]The Compustat dataset is most commonly used for accounting research on U.S. equities. Researchers can download this pre-built dataset without any modification. In contrast, Refinitiv Eikon Worldscope does not provide such a pre-built dataset but offers more flexibility in constructing a dataset from the 1,257 financial statement variables.

whether restated data are available. From the industry section, we use the static industry identifiers General Industry Classification (WC #06010), Industry Group (WC #06011), and SIC Code 1 through 3 (WC #07021, WC #07022, and WC #07023). From the auditor section, we use the static variable Parent Auditor 1 (WC #07800) and the time series variable Auditor Fees (WC #01801).

We omit the "Key Items in U.S. Dollars" section because the U.S. firms in our sample all report in U.S. dollars, and the key items are duplicates of variables we include in the following sections. From the "Stock Data" annual statistics, we extract the Market Capitalization (WC #08001), Trading Volume ($ Amount) (WC #08006), and Closely Held Shares in % (WC #08021). From the "Company Specific Accounting Practices" section, we use the character variable Accounting Standards Followed (WC #07536).

The next category consists of financial statement items. We use all items of the balance sheet and income statement in full, including all supplementary items in these sections. Worldscope provides different structures and layouts for the cash flow variables. We use only the first cash flow statement structure, "Funds Flow Statement - per FASB 95" (Thomson-Reuters, 2013, pp. 98–99), and disregard the slightly different structures of "Cash" and "Total Sources/Total Uses" (Thomson-Reuters, 2013, pp. 100–102). Only the FASB 95 layout structure includes the subtotal of the net cash flow of operating, investing, and financing activities. Consistent with the selection patterns above, we include all "Supplementary Cash Flow Fields". The third category of variables consists of Worldscope's key ratios, which we also include.

The remaining items are grouped in the categories Pension and Postretirement, Product Segment Data, Geographic Segment Data, Foreign Business, Industry Metrics, Other Fields, Annual Series Interim Data, Monthly Price Information, Multiple Share Data, and Monthly Foreign Exchange Rates. Since this information is firm-specific and most of the data is missing for most firms, we do not include any of these items. Finally, we result with 364 selected variables.

Firm Selection

Not all of the more than 80,000 dead and live firms in the Worldscope universe are fully covered in Worldscope. In fact, for most firms, only the most essential items are available in the database. For large U.S. firms, most Worldscope items are covered, while only a few data items are available for most smaller firms. Therefore, we focus only on large U.S. firms.

First, we select all 181,823 U.S. firm-years from 1986 to 2020 with available Market Capitalization (WC #08001) reported in U.S. dollars, Total Assets (WC #02999), and SIC Code 1 (WC #07021).[14] We then select the largest 20% of firm-years by market capitalization, which shrinks our sample to 36,351 firm-years and 4,205 unique firms. Finally, we remove financial firms with a SIC Code 1 (WC #07021) between 6000 and 6999, consistent with Larson et al. (2018).

After this step, our uncleaned dataset consists of 26,777 firm-years and 364 variables. However, some variables contain NA or zero values that cannot be fed into a regression, leaving us with two options. Either we remove the entire firm-year or the entire variable. Because of the different characteristics of the variables, we treat financial statement, static, and ratio variables differently.

Financial Statement Variables

We begin our data cleaning with several manipulations of the financial statement variables. Some variables contain extremely few observations. For example, the supplementary income statement variable Ordinary Profit Japan (WC #18175) contains only one entry for "DEERE & COMPANY" for the year 1999. If this variable were included in the regression, the regression coefficient would simply be adjusted for this specific firm-year. This might increase the in-sample explanatory power but not the out-of-sample predictive power of total accruals. To ensure that the variables add useful information, we require

[14] See GitHub for a detailed description of how the base sample was downloaded and constructed: https://github.com/chardonnensp/Datastream-Worldscope-fundamental-dataset

that at least 5% of the values be nonzero and not NA's.[15] This results in 74 of the initial 272 financial statement variables being dropped due to this selection criterion, leaving us with 198 financial statement variables.

Of the remaining variables, some still contain a considerable number of zeros and NA's. It is important to note that not all firms in our sample use the same accounting items. For example, Amazon's 2020 10-k report (Amazon.com, 2020) does not include any position for the Worldscope items Raw Materials (WC #02097) and Work in Progress (WC #02098). For 2012 through 2020, the Worldscope database reports these items as NA. For years prior to 2012, the database returns a zero. A missing value reported as NA would lead to exclusion from Amazon, since regressions cannot handle missing values, but a zero does not lead to exclusion. We expect the non-zero financial statement items for the largest 20% of firm-years were fully reported. Therefore, we recode missing values in the balance sheet, income statement, and cash flow statement as zeros.

Next, we scale all 198 financial statement variables by the same year's Total Assets (WC #02999).[16] Given the large number of variables, standardizing the values makes them more comparable across different firms.[17] The computation of accruals requires a one-year lag of some financial statement variables. Consequently, we lag and include all financial statement variables, which results in a doubling of the financial variables.[18]

Computation of Accruals

The available financial statement variables, including the one-year lag, now allow us to compute accruals. We run our computations using two different definitions of accruals.

[15] In the robustness test Section 6.5.4, we increase this requirement to 95% for financial statement variables to be included.

[16] Most earnings management models scale their variables by lagged total assets. We believe that in this research setting, the standardization with the same year's total assets is more appropriate.

[17] Due to the scaling, the total assets for each observation of a firm-year are equal to one. Therefore, we need to remove total assets from the sample to avoid multicollinearity. We also remove the variable Total Liabilities & Shareholders' Equity (WC #03999) since it represents the sum of liabilities and common equity and is equal to the sum of total assets.

[18] In the robustness test Section 6.5.4, we rerun the computations with three lags instead of one.

The first consists of WCA following the traditional literature (e.g., Dechow et al., 1995; Jones, 1991), while the second follows a more comprehensive approach (Larson et al., 2018; Richardson & Sloan, 2005). Both accrual definitions have in common that they are based on broadly available accounting items.

WCA consist of the annual change of short-term accounting items, excluding the change in cash and short-term debt. The following equation presents the calculation of WCA, with all values scaled by Total Assets (WC #02999):

$$WCA_{i,t} = \Delta CA_{i,t} - \Delta Cash_{i,t} - (\Delta CL_{i,t} - \Delta STD_{i,t}), \tag{6.13}$$

where $WCA_{i,t}$ are the annual changes in WCA for the individual firm i in year t. $CA_{i,t}$ represents the Current Assets (WC #02201), $CL_{i,t}$ represents the Current Liabilities (WC #03101), $Cash_{i,t}$ represents Cash & Short Term Investments (WC #02001), and $STD_{i,t}$ represents Short-Term Debt (WC #03051).

Larson et al. (2018) and Richardson and Sloan (2005) define accruals more comprehensively. Following their idea, all non-cash assets and liabilities accounting items could contain some degree of earnings management. In other words, only the annual change in cash and common equity is expected to be free of earnings management. The following equation presents the calculation of CompAcc, with all values scaled by Total Assets (WC #02999):

$$CompAcc_{i,t} = \Delta CE_{i,t} - \Delta Cash_{i,t}, \tag{6.14}$$

where $CompAcc_{i,t}$ is the annual change in CompAcc for the individual firm i in year t. $CE_{i,t}$ stands for Common Equity (WC #03501), and $Cash_{i,t}$ for Cash & Short Term Investments (WC #02001).

Since we include all available financial variables, we cannot use accruals of the same year. This would lead to multicollinearity, as the model would find the accrual components. Simply excluding the accrual components does not solve the problem, as other accounting items may be associated with the accrual components, which would still result in a very high, but unjustified, explanatory power. To solve this problem, we shift our estimated and predicted accruals forward by one year. For example, the accrual components of 2019 included in the explanatory variables must explain (for in-sample data) or predict (for out-of-sample data) the accruals of 2020. The additionally required lead reduces our dataset to 22,449 firm-years.

Ratio Variables

Our dataset contains 82 financial ratios of interest constructed and provided by Worldscope. Again, as with the financial statement variables, some ratios have frequent $NA's$. Unlike the financial statement items, we cannot assume that a missing value equals zero. Therefore, we must decide whether to remove the firm-year observation or the variable with missing values. We require that the ratios have at least 95% non-NA and non-zero values. If fewer data are available, these variables are removed entirely. Due to the comparatively high proportion of NA values, 38 of 83 variables are removed, leaving 45 financial ratios.

Of the ratios with less than 5% missing values, we must delete the firm-years that contain missing values, which reduces our sample to 18,571 firm-years. We then merge the financial statements and the ratios into a dataset containing all numeric items. Finally, all numerical values are winsorized at the 1% level.

Static Variables

The static variables consist of the year, industry group, SIC codes (SIC1, SIC2, SIC3), auditor, and accounting standards followed. The static variables are factorized, meaning that in the regression models each of these factor levels represents a separate variable with a value of 1 if the factor level matches and 0 otherwise.

We exclude variables with more than 5% of missing values for each variable, resulting in the exclusion of SIC2 and SIC3, but not SIC1. We also exclude the accounting standards followed, as only 15 firms have a value other than "US standards (GAAP)".

Some factors occur infrequently. SIC1 and the Industry Group have many different factor levels, but partially only a few observations per factor level. Therefore, we group them by using only the first digit, which is the broad industry classification. Typical approaches to controlling for industry differences often include the first two digits of the SIC code (Kothari et al., 2005). However, our sample of 4,205 individual firms does not allow for a more specific industry comparison. For the other variables, we set a minimum requirement of 50 firm-years per factor level. One hundred fifty-nine firm-years that are not in one of these selected factor levels are dropped. Table 6.1 shows the firm-year selection process.

Firm-years selection	**Dropped FY**	**Remaining FY**
All public U.S. firm-years from 1986 to 2020		181,755
Firm-years in the top 20% by market cap	(145,404)	36,351
SIC1 codes below 6000 and above 6999	(9,574)	26,777
Firm-years with missing leads and lags	(4,328)	22,449
Firm-years with missing ratio values	(3,878)	18,571
Firm-years with missing static values	(159)	18,412

Table 6.1: Selection Process for Firm-Years. The table shows the firm-year selection process of the Refinitiv Worldscope database. Our final dataset consists of 18,412 firm-year observations.

At the end of our variable selection process, we remove the firm-specific identifiers such as the name and the stock quote symbol. Our final dataset contains 447 variables. If we count the different factor levels as separate variables, the models use 503 explanatory variables. Table 6.2 shows the breakdown of variables into financial statement, ratio, and factor variables. Appendix D lists all Refinitiv Worldscope variables used.

Variable type	**Number of variables**
Financial statement variables	397
Ratio variables	45
Factor variables	5
Total variables	**447**

Table 6.2: Type of Selected Variables. The table shows the number of variables selected for each of the categories financial statement, ratio, and factor variables. Our final dataset consists of 447 distinct variables.

We split our final dataset into a training dataset and a test dataset. We assign half of the data to the training data and half to the test data. The R function *sample* draws the observations randomly. To ensure reproducibility, we manually set the seed to “12345”.[19]

[19] We ran our results for different seeds and got nearly identical results.

We use the R function *predict* in combination with the “newdata” option (or “newx” if the required input must be a matrix). The function can be used for both in-sample and out-of-sample tests. It is important to note that for in-sample MSE, the “newdata” option must be applied to the training dataset, which is identical to selecting the fitted values of the training data. Accordingly, for out-of-sample predictions, the “newdata” option must be applied to the still unused test data, so that the calculated regression parameters of the training sample are predictions on the test data.

6.4.2 Descriptive Statistics

This section presents the descriptive statistics for our variables. Table 6.3 contains the variables required to compute the accruals. Figure 6.5 shows the histograms of the distributions of accruals, their components, and the variables of the modified Jones model. Figure 6.6 shows the annual distribution of WCA and CompAcc. The distribution of WCA is narrower compared to the distribution of CompAcc.

Figure 6.7 shows the correlation matrices for the selected variables in Table 6.3 from 1989 to 2004 in Panel (a) and 2005 to 2019 in Panel (b). Pearson correlations are in the upper right part, and Spearman correlations are in the lower left part. Generally, the correlations of our accrual definitions with their components are low.

(a) WCA Components

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
WCA	18,412	-0.003	0.040	-0.150	-0.019	0.015	0.122
Net Working Capital	18,412	0.073	0.132	-0.502	-0.007	0.142	0.780
Current Assets	18,412	0.394	0.221	0.045	0.207	0.555	0.902
Cash	18,412	0.134	0.159	0.001	0.020	0.188	0.701
Current Liabilities	18,412	0.219	0.117	0.042	0.130	0.285	0.590
Short Term Debt	18,412	0.032	0.046	0.000	0.001	0.045	0.249

(b) CompAcc Components

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
CompAcc	18,412	-0.001	0.091	-0.308	-0.039	0.038	0.314
Common Equity	18,412	0.456	0.197	0.058	0.312	0.586	0.918
Cash	18,412	0.134	0.159	0.001	0.020	0.188	0.701

(c) Modified Jones Model Components

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Sales	18,412	0.992	0.699	0.147	0.498	1.244	3.881
Receivables	18,412	0.126	0.093	0.007	0.056	0.172	0.482
PPE	18,412	0.592	0.390	0.035	0.261	0.887	1.618

Table 6.3: Summary Statistics for Accruals and Their Components. The table shows the summary statistics for the variables required to compute accruals.

Panel (a) shows the summary statistics of the change in annual WCA:
$WCA_{i,t} = \Delta\ Current\ Assets\ (WC\ \#02201)_{i,t} - \Delta\ Cash\ \&\ Short\ Term\ Investments\ (WC\ \#02001)_{i,t} - (\Delta\ Current\ Liabilities\ (WC\ \#03101)_{i,t} - \Delta\ Short\text{-}Term\ Debt\ (WC\ \#03051)_{i,t})$.

Net working capital represents the total share of total assets computed as follows:
$Net\ Working\ Capital_{i,t} = Current\ Assets\ (WC\ \#02201)_{i,t} - Cash\ \&\ Short\ Term\ Investments\ (WC\ \#02001)_{i,t} - (Current\ Liabilities\ (WC\ \#03101)_{i,t} - Short\text{-}Term\ Debt\ (WC\ \#03051)_{i,t})$.

Panel (b) shows the CompAcc, which are calculated as annual changes:
$CompAcc_{i,t} = \Delta\ Common\ Equity\ (WC\ \#03501)_{i,t} - \Delta\ Cash\ \&\ Short\ Term\ Investments\ (WC\ \#02001)_{i,t}$.

Panel (c) shows the summary statistics for the explanatory variables of the modified Jones model, consisting of $Sales\ (WC\ \#01001)_{i,t}$, $Receivables\ (WC\ \#02051)_{i,t}$, and $PPE\ (WC\ \#02301)_{i,t}$.

All variables are scaled by $Total\ Assets\ (WC\ \#02999)$.

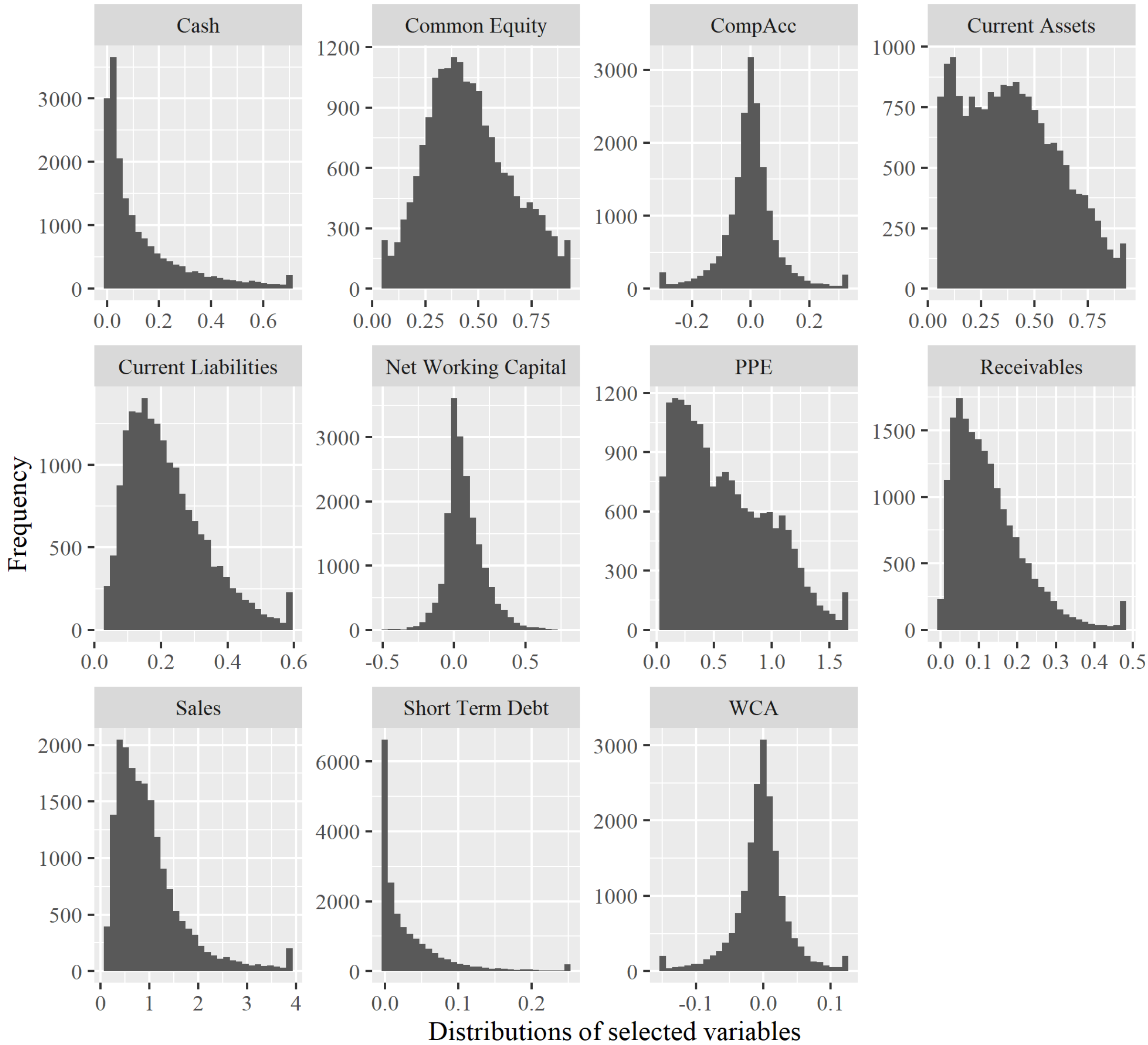

Figure 6.5: Histograms for Accruals and Their Components. The figure shows the histograms of accruals, their components, and the variables of the modified Jones model. The variable definitions correspond to those in Table 6.3.

(a) WCA Distribution by Year **(b) CompAcc Distribution by Year**

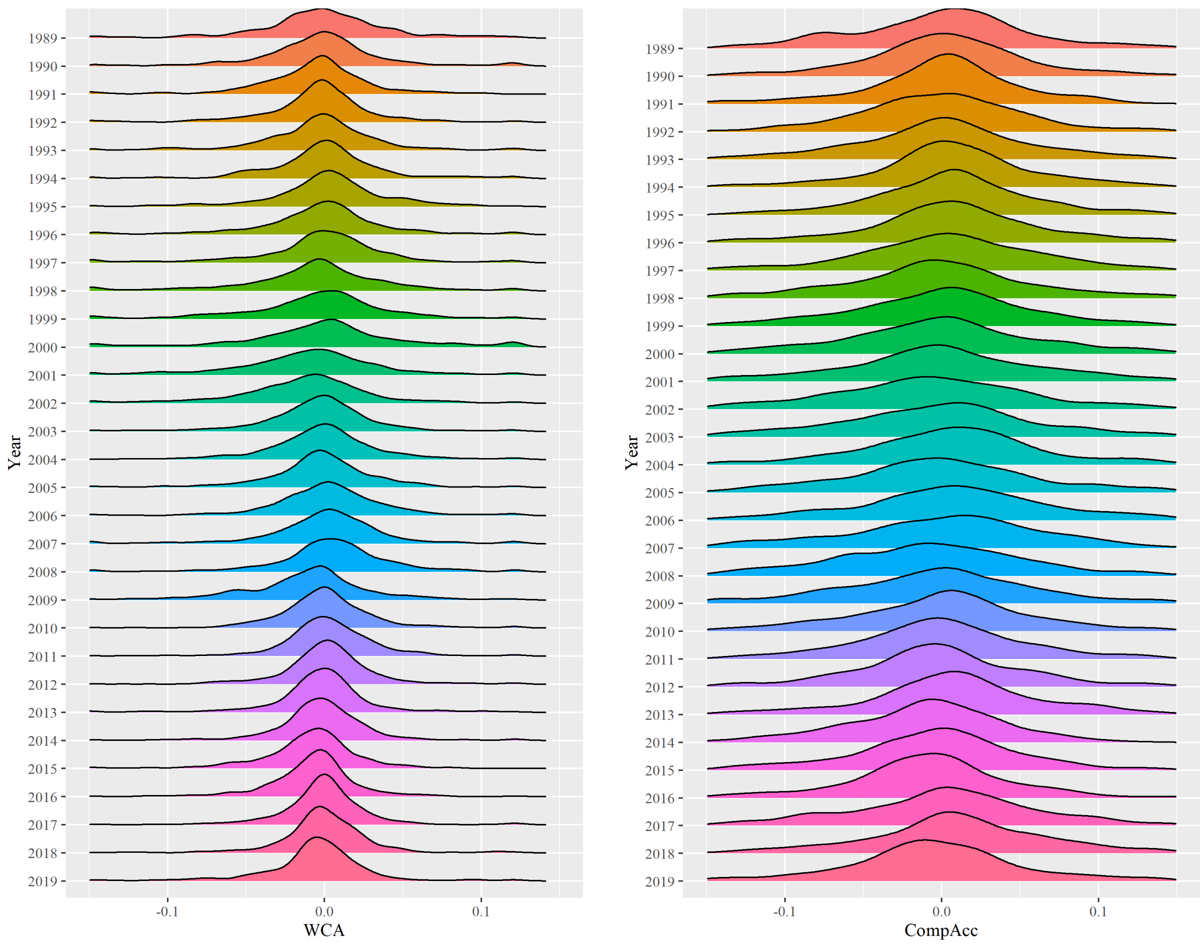

Figure 6.6: Yearly Accruals Distributions. The figure shows the annual distribution of WCA and CompAcc from 1989 to 2019. The distribution of WCA is narrower compared to the distribution of CompAcc. The variable definitions correspond to those in Table 6.3.

(a) **Correlations From 1989 to 2004** (b) **Correlations From 2005 to 2019**

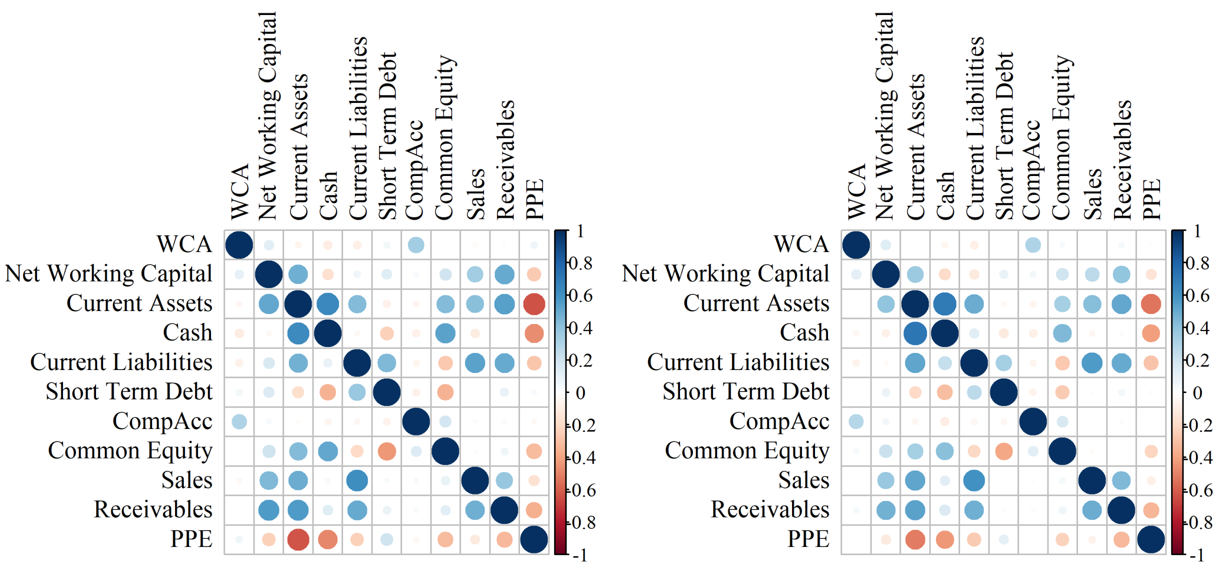

Figure 6.7: Correlation Matrices for Accruals and Their Components. The figure shows the correlation matrices for the selected variables in Table 6.3 from 1989 to 2004 in Panel (a) and 2005 to 2019 in Panel(b). Pearson correlations are shown in the upper right part, and Spearman correlations in the lower left part. The colors and the size of the circles depend on the strength of the correlation. The colors and the size of the circles depend on the strength of the correlation. The variable definitions correspond to those in Table 6.3.

6.5 Results

The results section proceeds as follows. First, we use the explanatory variables of the modified Jones model to explain next year's WCA and measure the in-sample and out-of-sample MSE. Second, we use all our explanatory variables on next year's WCA and evaluate the results. Thereafter, we use machine learning models and test whether this leads to more precise predictions for next year's accruals. Finally, we use the five most important variables selected by the LASSO model and evaluate if they have predictive power on next year's accruals.

6.5.1 Jones-Type Explanatory Variables

Regression Results

In this subsection, we present the regression results of the explanatory variables of the modified Jones model on next year's WCA and compare the in-sample and out-of-sample MSE. The results of the in-sample regression are presented in Regression (1) of Table 6.4. Although the coefficients of the modified Jones model are significant, the adjusted R^2 of our base modified Jones model is close to zero (0.2%). Typical adjusted R^2 values for modified Jones models are in the higher single digits. For example, in the study by Ball and Shivakumar (2006, p. 217), the modified Jones model yields an explanatory power of 8.8%. There are several differences in research design, but the main difference is that we use the one-year ahead WCA instead of the same-year accruals.[20]

The explanatory power of typical earnings models increases with added cash flows. Accordingly, Regression (2) adds cash flow from operations and its lag. In our model, however, the adjusted R^2 increases only marginally to 0.3%. Cash flow from operations of

[20]We found further differences from the study by Ball and Shivakumar (2006). First, they compute accruals as earnings minus cash flow from operations. Second, they do not select the largest 20% of firm-years in terms of market capitalization. Third, they use the Compustat database instead of Worldscope. Fourth, they scale the variables by average total assets, whereas we use total assets of the same year, and finally, they have a different sample period.

	Dependent variable:	
	Lead WCA	
	(1)	(2)
Intercept	-0.006^{***}	-0.008^{***}
	(0.001)	(0.001)
ΔRev - ΔRec	-0.006^{**}	-0.007^{***}
	(0.003)	(0.003)
PPE	0.004^{***}	0.004^{***}
	(0.001)	(0.001)
CFO		0.030^{***}
		(0.008)
CFO lag		-0.012
		(0.008)
Observations	9,206	9,206
R^2	0.002	0.004
Adjusted R^2	0.002	0.003
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 6.4: Regression Results of Modified Jones Model Variables on Lead WCA. The table shows the regression that explains lead WCA. Regression (1) uses the components of the modified Jones model as explanatory variables. Regression (2) additionally includes cash flow from operations and lagged cash flow from operations. While most of the variables are significant, the explanatory power of the adjusted R^2 remains close to zero, indicating that the explanatory variables cannot explain next year's WCA. The variable definitions correspond to those in Table 6.3.

the actual year shows a significant explanatory power for next year's WCA. The one-year lag in cash flows has no significance for the one-year lead of WCA. We do not replicate the model of Larson et al. (2018) on CompAcc because their model includes multiple leads and lags of cash flows, which would drastically reduce our sample size.

In general, the frequently used variables in Jones-type models have little explanatory power to explain next year's WCA. Consequently, it appears tempting to include additional variables that might increase the explanatory power.

Error Measures

Table 6.5 reports the MAE, the $\sqrt{MSE}$, the MSE, and the MSE components K1, K2, and K3 of the in-sample and out-of-sample results with the explanatory variables of the modified Jones model for lead WCA as in Regression (1) of Table 6.4. The in-sample MSE (0.001503) is similar to the out-of-sample MSE (0.001548), which is likely because the model has only two explanatory variables, the change in revenue minus the change in accounts receivable and the gross PPE.[21] The $\sqrt{MSE}$ for the in-sample value (0.038764) is only slightly lower than the out-of-sample value (0.039349). The MAE directly indicates the measurement error with respect to a firm's total assets. In other words, the average error of the fitted in-sample values for next year's WCA is 2.7% of total assets.

The prediction error appears relatively large, considering that the average net working capital is 7.3% of the balance sheet, as shown in Table 6.3. This indicates that the average error of the fitted values is more than one-third of WCA.

Model	**MAE**	$\sqrt{MSE}$	**MSE**	**K1**	**K2**	**K3**
In-sample WCA	0.026551	0.038764	0.001503	0.00	91.20	8.80
Out-of-sample WCA	0.026896	0.039349	0.001548	0.02	90.90	9.07

Table 6.5: OLS Error Measures of Modified Jones Model Variables on Lead WCA. The table presents the MAE, $\sqrt{MSE}$, and the MSE for in- and out-of-sample lead WCA resulting from Regression (1) of Table 6.4. The MSE and MAE computation is presented in Section 6.3.1. The in-sample MSE consists of the mean squared difference of the dependent variable, the lead WCA, to the fitted values of the in-sample data. The out-of-sample MSE uses data from a training sample to predict the values of the dependent variable for a set of test data not yet seen.
The right side of the table shows the MSE decomposition into K1, K2, and K3. The largest source of error is K2.

We measure whether the MSE of the in-sample and the out-of-sample show significant differences. For medium and large samples, the differences in MSE can be calculated conservatively using a large sample test (Holst & Thyregod, 1999). Following Devore and Berk (2012, pp. 490–491), we compute the z-score:

[21] We have run the results for different random selections of the train and test data. In almost all cases, the in-sample MSE is lower than the out-of-sample MSE because OLS minimizes the squared errors of the underlying training data. In rare cases where the training sample data is more dispersed than the test data, the in-sample MSE may be higher. Machine learning models inherently incorporate sampling techniques such as n-fold validation or bootstrapping to marginalize the impact of individual outliers in the training data.

$$z = \frac{MSE_{is} - MSE_{oos}}{\sqrt{\frac{s_{is}^2}{N_{is}} + \frac{s_{oos}^2}{N_{oos}}}}, \tag{6.15}$$

where MSE_{is} (MSE_{oos}) is the in-sample (out-of-sample) MSE, s_{is}^2 (s_{oos}^2) the variance, and N_{is} (N_{oos}) the number of observations. From the above formula, the computed z-score of -0.079 is insignificant, which indicates that in- and out-of-sample MSE cannot be interpreted as different.

The MSE can be further decomposed into its components K1, K2, and K3, as described in Section 6.3.1. The right side of Table 6.5 reports the percentages of each component, which means that the sum of K1, K2, and K3 equals 100%.

The zero bias of K1 for the in-sample is by definition because the mean of lead WCA corresponds to the mean of the fitted values of lead WCA. By far, the largest source of error arises from K2 for both, in-sample and out-of-sample approaches, as the standard deviation of the training is much higher than the fitted (predicted) values. The standard deviation for the in-sample lead WCA is 0.0388, while it is only 0.0013 for the fitted values. This low standard deviation of the fitted values is because all fitted values are only within a marginal range of -0.009 and 0.003, most likely because the variables of the modified Jones model cannot account for the variation in next year's accruals, as shown in Panel (a) of Figure 6.8. In contrast, the actual lead WCA range from -0.150 to 0.122 of total assets.

K3 accounts for the correlation of actual and predicted values, which is near zero. The error resulting from K2 is larger than the error from K3. This is because the particularly small standard deviation in K3 is *multiplied* with the other elements of the term. In contrast to K2, the large $difference$ between the actual and the predicted standard deviation is squared.

Actual Values and Fitted Values

The actual and fitted values can be shown in a dot plot. In in-sample measures, the lead WCA values from the training sample are plotted on the x-axis, and the fitted lead

WCA values from the same training sample are plotted on the y-axis. For out-of-sample measures, the x-axis plots the lead WCA from the test sample, and the y-axis plots the predicted lead WCA for the test sample, computed based on training sample data.

Panel (a) of Figure 6.8 plots the actual lead WCA to the fitted lead WCA of the training sample.[22] The actual lead WCA ranges between -0.15 and 0.12 of total assets, bounded by the 1% winsorization. Ideally, the model's fitted values in Panel (a) would be tightly distributed around the slope of one corresponding to the dashed line. However, the in-sample data reveal that the fitted values in Panel (a) are all tightly clustered around the zero lead WCA, regardless of whether the actual lead WCA data points are negative or positive. Accordingly, the adjusted R^2 of the underlying data points is close to zero. Panel (b) plots the residuals, i.e., the difference between the actual and the fitted values, which are tightly distributed around the slope of one. If the model had high explanatory power, a uniform distribution of the residuals around zero would have been expected.

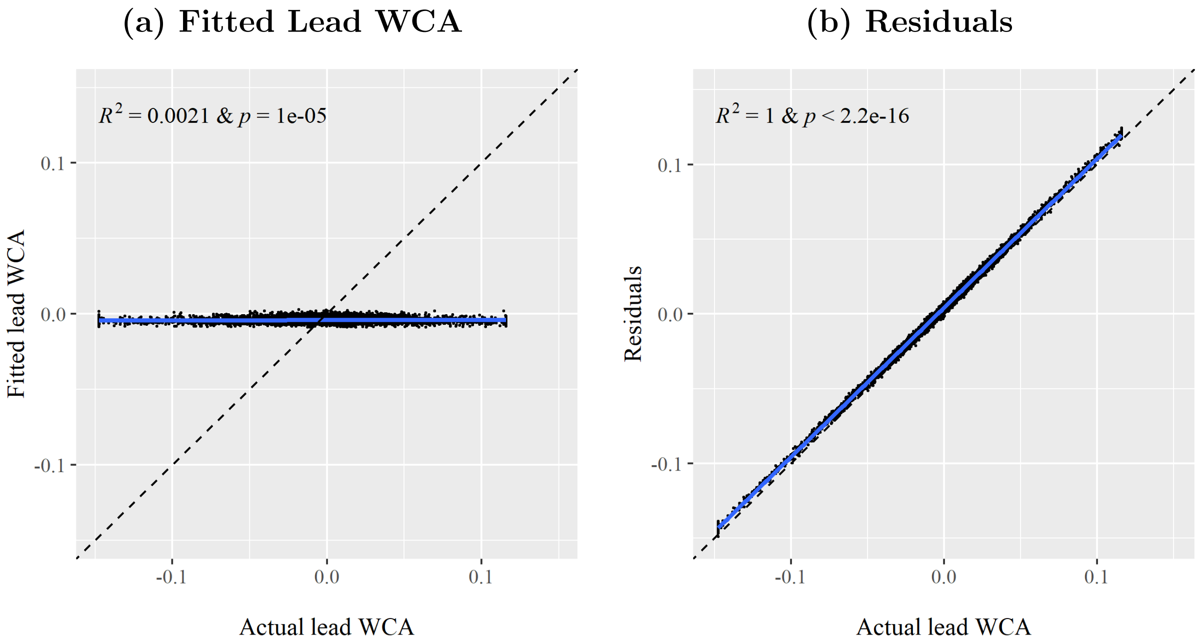

Figure 6.8: Scatter Plot of Fitted Lead WCA and Residuals. The figure presents the distribution of lead WCA on their fitted values in Panel (a) and residuals in Panel (b). Panel (a) shows that the fitted values do not correspond to the lead WCA. Instead, the residuals are correlated with the lead WCA, indicating the poor explanatory performance of the model.

[22]This figure reports only the fitted in-sample values and omits the out-of-sample predictions because they are graphically nearly identical.

It can be concluded that the variables in the modified Jones model cannot explain the variations in next year's WCA. The near-zero slope in Panel (a) suggests that the model is not able to capture the variations in lead WCA across firm-years. Therefore, it seems intuitive to include additional variables that could better capture the explanatory power of next year's WCA.[23]

6.5.2 OLS Regression With all Variables

This subsection uses all selected Refinitiv Worldscope variables, as described in the data section, predicts the following year's WCA and CompAcc, and compares the in-sample and out-of-sample results.

Regression Results

First, we run a within-sample regression including all variables on the one-year lead of WCA and the one-year lead of CompAcc. We obtain a comparatively high explanatory power with an adjusted R^2 of 13.5% (F-statistic: 3.9 on 496 and 8709 degrees of freedom) for lead WCA and 16.5% (F-statistic: 4.7) for lead CompAcc. The variables in the model appear to fit the available in-sample data better than the explanatory variables in the modified Jones model.[24] Next, we use the coefficients of the in-sample estimation to predict the out-of-sample accruals.

Error Measures

Table 6.6 presents the MAE, $\sqrt{MSE}$, MSE, and the decomposition of the MSE into K1, K2, and K3 for the lead WCA and the lead CompAcc, including all explanatory variables. The in-sample MSE for lead WCA (0.001232) yields a much smaller error

[23]Unreported results of the plots that include cash flow and its lag, as in formula (2) of the regression Table 6.4, show a marginally higher correlation, adjusted R^2, and therefore a steeper slope of the actual lead WCA training data with the fitted values.

[24]Section 6.5.5 discusses the importance of individual variables in more detail.

than that of the modified Jones model (0.001503). The out-of-sample MSE (0.001456) is considerably higher than the in-sample value but still yields a lower MSE than the out-of-sample modified Jones model (0.001548). The poorer out-of-sample performance, compared to the in-sample performance, is likely due to the OLS model's tendency to overfit the numerous variables. Nevertheless, the z-score remains insignificant, with a value of -0.3892.

Model	**MAE**	$\sqrt{MSE}$	**MSE**	**K1**	**K2**	**K3**
In-sample lead WCA	0.024689	0.035104	0.001232	0.00	40.23	59.77
Out-of-sample lead WCA	0.026833	0.038162	0.001456	0.03	35.68	64.28
In-sample lead CompAcc	0.057599	0.082242	0.006764	0.00	37.13	62.87
Out-of-sample lead Compacc	0.061343	0.087622	0.007678	0.02	29.95	70.03

Table 6.6: OLS Error Measures of all Explanatory Variables on Lead Accruals. The table presents the MAE, $\sqrt{MSE}$, and the MSE for in- and out-of-sample lead WCA resulting from a regression including all explanatory variables. The calculation of MSE and MAE is presented in Section 6.3.1. The right side of the table shows the MSE decomposition in K1, K2, and K3. The largest source of error is K3.

Similar results are found for lead CompAcc. The in-sample MSE (0.006764) is considerably lower than the out-of-sample MSE (0.007678), indicating that the in-sample regression is overfitted. The z-score is not significant, with a value of -0.6752.

The right side of Table 6.6 shows the decomposition of the MSE as the percentage of error attributable to K1, K2, and K3, where all variables are used to explain (for in-sample) or predict (for out-of-sample) next year's accruals. The decomposition of the MSE reveals that K2, the standard deviation of the train values minus the standard deviation of the fitted (predicted) values of the train (test) sample is with 40.2% (35.7%) clearly lower than in the regression with the modified Jones model. The variance and, thus, the standard deviation of the fitted (predicted) values are much higher and also more comparable to the actual train (test) data. This is because the models that include all variables can better adjust to the underlying data, as visible in Figure 6.9. Consequently, the K2 error is reduced, and a higher proportion of error is now attributed to K3.

In general, our broad set of variables appears to be better able to predict lead WCA than if only the variables of the modified Jones model were included. The out-of-sample lead WCA model, including all variables, is, with 0.001456, even lower than the corresponding value of the modified Jones model (0.001548). This finding is important because it shows that including a broad set of variables, in this case, is better than restricting the dataset to the variables of the modified Jones model. Accordingly, the standard OLS appears to assign a higher weight to in-sample variables that are also relevant in out-of-sample measurements. Section 6.5.5 extracts the most important variables for predicting the lead WCA.

Distribution of Actual to Fitted and Predicted Accruals

Figure 6.9 shows the lead WCA and lead CompAcc with in-sample results on the left and out-of-sample results on the right. The extensive use of explanatory variables results in a comparatively higher in-sample adjusted R^2 for both, lead WCA (18%) and lead CompAcc (21%). The in-sample fitted values tend to have positive skewness, which contrasts with the above Figure 6.8, which includes only the explanatory variables of the modified Jones model. The larger standard deviation of the fitted (predicted) values results in a lower fraction of the error term K2 of the overall MSE.

The slopes of the out-of-sample results appear to be weaker than the in-sample slopes. Accordingly, the out-of-sample adjusted R^2 (7.8% for lead WCA and 10% for lead CompAcc) is lower. The OLS model tends to overfit the training data because some variables may be useful for in-sample explanatory purposes but not for predicting out-of-sample accruals. This is visible on some data points in Panels (b) and (d), where they appear to be far from the diagonal dashed line.

We conclude from this subsection that adding more variables increases the explanatory power of in-sample methods as measured by the adjusted R^2. The improvements for out-of-sample predictions may be substantially smaller, although still superior to using only the explanatory variables of the modified Jones model. The main weakness of the OLS is that it does not include any procedure that accounts for the optimal trade-off between bias and variance, which leads to overfitting in out-of-sample data. Therefore, it seems reasonable to apply enhanced methods that automatically account for the overfitting.

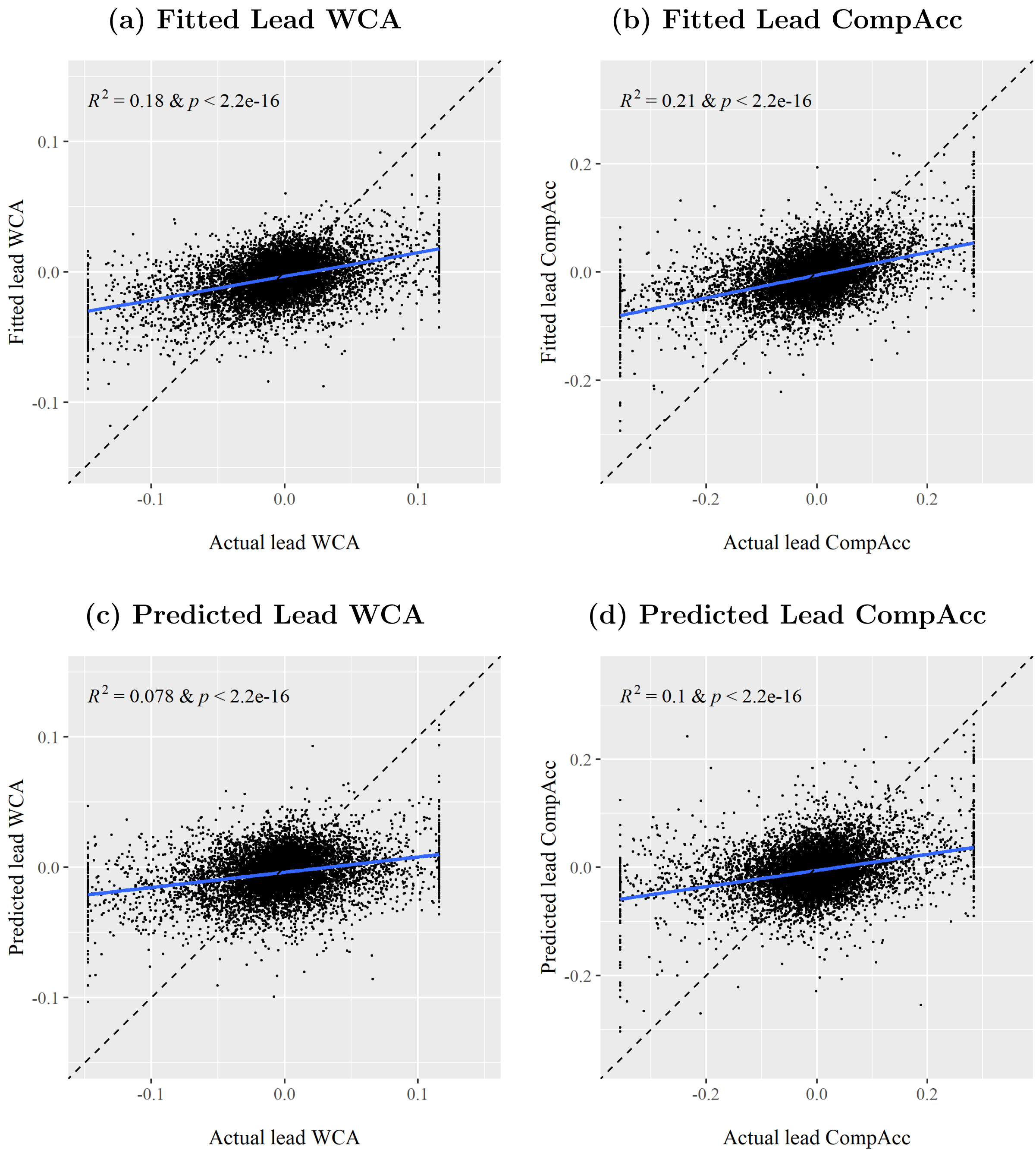

Figure 6.9: Scatter Plot of Fitted and Predicted Lead WCA. The figure presents the distribution of lead WCA to their fitted values in Panel (a) and (b) and predicted values in Panel (c) and (d). The fitted accruals show a partial alignment with the actual accruals. However, this effect is weaker for the predicted values.

6.5.3 Machine Learning Methods

This subsection evaluates the accuracy of the various machine learning models presented in the methods section and compares the evaluation metrics to the OLS models. We begin by reporting the out-of-sample error measures, followed by the decomposition elements of the MSE, and end by plotting the actual versus predicted values of the lead accruals.

Error Measures on Lead WCA

Table 6.7 presents the out-of-sample MAE, $\sqrt{MSE}$, and the MSE for next year's WCA for the OLS model and our machine learning models, as presented in the methods section. The machine learning models generally outperform the OLS model because they all have a systematic method for distinguishing important and unimportant variables that prevents overfitting.

Model	**MAE**	$\sqrt{MSE}$	**MSE**	**K1**	**K2**	**K3**
OLS Lead WCA	0.026833	0.038162	0.001456	0.03	35.68	64.28
LASSO Lead WCA	0.025849	0.037412	0.001400	0.02	53.08	46.90
Random Forest Lead WCA	0.025826	0.037731	0.001424	0.04	61.94	38.02
SVM Lead WCA	0.026278	0.038156	0.001456	0.00	50.63	49.36

Table 6.7: Machine Learning Models Error Measures of all Variables on Lead WCA. The table presents the MAE, $\sqrt{MSE}$, and the MSE for in- and out-of-sample lead WCA resulting from the machine learning models, including all explanatory variables. The computation of MSE and MAE is presented in Section 6.3.1.
The right side of the table shows the MSE decomposition into K1, K2, and K3.

The LASSO model has the lowest out-of-sample MSE error and thus can be considered the most accurate model. The variable selection with the built-in penalty performs well, resulting in a 3.8% $(1 - \frac{0.001400}{0.001456})$ reduction in MSE compared to the OLS model. The MAE indicates that the average prediction error of LASSO is 2.6% of total assets, which is approximately a third of the net working capital balance sheet value.

Our LASSO model does not include any higher degree polynomials or interactions of descriptive variables. We also do not split the variables at zero to measure the effects of positive and negative variables separately, as Ball and Shivakumar (2006) did for cash flows. Thus, our LASSO model measures only linear relationships, although there is evidence that the correlations of accounting items are nonlinear, as reported in the literature section. Nevertheless, random forests and SVM do not outperform LASSO. This finding is interesting because it could have different explanations. One explanation would be that the nonlinearities could be less severe if many variables were included. Potentially, other variables not included in typical accrual models could capture some of the nonlinear relationships. The second explanation could be that the random forest and SVM model might account for nonlinearities but perform poorly on linear variables, so they do not result in a lower MSE than the LASSO model. Finally, both, the LASSO and nonlinear methods have their weaknesses, and the MSE could possibly be further reduced by using more sophisticated models.

As previous research has shown, ridge regression usually performs better when some variables are highly correlated (Dai et al., 2018; Z. Liu et al., 2017). We find only few highly correlated variables in our dataset. In particular financial statement variables with many zeros have a low correlation precisely because of the many zeros. We run an additional unreported test using a simple ridge regression (in the R function *glmnet* the parameter alpha = 0 for ridge, instead of 1 for LASSO), which results in a slightly worse MSE than for the standard LASSO model.

The random forest model is particularly well suited for capturing nonlinear relationships because it splits the variables and groups them into subgroups. The results of our random forest show that the model cannot outperform the LASSO in terms of our accuracy measure, the MSE, but still performs better than the standard OLS and SVM. The MAE is the lowest for the random forest. Likely the random forest model was unable to account for outliers as well as the LASSO model, which results in a higher MSE but a lower MAE. Figure 6.10 illustrates an example of the splitting into different variables of one regression tree. The numbers in the boxes represent the predicted lead WCA at the actual node. The percentages represent the share of the sample for each node.

The SVM does not appear to consistently outperform the OLS model. The MSE is the same as the OLS model, indicating that the SVM did not result in an improvement. Apparently, our SVM maximum margin model cannot use its nonlinear properties to improve the prediction accuracy.[25] The MAE is marginally lower than for the OLS model.

The disaggregation of the MSE into K1, K2, and K3 reveals that the machine learning models produce a higher error in variance (K2) but better capture the correlation between

[25]The SVM uses the radial kernel by default. Unreported results using the polynomial kernel are slightly worse and are not included in the results.

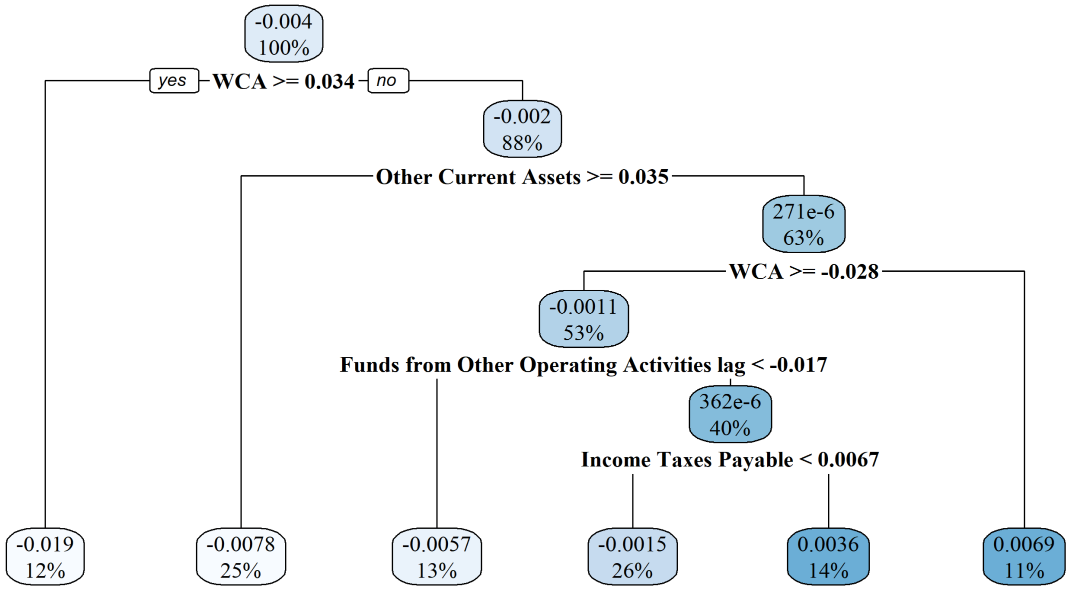

Figure 6.10: Regression Tree of Lead WCA. The figure shows the splitting of the explanatory variables to predict the lead WCA for each leave node. The numbers in the boxes represent the predicted lead WCA at the actual node. The percentages represent the share of the sample for each node.

actual and predicted values (K3). More precisely, the higher K2 value can be explained due to the fact that the standard deviation of the predicted values is lower (0.0123 for LASSO and 0.0165 for OLS) compared to the standard deviation of the test sample (0.0393). Therefore, the higher difference between the two standard deviations leads to a relatively higher percentage error of K2 of the total MSE. In turn, the slightly higher correlation of predicted and test data of the LASSO model, combined with the lower standard error in predictions, leads to a clearly smaller K3 value.

Actual and Predicted Accruals

Figure 6.11 presents the actual test (x-axis) to the out-of-sample predicted observations (y-axis) for the OLS (top left), LASSO (top right), random forest (bottom left), and SVM (bottom right). Ideally, the observations would be narrowly distributed around the slope of one, which is represented by the dashed line. All models predict values that do not fully reflect the following year's level of WCA in the test sample. Rather, they predict lead WCA close to zero with slightly higher values for more positive actual lead WCA.

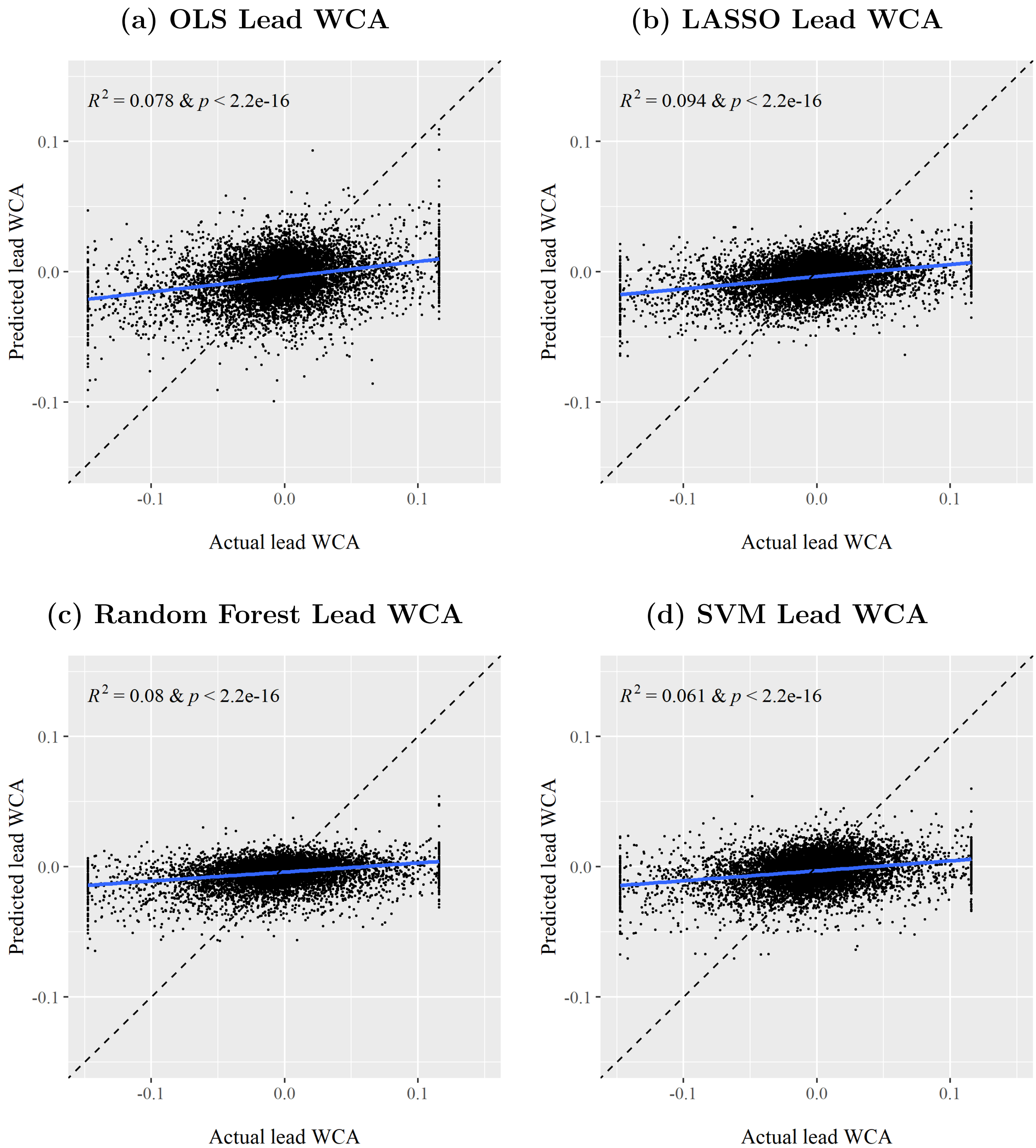

Figure 6.11: Scatter Plot of Predicted Lead WCA. The figure presents the out-of-sample performance of lead WCA for machine learning models. The adjusted R^2 values represent the explanatory power of a regression function of test data on predicted values of training data.

The OLS regression shows a comparatively broad distribution of predicted lead WCA on the y-axis. The extreme values could arise due to the overfitting of some variables to individual observations of the in-sample fitted values. Machine learning methods include a penalty or validation mechanism to prevent overfitting. The LASSO method shows the highest adjusted R^2 (9.4%) of all models. It appears that LASSO can shrink certain variables to result in fewer extreme outliers. The random forest model also shows a small positive relationship between actual and test values combined with a relatively narrow distribution of individual observations. The adjusted R^2 of the SVM model is lower than for the OLS model.

In summary, the LASSO and random forest represent an improvement over the standard OLS model in predicting next year's accruals. The SVM obtains the same MSE as the OLS regression. Nevertheless, all models perform relatively poorly in predicting next year's accruals with a relatively high error. The MAE of roughly one-third of the net working capital shows that a large share of the variation remains unpredicted. In addition, all models issue too high (low) predictions for future negative (positive) lead WCA and lead CompAcc.

Error Measures on Lead CompAcc

Table 6.8 presents the error measures for lead CompAcc. LASSO is again the most accurate model for out-of-sample prediction. The reduction in MSE is 5.4% $(1 - \frac{0.007264}{0.007678})$, which is larger than for the lead WCA. The MAE is 5.9% of total assets for the one-year forecasted CompAcc using the LASSO model. The random forest model performs best after the LASSO model but obtains the lowest MAE value. The SVM model performs the worst of the three supervised machine learning models but is still better than the OLS model. Interestingly, the SVM, that performs worst of the supervised machine learning models, has the lowest MAE. Potentially, the SVM does not account for outliers as strong as the LASSO model and predicts more "average" values. The graphs for lead CompAcc are similar to the lead WCA actual and predicted figures and, therefore, we do not report them separately.

6.5.4 Robustness Tests

Our results may have been affected by the data selection choices. To assess the robustness of our models, we run two robustness tests in which we modify our dataset.

Model	MAE	$\sqrt{MSE}$	MSE	K1	K2	K3
OLS Lead CompAcc	0.061343	0.087622	0.007678	0.02	29.95	70.03
LASSO Lead CompAcc	0.058635	0.085230	0.007264	0.02	48.11	51.87
Random Forest Lead CompAcc	0.058292	0.085546	0.007318	0.14	53.30	46.56
SVM Lead CompAcc	0.058086	0.086130	0.007418	0.22	47.55	52.23

Table 6.8: Machine Learning Models Error Measures of all Variables on Lead CompAcc. The table presents the MAE, $\sqrt{MSE}$, and the MSE for in- and out-of-sample lead WCA, resulting from the machine learning models including all explanatory variables. The MSE and MAE computation is presented in Section 6.3.1.
The right side of the table shows the MSE decomposition into K1, K2, and K3.

Exclude More Exact Zeros in Balance Sheet Items

In the data section of the financial statement variables, we have required at least five percent of non-zero and non-NA, which resulted in 74 variables being dropped. For the remaining 198 financial statement variables, we assumed that a missing value in a financial statement variable means that this accounting item does not exist at the firm level and therefore coded it as zero. In other words, all available financial statement variables in the Worldscope database are assumed to be fully reported for the 20% of each year's largest firm-years and are therefore independent of whether Worldscope enters a zero or marks them as missing.[26]

Although this assumption might appear plausible, some variables consist of a large proportion of zeros, which could affect the MSE of the models used above. To account for this potential drawback, we rerun our models with a dataset in which at least 95% of the financial statement variables must consist of non-zero and non-NA values.

The modified dataset contains 37 instead of 198 financial statement variables, while 135 financial statement variables are dropped. The stricter data requirements do not affect the number of firm-years, but only the variables. Accordingly, the descriptive statistics in Table 6.3 on accruals and their components remain unchanged.

Table 6.9 reports the MAE, $\sqrt{MSE}$, the MSE, and the percentage change of the MSE in relation to the base Table 6.7. The MSE of the OLS regression is lower by 2.8% compared to the main results in Table 6.7. This is interesting because, in this specific

[26]See the data Section 6.4 for an example regarding Amazon's raw materials.

case, fewer variables led to a more precise out-of-sample OLS prediction. It is possible that in the initial setting with some variables consisting of mostly zeros, the OLS model used these variables to account for specific outliers in the training sample, also known as overfitting, which then results in a higher error for out-of-sample predictions.

Model	**MAE**	$\sqrt{MSE}$	**MSE**	**Δ MSE (%)**
OLS Lead WCA	0.026151	0.037626	0.001416	-2.79
LASSO Lead WCA	0.025909	0.037448	0.001402	0.19
Random Forest Lead WCA	0.025846	0.037745	0.001425	0.08
SVM Lead WCA	0.025729	0.037656	0.001418	-2.60

Table 6.9: Machine Learning Models Error Measures With Fewer Financial Variables on Lead WCA. The table presents the MAE, $\sqrt{MSE}$, and the MSE for in- and out-of-sample lead WCA, resulting from the machine learning models that include all explanatory variables for financial variables with equal to or more than 95% of non-zero and non-NA data. The MSE and MAE computation is presented in Section 6.3.1.
The right side of the table shows the MSE decomposition into K1, K2, and K3.

The LASSO has a very similar MSE (0.001402 instead of 0.001400). It appears that the LASSO could identify the important variables for out-of-sample predictions in the baseline dataset, although many zeros were included. The random forest is marginally less accurate than the LASSO. The comparatively large reduction in the MSE for the OLS model results in the OLS model being more accurate than the random forest and the SVM, of which the SVM model results in a lower MSE of 2.6%, compared to the main results in Table 6.7. This shows that machine learning models are not necessarily superior to standard OLS regression. In general, the results show that a more restrictive choice of variables can increase the out-of-sample prediction accuracy.

Table 6.10 reports the error measures for the different models used to predict the lead CompAcc. All models perform better with fewer variables, although the increase in accuracy is small for the machine learning models. The OLS model shows the largest increase in predictive accuracy compared to the baseline sample. It remains to be discussed why the LASSO model performs slightly better in this particular case, since the model should be able to identify the relevant variables and shrink the others. One possible explanation could be that the training and test samples differ marginally, with specific individual outliers occurring in the test but not in the training dataset.

Model	MAE	$\sqrt{MSE}$	MSE	Δ MSE (%)
OLS Lead CompAcc	0.059331	0.085705	0.007345	-4.33
LASSO Lead CompAcc	0.058738	0.085156	0.007252	-0.17
Random Forest Lead CompAcc	0.058079	0.085194	0.007258	-0.82
SVM Lead CompAcc	0.057445	0.085551	0.007319	-1.34

Table 6.10: Machine Learning Models Error Measures With Fewer Financial Variables on Lead CompAcc. The table presents the MAE, $\sqrt{MSE}$, and the MSE for in- and out-of-sample lead CompAcc, resulting from the machine learning models that include all explanatory variables for financial variables with equal to or more than 95% non-zero and non-NA data. The MSE and MAE computation is presented in Section 6.3.1.
The right side of the table shows the MSE decomposition into K1, K2, and K3.

Including More Lags

Not only can additional variables of the actual year increase the prediction accuracy, but also leads and lags result in higher explanatory power R^2. For example, Dechow et al. (2012, p. 299) show that the Dechow and Dichev (2002) and McNichols (2002) models, both of which include a lead and a lag of cash flow from operations, result in a higher adjusted R^2 compared to models without leads and lags of cash flows. Similarly, Larson et al. (2018, p. 855) add up to eight leads and lags of cash flows to explain CompAcc. The three years before and after the actual year are highly significant. Cash flows further in the future or past are steadily losing significance.

Given the literature above, additional lags could increase the predictive power of the out-of-sample models. Since we want to predict the following year's accruals, we only include additional lags, not leads.

One of the drawbacks of using data from other periods is that it reduces the number of firm-years. Creating one lead and three lags implicitly means that a publicly traded firm must be among the 20% with the highest market capitalization in the U.S. for five consecutive years to be included in the sample. At the same time, however, the number of explanatory variables increases. Each lag of the 196 financial statement variables adds another 196 variables. With three lags, our dataset increases accordingly to 784 financial statement variables.

We rerun our data selection process as described in the data Section 5.5, with the only change being that we require three lags of data instead of one. The addition of two lags cuts the sample to 15,802 firm-years but increases the variables to 839. The different firm-years selection slightly changes our summary statistics used to calculate accruals, as shown in Table 6.11.

(a) WCA Components

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
WCA	15,802	-0.004	0.037	-0.138	-0.019	0.014	0.110
Net Working Capital	15,802	0.071	0.130	-0.503	-0.006	0.138	0.753
Current Assets	15,802	0.383	0.213	0.045	0.203	0.536	0.879
Cash	15,802	0.125	0.148	0.001	0.020	0.175	0.671
Current Liabilities	15,802	0.221	0.117	0.045	0.132	0.288	0.591
Short Term Debt	15,802	0.034	0.047	0.000	0.001	0.048	0.249

(b) CompAcc Components

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
CompAcc	15,802	-0.004	0.082	-0.289	-0.038	0.033	0.260
Common Equity	15,802	0.445	0.190	0.060	0.306	0.568	0.897
Cash	15,802	0.125	0.148	0.001	0.020	0.175	0.671

Table 6.11: Summary Statistics of Robustness Test. The table shows the summary statistics for the accrual variables with three-year lags.
Panel (a) presents the summary statistics of WCA as defined in Table 6.3
Panel (b) represents the CompAcc computation method as defined in Table 6.3
All variables are scaled by Total Assets (WC #02999).

Table 6.12 reports the error measures for our modified dataset. Panel (a) presents the lead WCA and Panel (b) shows the lead CompAcc. It is important to note that the MSE cannot be directly compared to our baseline dataset because the firm-year sample in this robustness test is reduced by 14.0% ($1 - \frac{15,843}{18,412}$). The additional requirement of two consecutive years of available data could introduce a selection bias towards larger firms with different accrual structures, as previous research shows that larger firms tend to have a lower proportion of accruals relative to their assets (Watts & Zimmerman, 1978).

For lead WCA, the MSE of the LASSO model is the best predictor, while the random forest and SVM are close. The mean error relative to total assets is 2.4% for the LASSO model and the random forest. The OLS model has the largest MAE of 2.6% of total assets. Still, all models have considerable errors. The average net working capital is 7.3% of total assets. This means that the average prediction error is roughly one-third of net working capital.

(a) Machine Learning Models Error Measures With Three Lags on Lead WCA

Model	MAE	$\sqrt{MSE}$	MSE	Δ MSE (%)
OLS Lead WCA	0.026454	0.037334	0.001394	-4.29
LASSO Lead WCA	0.024469	0.035481	0.001259	-10.06
Random Forest Lead WCA	0.024446	0.035602	0.001267	-10.97
SVM Lead WCA	0.024840	0.035959	0.001293	-11.19

(b) Machine Learning Models Error Measures With Three Lags on Lead CompAcc

Model	MAE	$\sqrt{MSE}$	MSE	Δ MSE (%)
OLS Lead CompAcc	0.060388	0.085412	0.007295	-4.98
LASSO Lead CompAcc	0.055076	0.079800	0.006368	-12.34
Random Forest Lead CompAcc	0.054973	0.080345	0.006455	-11.79
SVM Lead CompAcc	0.055060	0.081186	0.006591	-11.15

Table 6.12: Machine Learning Models Error Measures With Three Lags on Lead Accruals. The table presents the MAE, $\sqrt{MSE}$, and the MSE for in- and out-of-sample lead CompAcc, resulting from the machine learning models that include all explanatory variables for financial variables with three-year lags. The computation of MSE and MAE is presented in Section 6.3.1.
The right side of the table shows the MSE decomposition into K1, K2, and K3.

For lead CompAcc reported in Panel (b), the MSE and $\sqrt{MSE}$ are again lowest for the LASSO model, followed by the random forest and the SVM. In contrast, the random forest performs best for the MAE, followed by the SVM. The average measurement error of next year's CompAcc remains relatively high at more than 5.5% of total assets.

6.5.5 Most Important Variables in Predicting Next Year's Accruals

One of the aims of this chapter is to select the variables that best predict the following year's accruals. James et al. (2021, p. 433) achieved the best predictive performance and thus the lowest MSE of baseball players' log salary in the "Hitters" dataset by selecting only four variables with the LASSO, which are then simply used in an OLS model. We aim to use a similar variable selection procedure in which we select the five most important variables for predicting next year's accruals and evaluate the predictive performance of an OLS model.

It may appear easiest to simply extract the most important variables of an OLS model. However, depending on the train and test samples split, the most important variables vary frequently. The LASSO model includes an in-sample cross-validation, which prevents the overfitting of the training sample. Therefore, the most important variables of the LASSO model are much more stable across different sampling seeds.

We use a slightly different data selection for the computation than those used previously. The new dataset combines both selection criteria of the robustness tests above, i.e., three lags and variables with 95% or more of non-zeros and non-missing values for financial statement items. It is particularly important to have a limited proportion of zeros to prevent LASSO from assigning high importance to variables with predominantly zeros to achieve a stable selection of the most important variables through multiple sampling procedures. Variables with a high proportion of zeros tend to return different variables for different sample seeds, even for the LASSO model.

For the variable selection, we use the R-package "vip"(Greenwell & Boehmke, 2020). In linear regressions such as the OLS and LASSO models, the variable importance is simply expressed as the t-statistic of the estimated coefficient (Greenwell & Boehmke, 2020, p. 348). Accordingly, we use the function *vi_model* to identify the five variables with the largest t-statistics of the LASSO model variables on the year ahead WCA and the one year ahead CompAcc.

Table 6.13 reports the most important variables of the LASSO lead WCA model in Panel (a) and the lead CompAcc in Panel (b). The variable with the largest weight is WCA of the actual year. Other important variables are Funds from/for Other Operating Activities (WC #4831), Short Term Debt & Current Portion of Long Term Debt (WC #3051), cash flow from operations labeled as Funds from Operations (WC #4201), and lagged Cash & Short Term Investments (WC #2001). All variables, except for lagged cash, negatively correlate with next year's WCA.

It appears plausible that high WCA of the current year will lead to a reduction of WCA and reverse to the mean. An exceptionally high actual WCA of the current year could be caused by an intended sale of assets within the next year. If these assets are sold

(a) Lead WCA LASSO Selected Most Important Variables

Variable	Importance	Sign
WCA	0.1790	NEG
Funds from Other Operating Activities	0.0701	NEG
Short Term Debt	0.0655	NEG
CFO	0.0622	NEG
Cash lag	0.0500	POS

(b) Lead CompAcc LASSO Selected Most Important Variables

Variable	Importance	Sign
Short Term Debt	0.1903	POS
Cash	0.1793	POS
Capital Expenditure	0.1748	POS
Change in Cash	0.1413	POS
Common Equity	0.0933	NEG

Table 6.13: Most Important Variables for Predicting Accruals. The table shows the most important variables selected by the LASSO model. For lead WCA, these variables are WCA of the current year, Funds from/for Other Operating Activities (WC #4831), Short Term Debt & Current Portion of Long Term Debt (WC #3051), cash flow from operations labeled as Funds from Operations (WC #4201), and lagged Cash & Short Term Investments (WC #2001).
The most important lead CompAcc variables are Short Term Debt & Current Portion of Long Term Debt (WC #3051), Cash & Short Term Investments (WC #2001), Capital Expenditures Additions to Fixed Assets (WC # 04601), Increase/Decrease in Cash & Short Term Investments (WC #4851), and Common Equity (WC #3501).

during the next year, the WCA, as a percentage of total assets, will decrease in relation to the actual year. The negative correlation of cash flows of operations to lead WCA is consistent with previous research (Dechow et al., 2012). For example, if this year‘s debtors paid off their debt relatively quickly, this would result in higher cash flow this year. If this returns to normal levels next year, the cash flow will be lower, and next year's WCA consequently higher.

For lead CompAcc, the most important variables are Short Term Debt & Current Portion of Long Term Debt (WC #3051), Cash & Short Term Investments (WC #2001), Capital Expenditures Additions to Fixed Assets (WC # 04601), Increase/Decrease in Cash & Short Term Investments (WC #4851), and Common Equity (WC #3501).

Subsequently, these variables are employed in two OLS regressions. Table 6.14 presents the in-sample regression results of lead WCA (1) and lead CompAcc (2). For lead WCA, the in-sample adjusted R^2 is 2.8%, which is relatively low but not negligible. The explanatory variables of the modified Jones model in Table 6.4 yielded almost zero explanatory

	Dependent variable:	
	Lead WCA	Lead CompAcc
	(1)	(2)
Intercept	-0.003^{***}	0.011^{***}
	(0.001)	(0.003)
WCA	-0.142^{***}	
	(0.011)	
Funds from/for Other Operating Activities	0.021^{*}	
	(0.013)	
Short Term Debt	-0.038^{***}	0.085^{***}
	(0.009)	(0.021)
Cash lag	0.005^{**}	
	(0.002)	
CFO	-0.007	
	(0.006)	
Cash		0.111^{***}
		(0.007)
Capital Expenditures		0.078^{***}
		(0.019)
Change in Cash		0.236^{***}
		(0.017)
Common Equity		-0.094^{***}
		(0.006)
Observations	9,206	9,206
R^2	0.028	0.072
Adjusted R^2	0.028	0.072
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 6.14: Regression Results of Most Important Variables for Predicting Accruals. The regression table shows the most important variables in predicting lead accruals. Regression (1) uses the most important variables in predicting lead WCA, selected by the LASSO model. Regression (2) uses the most important variables in predicting lead CompAcc, selected by the LASSO model. The variable definitions correspond to those in Table 6.3.

power, while in contrast, the regression explaining lead WCA including all variables yielded an in-sample adjusted R^2 of 13.5%. We caution against comparing the results with other adjusted R^2 values because this dataset contains fewer firm-years.

Interestingly not all variables of the lead WCA model are significant. Cash flow from operations does not appear to have high predictive power for lead WCA in this model, even though the LASSO model has selected it as one of the five most important variables. It is challenging to draw pertinent conclusions because it would require an in-depth analysis, which we leave for future research. For lead CompAcc, the in-sample adjusted R^2 at 7.2% is clearly higher than for the lead WCA. Further, all lead CompAcc variables are significant at the 1% level.

6.5.6 Discussion of Results

Höglund (2012) used neural networks to predict same-year accruals and found that his model outperformed standard OLS. In his conclusion, he addresses whether neural network models should be used more frequently in accounting research. He concludes that "it is likely that accounting researchers will continue using models based on linear and piecewise-linear regression for estimating discretionary accruals" (Höglund, 2012, p. 9570) because the application of neural networks would require more advanced skills, and the results would be less intuitive to interpret.

Indeed, the question is whether machine learning models add sufficient value at a reasonable cost. The benefit of the enhanced models is an increased prediction accuracy. Our LASSO model improves the prediction accuracy, as measured by the MSE, by 3.8% over the standard out-of-sample OLS model. The cost is that additional understanding of statistical models is required, and the models take more time to run on computers.

Although the above results may not appear to be a substantial improvement, it is important to consider that for many explanatory variables, the OLS model may tend to overfit the in-sample data. The MSE of the in-sample lead WCA is 15.4% ($1 - \frac{0.001232}{0.001456}$) lower than that of the out-of-sample lead WCA. The overfitting of the coefficients to the in-sample data means that the results are not generalizable to data outside the dataset.

Nowadays, programming packages such as the "caret" package for R and Python provide a simple implementation of train-test splits and numerous statistical models (Kuhn, 2019; Moez, 2020). They require only marginally more code than running a standard in-sample OLS model. Conclusive, accounting research on accruals that use many explanatory variables should at least use a LASSO model because it increases the external validity of the findings. The LASSO model is easy to understand because it is based on the OLS model and extended with a penalty to shrink unimportant variables, it accounts for overfitting, and the interpretation of the linear coefficients is straightforward.

6.6 Conclusion

This paper aims to predict next year's accruals. We started by creating an extensive dataset consisting of 447 variables from the Refinitiv Worldscope database. We applied the supervised machine learning models LASSO, random forest, and SVM to next year's accruals and compared them with the prediction accuracy of the OLS model. Finally, we selected the five most important variables from the LASSO model and tested the explanatory power for next year's accruals. Our findings can be summarized as follows.

1. The explanatory variables of the modified Jones model have almost no explanatory power for next year's WCA. Including more explanatory variables leads to a higher accuracy, which increases the in-sample adjusted R^2 to 13.5%.

2. With a broad set of explanatory variables, the standard OLS regression tends to overfit on in-sample explanatory tests. The in-sample MSE for next year's WCA is 15.4% lower than the out-of-sample MSE. Although the z-score is not significant, there is likely an effect of overfitting the training sample. Likewise, when comparing the actual with the fitted (for in-sample) or the predicted (for out-of-sample) next year's WCA, the in-sample adjusted R^2 is, at 18.0%, considerably higher than the out-of-sample prediction with 7.8%. Large datasets with numerous explanatory variables in accounting research require an out-of-sample prediction method that avoids overfitting. Out-of-sample measures that include

all variables are more accurate than out-of-sample measures that include only the variables of the modified Jones model. The additional predictive power of the supplementary variables seems to outweigh the effect of overfitting.

3. Basic supervised machine learning models generally increase the accuracy of the out-of-sample measurement in next year's accruals compared to a standard OLS model. Consistent through all tests, the LASSO model performs best, measured by the out-of-sample MSE, even though it does not account for nonlinear relationships. Random forests and SVM models, which would be able to capture nonlinear relationships, tend to perform better than the OLS model. However, machine learning models are not always more accurate than the OLS model. In a robustness test where we only include financial statement variables with at least 95% of non-zero and non-missing values, the OLS model performed slightly better than the SVM and random forest.

4. Even though machine learning models are an improvement over OLS models, changes in accruals for one year ahead remain challenging to predict. The LASSO model's most accurate prediction of next year's WCA with three lags of financial statement variables results in a MAE of 2.4% of total assets. In relation, the average net working capital is 7.3% of total assets. This means that the average error in the most precise prediction is still roughly one-third of net working capital.

5. The five most important variables can capture a small fraction of the one-year ahead WCA and a more decent fraction of the one-year ahead CompAcc. For the lead CompAcc model, all selected variables are significant at the one percent level.

This study is not free of limitations. First, we used standard machine learning models with default parameters that are not optimized for the data at hand. Most likely, the prediction accuracy would increase with optimal model parameters. Second, the Worldscope industrial template includes 1,257 individual variables. Although we went through all variables individually, some excluded variables could possibly contain relevant information for predicting accruals. Finally, other data from databases different from Worldscope could yield different results.

Future research could take several paths to increase the predictive power of future accruals. One way would be to account for the nonlinear relationships of accounting items by adding a partitioning variable to separate positive and negative values, as Ball and Shivakumar (2006) did. It is expected that this would increase the predictive power since our most accurate model, the LASSO, does not account for nonlinearities. In addition, it might be interesting to add some non-numerical data. Text-based information can now be processed with machine learning models and could result in interesting findings.

Another path would be to use more enhanced models. Our study does not use unsupervised machine learning models such as neural networks. With the optimal parameters of a neural network and considerably more computational power, the results could be increased substantially. Also, a combination of different machine learning methods, known as “ensemble learning”, could be of value.

If the intent is to understand and establish potentially causal relationships about how earnings management is actually conducted, it might be promising to evaluate why the five most important variables have predictive power for next year’s accruals. A thorough data analysis could lead to a better understanding of the accrual building and dissolution process. Finally, it could be insightful to conduct qualitative or behavioral studies in which researchers seek to better understand the incentives behind earnings management. After all, earnings management is actively practiced by humans.

Chapter 7
CONCLUSION

In this thesis, we critically examine the literature on earnings discontinuities and earnings management models. We identify several weaknesses in the existing research and aim to improve the understanding of earnings management. The conclusion first summarizes the key findings of each chapter and then ends with suggestions for future research.

Chapter 2 reviews the literature on earnings discontinuities. We emphasize that the standardized differences test statistic should not be used to compare samples of different sizes, as this may lead to erroneous conclusions about the magnitude of the kink. We analyze the U.S. literature that finds an abrupt disappearance of the zero earnings discontinuity around the Sarbanes-Oxley Act (SOX) introduction (Gilliam et al., 2015). In contrast, the European literature reports only a slight decline in discontinuity following the mandatory adoption of International Financial Reporting Standards (IFRS) in 2005 (Trimble, 2018). For earnings changes and earnings surprise benchmarks, we caution that discontinuities might be due to earnings management but to mechanical effects. To interpret the magnitude of the discontinuity, one would need to know how the distribution would look in the absence of earnings management.

Chapter 3 is a joint paper with Peter Fiechter and Martin Wallmeier that examines the disappearance of the zero earnings discontinuity found by Gilliam et al. (2015) in the U.S. around the introduction of the SOX. It identifies the dotcom boom at the turn of the millennium as a potential confounding factor when numerous firms went public with almost no sales but high market capitalization. This mechanically leads to more firms reporting small losses, which reduces the appearance of the zero earnings kink during this period. We exclude these firms with almost no revenues and reach two conclusions. First, it is unclear whether the zero earnings discontinuity has completely disappeared. There are still more firms reporting a small profit than a small loss, even after the introduction of the SOX. Second, if the SOX were causally responsible for the decline in the zero earnings discontinuity, one would expect an abrupt disappearance. However, our results indicate that such a decline occurred successively over time.

Chapter 4 is a joint paper with Martin Wallmeier that investigates the zero earnings discontinuity in Europe and finds a robust remaining kink contrasting U.S. findings. Country-level analyses show that countries with a higher Uncertainty Avoidance Index (UAI) also have higher zero earnings discontinuities. The results suggest that despite the adoption of the IFRS, managers of European firms still have room to manage their earnings to reach the zero earnings threshold. The cultural dimension seems more important than factors indicating a decline in the zero earnings kink.

Chapter 5 discusses firm-level earnings management models. The main criticism is that the discretionary portion assigned may not truly arise from earnings management but rather from misspecifications of earnings management models. We address this criticism and incorporate earnings into an extended model that successfully accounts for the undesirable correlation between earnings and discretionary accruals. We also derive a distributional model that measures firm-level earnings management based on earnings. We evaluate the validity of these three models using the Accounting and Auditing Enforcement Releases (AAER) sample of the Securities and Exchange Commission (SEC), which includes firms known to have manipulated earnings. First, we find that firms with overstated earnings have significantly higher discretionary accruals under the modified and extended Jones models than firms not included in the AAER sample. Our distributional model does not provide a clear indication. The regression-based modified and extended Jones models perform slightly better than a random selection. The distributional model does not outperform a random selection. Likely, the distributional model performs comparatively poorly because it assigns a higher level of discretionary accruals to specific earnings regions. However, no such clustering is observed in the AAER sample.

Chapter 6 predicts next year's accruals and aims to improve understanding of the accrual formation process. For this purpose, we use an extensive set of 447 explanatory variables with different supervised machine learning models. The main results can be summarized as follows. First, the explanatory variables of the modified Jones model have almost no explanatory power for next year's working capital accruals (WCA). However, the explanatory power increases when all explanatory variables are used. Second, the ordinary least squares (OLS) model overfits if all 447 explanatory variables are included. Third, the least absolute shrinkage and selection operator (LASSO) model performs best among all models, even though it cannot account for nonlinear relationships. The outperformance of the LASSO is likely related to its ability to separate important from unimportant variables. Finally, next year's accruals remain difficult to predict even with an extensive dataset and more sophisticated measurement methods. The lowest mean absolute error for WCA is 2.4% of total assets. For future research, it might be interesting to investigate why some variables have predictive power for next year's accruals. This could help identify causal relationships in the accrual formation process.

If an overall conclusion had to be drawn from this thesis, it would likely be that current earnings management methods can capture part, but clearly not all, of earnings management, and that caution should be exercised in interpreting the results. Still, advances in improving earnings management models must be made. With this in mind, it is not surprising that the models are not without controversy in the academic literature.

We highlight several paths that future research could focus on to improve the understanding of earnings management. More available data and more powerful computers lead researchers to use these data and machine learning methods. Current research is reaching the point where almost all financial statement data are included in earnings management models. Future studies could incorporate non-numerical data, such as textual information. Managers may justify some accrual bookings in the 10-K and 10-Q filings which purely numerical models might interpret as earnings management.

A common challenge in research, of which earnings management is no exception, is that it is difficult to measure the exact causal effect of an independent variable on a dependent variable (i.e., earnings management). To do so, one would have to imagine a parallel world identical to ours, with the only difference being that firms in this parallel world report only true, unmanaged earnings. Then the difference in earnings between the two worlds could be identified as earnings management. Unfortunately, this is impossible, so research must resort to practicable methods. Future studies could seek situations where a sudden exogenous change in an independent variable occurs while all other things remain constant. Econometric research has identified measurement tools where in some situations, it is plausible to believe that the independent variable is truly independent and free of correlated omitted variables. If such circumstances can be found and analyzed, we may be able to better understand the magnitude of earnings management.

One possible approach to improving our understanding of the earnings management process might be to observe earnings decisions over a firm's life. At two points in time, a firm is potentially free of earnings management: when it is founded and when it is liquidated. This lifelong approach could help understand the incentives behind financial statements, which are just the results of this extensive and complex process.

Earnings management models might appear attractive to use because they are relatively simple to apply. However, the drawbacks of earnings management models are sometimes ignored, which can lead to a misinterpretation of the earnings management effect. We argue that future research on earnings management should agree to a quality standard or guideline similar to the suggestions of McNichols and Stubben (2018, pp. 234–242). We highlight two of their suggestions, the first of which is to "articulate the story" in the sense that the mechanism that might lead to earnings management should be explained and predictions made about "how, when, and where" (McNichols & Stubben, 2018, pp. 234–235). The second suggestion is to carefully select the earnings management model for the specific study. Depending on the research question, not every model is appropriate for every situation.

If academic researchers could agree to follow such a quality standard or code of conduct, the many weaknesses of the models could be reduced, and eventually, Ray Ball (2013, p. 848) would no longer have to "fulminate against the earnings management literature" anymore.

REFERENCES

Amar, M., Chen, E., Gavious, I., & Weihs, H. (2022). Financial reporting misconduct: Evidence from the field. *Finance Research Letters*, *47*, 102540. https://doi.org/10.1016/j.frl.2021.102540

Amazon.com. (2020). *Form 10-K 2020*. United States Securities Exchange Commission. http://d18rn0p25nwr6d.cloudfront.net/CIK-0001018724/336d8745-ea82-40a5-9acc-1a89df23d0f3.pdf

Anagnostopoulou, S. C., & Tsekrekos, A. E. (2015). Earnings management in firms seeking to be acquired. *British Accounting Review*, *47*(4), 351–375. https://doi.org/10.1016/j.bar.2014.07.001

Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: empirical comparisons. *International Journal of Forecasting*, *8*, 69–80. https://doi.org/10.1016/0169-2070(92)90008-W

Ashbaugh-Skaife, H., Collins, D. W., & Kinney, W. R. (2009). The effect of SOX internal control deficiencies on firm risk and cost of equity. *Journal of Accounting Research*, *47*(1), 1–43. https://doi.org/10.1111/j.l475-679X.2008.00315.x

Atanasov, V., & Black, B. (2016). Shock-based causal inference in corporate finance and accounting research. *Critical Finance Review*, *5*(2), 207–304. https://doi.org/10.1561/104.00000036

Atkins, P. S., & 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*(3), 367–417. https://ir.lawnet.fordham.edu/jcfl/vol13/iss3/1/

Ayers, B. C., Jiang, J., & Yeung, P. E. (2006). Discretionary accruals and earning management: An analysis of pseudo earnings targets. *Accounting Review*, *81*(3), 617–652. https://doi.org/10.2308/accr.2006.81.3.617

Baber, W. R., & Kang, S. H. (2002). The impact of split adjusting and rounding on analysts' forecast error calculations. *Accounting Horizons*, *16*(4), 277–289. https://doi.org/10.2308/acch.2002.16.4.277

Baber, W. R., Kang, S. H., & Li, Y. (2011). Modeling discretionary accrual reversal and the balance sheet as an earnings management constraint. *Accounting Review*, *86*(4), 1189–1212. https://doi.org/10.2308/accr-10037

Ball, R. (2013). Accounting informs investors and earnings management is rife: Two questionable beliefs. *Accounting Horizons*, *27*(4), 847–853. https://doi.org/10.2308/acch-10366

Ball, R. (2016). IFRS – 10 years later. *Accounting and Business Research*, *46*(5), 545–571. https://doi.org/10.1080/00014788.2016.1182710

Ball, R., & Nikolaev, V. V. (2020). *FASB was right: Earnings beat cash flows when predicting future cash flows*. https://doi.org/10.2139/ssrn.3689802

Ball, R., Robin, A., & Sadka, G. (2008). Is financial reporting shaped by equity markets or by debt markets? An international study of timeliness and conservatism. *Review of Accounting Studies*, *13*(2-3), 168–205. https://doi.org/10.1007/s11142-007-9064-x

Ball, R., & Shivakumar, L. (2005). Earnings quality in UK private firms: Comparative loss recognition timeliness. *Journal of Accounting and Economics*, *39*(1), 83–128. https://doi.org/10.1016/j.jacceco.2004.04.001

Ball, R., & Shivakumar, L. (2006). The role of accruals in asymmetrically timely gain and loss recognition. *Journal of Accounting Research*, *44*(2), 207–242. https://doi.org/10.1111/j.1475-679X.2006.00198.x

Banker, R. D., Byzalov, D., Fang, S., & Jin, B. (2020). Operating asymmetries and nonlinear spline correction in discretionary accrual models. *Review of Quantitative Finance and Accounting*, *54*(3), 803–850. https://doi.org/10.1007/s11156-019-00808-5

Banker, R. D., & Chen, L. (2006). Predicting earnings using a model based on cost variability and cost stickiness. *The Accounting Review*, *81*(2), 285–307. https://doi.org/10.2308/accr.2006.81.2.285

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach. *Journal of Accounting Research*, *58*(1), 199–235. https://doi.org/10.1111/1475-679X.12292

Barth, M. E., Clinch, G., & Israeli, D. (2016). What do accruals tell us about future cash flows? *Review of Accounting Studies*, *21*(3), 768–807. https://doi.org/10.1007/s11142-016-9360-4

Barth, M. E., Cram, D. P., & Nelson, K. K. (2001). Accruals and the prediction of future cash flows. *The Accounting Review*, *76*(1), 27–58. https://doi.org/10.2308/accr.2001.76.1.27

Barth, M. E., Elliott, J. A., & Finn, M. W. (1999). Market rewards associated with patterns of increasing earnings. *Journal of Accounting Research*, *37*(2), 387–413. https://doi.org/10.2307/2491414

Barth, M. E., Landsman, W. R., & Lang, M. H. (2008). International accounting standards and accounting quality. *Journal of Accounting Research*, *46*(3), 467–498. https://doi.org/10.1111/j.1475-679X.2008.00287.x

Bartov, E. (1993). The timing of asset sales and earnings manipulation. *The Accounting Review*, *68*(4), 840–855. https://www.jstor.org/stable/248507

Bartov, E., & Cohen, D. A. (2009). The "Numbers Game" in the pre- and post-Sarbanes-Oxley eras. *Journal of Accounting, Auditing & Finance*, *24*(4), 505–534. https://doi.org/10.1177/0148558X0902400401

Bartov, E., Givoly, D., & Hayn, C. (2002). The rewards to meeting or beating earnings expectations. *Journal of Accounting and Economics*, *33*(2), 173–204. https://doi.org/10.1016/S0165-4101(02)00045-9

Basu, S. (1997). The conservatism principle and the asymmetric timeliness of earnings. *Journal of Accounting and Economics*, *24*, 3–37. https://doi.org/10.1111/j.1911-3846.2011.01151.x

Beaver, W. H., McNichols, M. F., & Nelson, K. K. (2007). An alternative interpretation of the discontinuity in earnings distributions. *Review of Accounting Studies*, *12*(4), 525–556. https://doi.org/10.1007/s11142-007-9053-0

Becker, C. L., DeFond, M. L., Jiambalvo, J., & Subramanyam, K. (1998). The effect of audit quality on earnings management. *Contemporary Accounting Research*, *15*(1), 1–24. https://doi.org/10.1111/j.1911-3846.1998.tb00547.x

Benabou, R., & Laroque, G. (1992). Using privileged information to manipulate markets: Insiders, gurus, and credibility. *Quarterly Journal of Economics*, *107*(3), 921–958. https://doi.org/10.2307/2118369

Ben-Amar, W., & Missonier-Piera, F. (2008). Earnings management by friendly takeover targets. *International Journal of Managerial Finance*, *4*(3), 232–243. https://doi.org/10.1108/17439130810878811

Beneish, M. D. (2001). Earnings management: A perspective. *Managerial Finance*, *27*(12), 3–17. https://doi.org/10.1108/03074350110767411

Beneish, M. D., & Vargus, M. E. (2002). Insider trading, earnings quality and accrual mispricing. *Accounting Review*, *77*(4), 755–791. https://doi.org/10.2308/accr.2002.77.4.755

Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*, *50*(2-3), 296–343. https://doi.org/10.1016/j.jacceco.2010.10.003

Bohrnstedt, G. W., & Carter, M. T. (1971). Robustness in regression analysis. *Sociological Methodology*, *3*, 118–146. https://doi.org/10.2307/270820

Brown, L. D., & Caylor, M. L. (2005). A temporal analysis of quarterly earnings thresholds: Propensities and valuation consequences. *The Accounting Review*, *80*(2), 423–440. https://doi.org/10.2308/accr.2005.80.2.423

Brown, L. D., & Higgins, H. N. (2001). Managing earnings surprises in the US versus 12 other countries. *Journal of Accounting and Public Policy*, *20*(4-5), 373–398. https://doi.org/10.1016/S0278-4254(01)00039-4

Burgstahler, D. C., & Chuk, E. (2015). Do scaling and selection explain earnings discontinuities? *Journal of Accounting and Economics*, *60*(1), 168–186. https://doi.org/10.1016/j.jacceco.2014.08.002

Burgstahler, D. C., & Chuk, E. (2017). What have we learned about earnings management? Integrating discontinuity evidence. *Contemporary Accounting Research*, *34*(2), 726–749. https://doi.org/10.1111/1911-3846.12301

Burgstahler, D. C., & Dichev, I. (1997). Earnings managment to avoid earnings decreases and losses. *Journal of Accounting and Economic*, *24*(1), 99–126. https://doi.org/10.1016/S0165-4101(97)00017-7

Burgstahler, D. C., & Eames, M. J. (2006). Management of earnings and analysts' forecasts to achieve zero and small positive earnings surprises. *Journal of Business Finance and Accounting*, *33*(5-6), 633–652. https://doi.org/10.1111/j.1468-5957.2006.00630.x

Burgstahler, D. C., Hail, L., & Leuz, C. (2006). Importance of reporting incentives: Earnings management in European firms. *The Accounting Review*, *81*(5), 983–1016. https://doi.org/10.2308/accr.2006.81.5.983

Callao, S., & Jarne, J. I. (2010). Have IFRS affected earnings management in the European Union? *Accounting in Europe*, *7*(2), 159–189. https://doi.org/10.1080/17449480.2010.511896

Callen, J. L., Robb, S. W., & Segal, D. (2008). Revenue manipulation and restatements by loss firms. *Auditing: A Journal of Practice & Theory*, *27*(2), 1–29. https://doi.org/10.2308/aud.2008.27.2.1

Cameron, A. B., & Stephens, L. (1991). The treatment of non-recurring items in the income statement and their consistency with FASB concept statements. *Abacus*, *27*(2), 81–96. https://doi.org/10.1111/j.1467-6281.1991.tb00260.x

Capkun, V., Collins, D., & Jeanjean, T. (2016). The effect of IAS/IFRS adoption on earnings management (smoothing): A closer look at competing explanations. *Journal of Accounting and Public Policy*, *35*(4), 352–394. https://doi.org/10.1016/j.jaccpubpol.2016.04.002

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Detecting management fraud in public companies. *Management Science*, *56*(7), 1146–1160. https://doi.org/10.1287/mnsc.1100.1174

Chardonnens, P., Fiechter, P., & Wallmeier, M. (2022). The disappearance of the zero-earnings discontinuity: SOX, dotcom boom or gradual decline? *Finance Research Letters*, *49*, 103033. https://doi.org/10.1016/j.frl.2022.103033

Chardonnens, P., & Wallmeier, M. (2022). *Kinked accounting? Small loss avoidance in Europe and (not) the US*. https://doi.org/10.2139/ssrn.3882901

Chen, C. C. S., Chou, Y. Y., & Wei, P. (2020). Country factors in earnings management of ADR firms. *Finance Research Letters*, *32*, 101146. https://doi.org/10.1016/j.frl.2019.04.003

Chen, F. H., Chi, D. J., & Wang, Y. C. (2015). Detecting biotechnology industry's earnings management using Bayesian network, principal component analysis, back propagation neural network, and decision tree. *Economic Modelling*, *46*, 1–10. https://doi.org/10.1016/j.econmod.2014.12.035

Chen, F. H., & Howard, H. (2016). An alternative model for the analysis of detecting electronic industries earnings management using stepwise regression, random forest, and decision tree. *Soft Computing*, *20*(5), 1945–1960. https://doi.org/10.1007/s00500-015-1616-6

Chen, H., Tang, Q., Jiang, Y., & Lin, Z. (2010). The role of international financial reporting standards in accounting quality: Evidence from the European Union. *Journal of International Financial Management and Accounting*, *21*(3), 220–278. https://doi.org/10.1111/j.1467-646X.2010.01041.x

Chen, S. K., Lin, B. X., Wang, Y., & Wu, L. (2010). The frequency and magnitude of earnings management: Time-series and multi-threshold comparisons. *International Review of Economics and Finance*, *19*(4), 671–685. https://doi.org/10.1016/j.iref.2010.02.005

Chen, S., Thomas, J., & Zhang, F. (2016). Spring-loading future performance when no one is looking? Earnings and cash flow management around acquisitions. *Review of Accounting Studies*, *21*(4), 1081–1115. https://doi.org/10.1007/s11142-016-9365-z

Chen, W., Hribar, P., & Melessa, S. (2018). Incorrect inferences when using residuals as dependent variables. *Journal of Accounting Research*, *56*(3), 751–796. https://doi.org/10.1111/1475-679X.12195

Cheng, Q., & Warfield, T. D. (2005). Equity incentives and earnings management. *The Accounting Review*, *80*(2), 441–476. https://doi.org/10.2308/accr.2005.80.2.441

Chung, H., Sheu, H. J., & Wang, J. L. (2009). Do firms' earnings management practices affect their equity liquidity? *Finance Research Letters*, *6*(3), 152–158. https://doi.org/10.1016/j.frl.2009.03.003

Coates IV, J. C. (2007). The goals and promise of the Sarbanes-Oxley Act. *Journal of Economic Perspectives*, *21*(1), 91–116. https://doi.org/10.1257/jep.21.1.91

Cohen, D. A., Dey, A., & Lys, T. Z. (2008). Real and accrual-based earnings management in the pre- and post-Sarbanes-Oxley periods. *The Accounting Review*, *83*(3), 757–788. https://doi.org/10.2308/accr.2008.83.3.757

Cohen, D. A., & Zarowin, P. (2010). Accrual-based and real earnings management activities around seasoned equity offerings. *Journal of Accounting and Economics*, *50*(1), 2–19. https://doi.org/10.1016/j.jacceco.2010.01.002

Dai, L., Chen, K., Sun, Z., Liu, Z., & Li, G. (2018). Broken adaptive ridge regression and its asymptotic properties. *Journal of Multivariate Analysis*, *168*, 334–351. https://doi.org/10.1016/j.jmva.2018.08.007

Daske, H., Gebhardt, G., & McLeay, S. (2006). The distribution of earnings relative to targets in the European Union. *Accounting and Business Research*, *36*(3), 137–167. https://doi.org/10.1080/00014788.2006.9730019

DeAngelo, L. E. (1986). Accounting numbers as market valuation substitutes: A study of management buyouts of public stockholders. *The Accounting Review*, *61*(3), 400–420. https://www.jstor.org/stable/247149

Dechow, P. M., & Dichev, I. D. (2002). The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review*, *77 (s-1)*, 35–59. https://doi.org/10.2308/accr.2002.77.s-1.35

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting material accounting misstatements. *Contemporary Accounting Research*, *28*(1), 17–82. https://doi.org/10.1111/j.1911-3846.2010.01041.x

Dechow, P. M., Ge, W., & Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, *50*(2-3), 344–401. https://doi.org/10.1016/j.jacceco.2010.09.001

Dechow, P. M., Hutton, A. P., Kim, J. H., & Sloan, R. G. (2012). Detecting earnings management: A new approach. *Journal of Accounting Research*, *50*(2), 275–334. https://doi.org/10.1111/j.1475-679X.2012.00449.x

Dechow, P. M., Hutton, A. P., & Sloan, R. G. (1995). Detecting earnings management. *Journal of Accounting Research*, *50*(2), 275–334. https://doi.org/10.1111/j.1475-679X.2012.00449.x

Dechow, P. M., Richardson, S. A., & Tuna, I. (2003). Why are earnings kinky? An examination of the earnings management explanation. *Review of Accounting Studies*, *8*(2/3), 355–384. https://doi.org/10.1023/A:1024481916719

DeFond, M. L., & Jiambalvo, J. (1994). Debt covenant violation and manipulation of accruals. *Journal of Accounting and Economics*, *17*(1-2), 145–176. https://doi.org/10.1016/0165-4101(94)90008-6

DeFond, M. L., & Park, C. W. (1997). Smoothing income in anticipation of future earnings. *Journal of Accounting and Economics*, *23*(2), 115–139. https://doi.org/10.1016/S0165-4101(97)00004-9

DeFond, M. L., & Park, C. W. (2001). The reversal of abnormal accruals and the market valuation of earnings surprises. *Accounting Review*, *76*(3), 375–404. https://doi.org/10.2308/accr.2001.76.3.375

Degeorge, F., Patel, J., & Zeckhauser, R. (1999). Earnings management to exceed thresholds. *The Journal of Business*, *72*(1), 1–33. https://doi.org/10.1086/209601

Desai, H., Rajgopal, S., & Venkatachalam, M. (2004). Value-glamour and accruals mispricing: One anomaly or two? *Accounting Review*, *79*(2), 355–385. https://doi.org/10.2308/accr.2004.79.2.355

Devore, J. L., & Berk, K. N. (2012). *Modern mathematical statistics with applications* (2nd ed.). Springer. https://doi.org/10.1007/978-1-4614-0391-3

Dhaliwal, D. S., Gleason, C. A., & Mills, L. F. (2004). Last chance earnings management: Using the tax expense to meet analysts' forecasts. *Contemporary Accounting Research*, *21*(2), 431–459. https://doi.org/10.1506/TFVV-UYT1-NNYT-1YFH

Dichev, I. D., Graham, J. R., Harvey, C. R., & Rajgopal, S. (2013). Earnings quality: Evidence from the field. *Journal of Accounting and Economics*, *56*(2-3), 1–33. https://doi.org/10.1016/j.jacceco.2013.05.004

Donelson, D. C., Mcinnis, J. M., & Mergenthaler, R. D. (2013). Discontinuities and earnings management: Evidence from restatements related to securities litigation. *Contemporary Accounting Research*, *30*(1), 242–268. https://doi.org/10.1111/j.1911-3846.2012.01150.x

Doukakis, L. C. (2014). The effect of mandatory IFRS adoption on real and accrual-based earnings management activities. *Journal of Accounting and Public Policy*, *33*(6), 551–572. https://doi.org/10.1016/j.jaccpubpol.2014.08.006

DuCharme, L. L., Malatesta, P. H., & Sefcik, S. E. (2001). Earnings management: IPO Valuation and subsequent performance. *Journal of Accounting, Auditing & Finance*, *16*(4), 369–396. https://doi.org/10.1177/0148558X0101600409

Durtschi, C., & Easton, P. (2005). Earnings management? The shapes of the frequency distributions of earnings metrics are not evidence ipso facto. *Journal of Accounting Research*, *43*(4), 557–592. https://doi.org/10.1111/j.1475-679X.2005.00182.x

Durtschi, C., & Easton, P. (2009). Earnings management? Erroneous inferences based on earnings frequency distributions. *Journal of Accounting Research*, *47*(5), 1249–1281. https://doi.org/10.1111/j.1475-679X.2009.00347.x

Dyck, A., Morse, A., & Zingales, L. (2010). Who blows the whistle on corporate fraud? *Journal of Finance*, *65*(6), 2213–2253. https://doi.org/10.1111/j.1540-6261.2010.01614.x

Easterwood, C. M. (1998). Takeovers and incentives for earnings management: An empirical analysis. https://doi.org/10.19030/jabr.v14i1.5726

Ecker, F., Francis, J., Olsson, P., & Schipper, K. (2013). Estimation sample selection for discretionary accruals models. *Journal of Accounting and Economics*, *56*(2-3), 190–211. https://doi.org/10.1016/j.jacceco.2013.07.001

Elliott, J. A., & Shaw, W. H. (1988). Write-offs as accounting procedures to manage perceptions. *Journal of Accounting Research*, *26*, 91. https://doi.org/10.2307/2491182

Engel, E., Hayes, R. M., & Wang, X. (2007). The Sarbanes-Oxley Act and firms ' going-private decisions. *Journal of Accounting and Economics*, *44*, 116–145. https://doi.org/10.1016/j.jacceco.2006.07.002

Enomoto, M., & Yamaguchi, T. (2017). Discontinuities in earnings and earnings change distributions after J-SOX implementation: Empirical evidence from Japan. *Journal of Accounting and Public Policy*, *36*(1), 82–98. https://doi.org/10.1016/j.jaccpubpol.2016.11.005

Erickson, M., & Wang, S. W. (1999). Earnings management by acquiring firms in stock for stock mergers. *Journal of Accounting and Economics*, *27*(2), 149–176. https://doi.org/10.1016/S0165-4101(99)00008-7

Ewert, R., & Wagenhofer, A. (2005). Economic effects of tightening accounting standards to restrict earnings management. *The Accounting Review*, *80*(4), 1101–1124. https://doi.org/10.2308/accr.2005.80.4.1101

Ewert, R., & Wagenhofer, A. (2012). Earnings management, conservatism, and earnings quality. *Foundations and Trends in Accounting*, *6*(2), 65–186. https://doi.org/10.1561/1400000025

Fama, E. F., & French, K. R. (2008). Dissecting anomalies. *Journal of Finance*, *63*(4), 1653–1678. https://doi.org/10.1111/j.1540-6261.2008.01371.x

Fan, Y., Barua, A., Cready, W. M., & Thomas, W. B. (2010). Managing earnings using classification shifting: Evidence from quarterly special items. *Accounting Review*, *85*(4), 1303–1323. https://doi.org/10.2308/accr.2010.85.4.1303

Fan, Y., & Liu, X. K. (2017). Misclassifying core expenses as special items: Cost of goods sold or selling, general, and administrative expenses. *Contemporary Accounting Research*, *34*(1), 400–426. https://doi.org/10.1111/1911-3846.12234

Financial Accounting Standards Board. (1985). Statement of financial accounting concepts no. 6. *Elements of Financial Statements*. https://www.fasb.org/Page/ShowPdf?path=aop%7B%5C_%7DCON6.pdf%7B%5C&%7Dtitle=CON+6+%7B%5C%%7D28AS+AMENDED%7B%5C%%7D29%7B%5C&%7DacceptedDisclaimer=true%7B%5C&%7DSubmit=

Fudenberg, D., & Tirole, J. (1995). A theory of income and dividend smoothing based on incumbency rents. *Journal of Political Economy*, *103*(1), 75–93. https://www.jstor.org/stable/2138719

Gaio, C. (2010). The relative importance of firm and country characteristics for earnings quality around the world. *European Accounting Review*, *19*(4), 693–738. https://doi.org/10.1080/09638180903384643

Gassen, J., Uwe Fülbier, R., & Sellhorn, T. (2006). International differences in conditional conservatism – The role of unconditional conservatism and income smoothing. *European Accounting Review*, *15*(4), 527–564. https://doi.org/10.1080/09638180601102107

Gilliam, T. A., Heflin, F., & Paterson, J. S. (2015). Evidence that the zero-earnings discontinuity has disappeared. *Journal of Accounting and Economics*, *60*(1), 117–132. https://doi.org/10.1016/j.jacceco.2014.07.001

Glaum, M., Lichtblau, K., & Lindemann, J. (2004). The extent of earnings management in the U.S. and Germany. *Journal of International Accounting Research*, *3*(2), 45–77. https://doi.org/10.2308/jiar.2004.3.2.45

Gore, P., Pope, P. F., & Singh, A. K. (2007). Earnings management and the distribution of earnings relative to targets: UK evidence. *Accounting and Business Research*, *37*(2), 123–149. https://doi.org/10.1080/00014788.2007.9730065

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics*, *40*(1-3), 3–73. https://doi.org/10.1016/j.jacceco.2005.01.002

Greenwell, B. M., & Boehmke, B. C. (2020). Variable importance plots - An introduction to the vip package. *R Journal*, *12*(1), 343–366. https://doi.org/10.32614/rj-2020-013

Guan, L., Wright, C. J., & Sun, L. (2004). Earnings management in targeted hostile takeover firms. *Journal of Forensic Accounting*, *5*(2), 421–432. https://doi.org/10.1108/17439130810878811

Guan, L., Wright, C. J., & Leikam, S. L. (2005). Earnings management and forced CEO dismissal. *Advances in Accounting*, *21*(5), 61–81. https://doi.org/10.1016/S0882-6110(05)21003-9

Guay, W. R. (2008). Conservative financial reporting, debt covenants, and the agency costs of debt. *Journal of Accounting and Economics*, *45*(2-3), 175–180. https://doi.org/10.1016/j.jacceco.2008.05.001

Haga, J., Huhtamaki, F., & Sundvik, D. (2017). *Myopia and earnings management strategies*. https://doi.org/hdl.handle.net/10125/51942

Haga, J., Huhtamäki, F., & Sundvik, D. (2019). Long-term orientation and earnings management strategies. *Journal of International Accounting Research*, *18*(3), 97–119. https://doi.org/doi.org/10.2308/jiar-52501

Haga, J., Siekkinen, J., & Sundvik, D. (2015). A neural network approach to measure real activities manipulation. *Expert Systems with Applications*, *42*(5), 2313–2322. https://doi.org/10.1016/j.eswa.2014.10.047

Han, S., Kang, T., Salter, S., & Yoo, Y. K. (2010). A cross-country study on the effects of national culture on earnings management. *Journal of International Business Studies*, *41*(1), 123–141. https://doi.org/10.1057/jibs.2008.78

Hansen, G. A. (1999). Bias and measurement error in discretionary accrual models. *SSRN Electronic Journal*, 1–38. https://doi.org/10.2139/ssrn.192748

Hansen, J. C. (2010). The effect of alternative goals on earnings management studies: An earnings benchmark examination. *Journal of Accounting and Public Policy*, *29*(5), 459–480. https://doi.org/10.1016/j.jaccpubpol.2010.06.002

Hastie, T., Tibshirani, R., & Tibshirani, R. J. (2017). Extended comparisons of best subset selection, forward stepwise selection, and the lasso. *arXiv preprint*, 1–53. https://doi.org/10.48550/arXiv.1707.08692

Haw, I. M., Hu, B., Hwang, L. S., & Wu, W. (2004). Ultimate ownership, income management, and legal and extra-legal institutions. *Journal of Accounting Research*, *42*(2), 423–462. https://doi.org/10.1111/j.1475-679X.2004.00144.x

Hayn, C. (1995). The information content of losses. *Journal of Accounting and Economics*, *20*(2), 125–153. https://doi.org/10.1016/0165-4101(95)00397-2

Healy, P. M., & Wahlen, J. M. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting Horizons*, *13*(4), 365–383. https://doi.org/10.2308/acch.1999.13.4.365

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind* (3rd ed.). McGraw-Hill, New York.

Höglund, H. (2012). Detecting earnings management with neural networks. *Expert Systems with Applications*, *39*(10), 9564–9570. https://doi.org/10.1016/j.eswa.2012.02.096

Höglund, H. (2013). Fuzzy linear regression-based detection of earnings management. *Expert Systems with Applications*, *40*(15), 6166–6172. https://doi.org/10.1016/j.eswa.2013.05.046

Höglund, H. (2015). Uniform accrual generating process grouping with self-organizing maps. *Expert Systems with Applications*, *42*(1), 554–561. https://doi.org/10.1016/j.eswa.2014.08.010

Holst, E., & Thyregod, P. (1999). A statistical test for the mean squared error. *Journal of Statistical Computation and Simulation*, *63*(4), 321–347. https://doi.org/10.1080/00949659908811960

Ibrahim, S. S. (2009). The usefulness of measures of consistency of discretionary components of accruals in the detection of earnings management. *Journal of Business Finance and Accounting*, *36*(9-10), 1087–1116. https://doi.org/10.1111/j.1468-5957.2009.02171.x

Iliev, P. (2010). The effect of SOX section 404: Costs, earnings quality, and stock prices. *Journal of Finance*, *65*(3), 1163–1196. https://doi.org/10.1111/j.1540-6261.2010.01564.x

Ipino, E., & Parbonetti, A. (2017). Mandatory IFRS adoption: The trade-off between accrual-based and real earnings management. *Accounting and Business Research*, *47*(1), 91–121. https://doi.org/10.1080/00014788.2016.1238293

Jackson, A. B. (2018). Discretionary accruals: Earnings management.. or not? *Abacus*, *54*(2), 136–153. https://doi.org/10.1111/abac.12117

Jacob, J., & Jorgensen, B. N. (2007). Earnings management and accounting income aggregation. *Journal of Accounting and Economics*, *43*(2-3), 369–390. https://doi.org/10.1016/j.jacceco.2007.01.007

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning* (2nd ed.). Springer Texts. https://hastie.su.domains/ISLR2/ISLRv2%7B%5C_%7Dwebsite.pdf

Jeanjean, T., & Stolowy, H. (2008). Do accounting standards matter? An exploratory analysis of earnings management before and after IFRS adoption. *Journal of Accounting and Public Policy*, *27*(6), 480–494. https://doi.org/10.1016/j.jaccpubpol.2008.09.008

Jeter, D. C., & Shivakumar, L. (1999). Cross-sectional estimation of abnormal accruals using quarterly and annual data: Effectiveness in detecting event-specific earnings management. *Accounting and Business Research*, *29*(4), 299–319. https://doi.org/10.1080/00014788.1999.9729590

Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of Accounting Research*, *29*(2), 193–228. https://doi.org/10.2307/2491047

Kamar, E., & Talley, E. (2008). Going-private decisions and the Sarbanes-Oxley Act of 2002: A cross-country analysis. *Journal of Law, Economics, & Organization*, *25*(1), 107–132. https://doi.org/10.1093/jleo/ewn019

Kaplan, R. S. (1985). Comments on Paul Healy evidence on the effect of bonus schemes on accounting procedure and accrual decisions. *Journal of Accounting and Economics*, *7*, 109–113. https://doi.org/10.1016/0165-4101(85)90030-8

Karpoff, J. M., Koester, A., Lee, D. S., & Martin, G. S. (2017). Proxies and databases in financial misconduct research. *Accounting Review*, *92*(6), 129–163. https://doi.org/10.2308/accr-51766

Kerstein, J., & Rai, A. (2007). Intra-year shifts in the earnings distribution and their implications for earnings management. *Journal of Accounting and Economics*, *44*(3), 399–419. https://doi.org/10.1016/j.jacceco.2007.04.004

Keung, E., & Shih, M. S. (2014). Measuring discretionary accruals: Are ROA-matched models better than the original Jones-type models? *Review of Accounting Studies*, *19*(2), 736–768. https://doi.org/10.1007/s11142-013-9262-7

Kirschenheiter, M., & Melumad, N. D. (2002). Can "big bath" and earnings smoothing coexist as equilibrium financial reporting strategies? *Journal of Accounting Research*, *40*(3), 761–796. https://doi.org/10.1111/1475-679X.00070

Kothari, S. P., Leone, A. J., & Wasley, C. E. (2005). Performance matched discretionary accrual measures. *Journal of Accounting and Economics*, *39*(1), 163–197. https://doi.org/10.1016/j.jacceco.2004.11.002

Kothari, S. P., Mizik, N., & Roychowdhury, S. (2016). Managing for the moment: The role of earnings management via real activities versus accruals in SEO valuation. *The Accounting Review*, *91*(2), 559–586. https://doi.org/10.2308/accr-51153

Kuhn, M. (2019). The caret Package. Retrieved July 11, 2022, from https://topepo.github.io/caret/

Lahr, H. (2014). An improved test for earnings management using kernel density estimation. *European Accounting Review*, *23*(4), 559–591. https://doi.org/10.1080/09638180.2013.860044

Lambert, R. A. (1984). Income smoothing as rational equilibrium behavior? *Accounting Review*, *59*(4), 604–618. https://doi.org/10.2308/TAR-2019-0671

Larson, C. R., Sloan, R., & Zha Giedt, J. (2018). Defining, measuring, and modeling accruals: A guide for researchers. *Review of Accounting Studies*, *23*(3), 827–871. https://doi.org/10.1007/s11142-018-9457-z

Leuz, C., Nanda, D., & Wysocki, P. D. (2003). Earnings management and investor protection: An international comparison. *Journal of Financial Economics*, *69*(3), 505–527. https://doi.org/10.1016/S0304-405X(03)00121-1

Leuz, C., Triantis, A., & Yue Wang, T. (2008). Why do firms go dark? Causes and economic consequences of voluntary SEC deregistrations. *Journal of Accounting and Economics*, *45*(2-3), 181–208. https://doi.org/10.1016/j.jacceco.2008.01.001

Lev, B., Li, S., & Sougiannis, T. (2010). The usefulness of accounting estimates for predicting cash flows and earnings. *Review of Accounting Studies*, *15*(4), 779–807. https://doi.org/10.1007/s11142-009-9107-6

Lev, B. I., & Nissim, D. (2006). The persistence of the accruals anomaly. *Contemporary Accounting Research*, *23*(1), 193–226. https://doi.org/10.2139/ssrn.546108

Levitt, A. (1998). The numbers game. *The CPA Journal*, *68*(12), 14.

Li, L. (2019). Is there a trade-off between accrual-based and real earnings management? Evidence from equity compensation and market pricing. *Finance Research Letters*, *28*, 191–197. https://doi.org/10.1016/j.frl.2018.04.021

Lin, S., Riccardi, W., & Wang, C. (2012). Does accounting quality change following a switch from U.S. GAAP to IFRS? Evidence from Germany. *Journal of Accounting and Public Policy*, *31*(6), 641–657. https://doi.org/10.1016/j.jaccpubpol.2012.10.006

Liu, Y., Ning, Y., & Davidson, W. N. (2010). Earnings management surrounding new debt issues. *Financial Review*, *45*(3), 659–681. https://doi.org/10.1111/j.1540-6288.2010.00265.x

Liu, Z., Sun, F., & McGovern, D. P. (2017). Sparse generalized linear model with L0 approximation for feature selection and prediction with big omics data. *BioData Mining*, *10*(1), 1–12. https://doi.org/10.1186/s13040-017-0159-z

Lobo, G. J., & Zhou, J. (2006). Did conservatism in financial reporting increase after the Sarbanes-Oxley Act? Initial evidence. *Accounting Horizons*, *20*(1), 57–73. https://doi.org/10.2308/acch.2006.20.1.57

Lobo, G. J., & Zhou, J. (2010). Changes in discretionary financial reporting behavior following the Sarbanes-Oxley Act. *Journal of Accounting, Auditing & Finance*, *25*(1), 1–26. https://doi.org/10.1177/0148558X1002500101

Louis, H. (2004). Earnings management and the market performance of acquiring firms. *Journal of Financial Economics*, *74*(1), 121–148. https://doi.org/10.1016/j.jfineco.2003.08.004

Makarem, N., Hussainey, K., & Zalata, A. (2018). Earnings management in the aftermath of the zero-earnings discontinuity disappearance. *Journal of Applied Accounting Research*, *19*(3), 401–422. https://doi.org/10.1108/JAAR-03-2017-0047

McMullin, J. L., & Schonberger, B. (2020). Entropy-balanced accruals. *Review of Accounting Studies*, *25*(1), 84–119. https://doi.org/10.1007/s11142-019-09525-9

McNichols, M. F. (2002). The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review*, *77(s-1)*, 35–59. https://doi.org/10.2308/accr.2002.77.s-1.35

McNichols, M. F., & Stubben, S. R. (2018). Research design issues in studies using discretionary accruals. *Abacus*, *54*(2), 227–246. https://doi.org/10.1111/abac.12128

McVay, S. E. (2006). Earnings management using classification shifting: An examination of core earnings and special items. *Accounting Review*, *81*(3), 501–531. https://doi.org/10.2308/accr.2006.81.3.501

Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics and Data Analysis*, *52*(1), 374–393. https://doi.org/10.1016/j.csda.2006.12.019

Mindak, M. P., Sen, P. K., & Stephan, J. (2016). Beating threshold targets with earnings management. *Review of Accounting and Finance*, *15*(2), 198–221. https://doi.org/10.1108/RAF-04-2015-0057

Missonier-Piera, F., & Spadetti, C. (2022). The consequences of earnings management for the acquisition premium in friendly takeovers. *Journal of Business Finance & Accounting*, 1–27. https://doi.org/10.1111/jbfa.12635

Moez, A. (2020). PyCaret: An open source, low-code machine learning library in Python. Retrieved July 11, 2022, from https://pycaret.gitbook.io/docs/

Myers, J. N., Myers, L. A., & Omer, T. C. (2003). Exploring the term of the auditor-client relationship and the quality of earnings: A case for mandatory auditor rotation? *The Accounting Review*, *78*(3), 779–799. https://doi.org/10.2308/accr.2003.78.3.779

Myers, J. N., Myers, L. A., & Skinner, D. J. (2007). Earnings momentum and earnings management. *Journal of Accounting, Auditing & Finance*, *22*(2), 249–284. https://doi.org/10.1177/0148558X0702200211

Myers, S. C., & Majluf, N. S. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, *13*(2), 187–221. https://doi.org/10.1016/0304-405X(84)90023-0

Nabar, S., & Boonlert-U-Thai, K. K. (2007). Earnings management, investor protection, and national culture. *Journal of International Accounting Research*, *6*(2), 35–54. https://doi.org/10.2308/jiar.2007.6.2.35

Namazi, M., & Maharluie, M. S. (2015). Detecting earnings management via statistical and neural network techniques. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, *9*(7), 2490–2498. https://doi.org/doi.org/10.5281/zenodo.1109241

Owens, E. L., Wu, J. S., & Zimmerman, J. (2017). Idiosyncratic shocks to firm underlying economics and abnormal accruals. *Accounting Review*, *92*(2), 183–219. https://doi.org/10.2308/accr-51523

Payne, J. L., & Robb, S. W. G. (2000). Earnings management: The effect of ex ante earnings expectations. *Journal of Accounting, Auditing & Finance*, *15*(4), 371–392. https://doi.org/10.1177/0148558X0001500401

Payne, J. L., & Thomas, W. B. (2003). The implications of using stock-split adjusted I/B/E/S data in empirical research. *Accounting Review*, *78*(4), 1049–1067. https://doi.org/10.2308/accr.2003.78.4.1049

Pereira, A., & Alves, M. d. C. G. (2017). Earnings management and European Regulation 1606/2002: Evidence from non-financial Portuguese companies listed in Euronext. *Revista de Contabilidad*, *20*(2), 107–117. https://doi.org/10.1016/j.rcsar.2017.05.002

Perry, S. E., & Williams, T. H. (1994). Earnings management preceding management buyout offers. *Journal of Accounting and Economics*, *18*(2), 157–179. https://doi.org/10.1016/0165-4101(94)00362-9

Pourciau, S. (1993). Earnings management and nonroutine executive changes. *Journal of Accounting and Economics*, *16*, 317–336. https://doi.org/10.1016/0165-4101(93)90015-8

Rakoff, J. S. (2014). The financial crisis: Why have no high-level executives been prosecuted? In *Enforcement of corporate and securities law: China and the world* (pp. 3–13). Cambridge University Press. https://doi.org/10.1017/9781316691397.002

Rangan, S. (1998). Earnings management and the performance of seasoned equity offerings. *Journal of Financial Economics*, *50*(1), 101–122. https://doi.org/10.1016/s0304-405x(98)00033-6

Refinitiv. (2022). Refinitiv Datastream. Retrieved September 6, 2022, from https://www.refinitiv.com/en/financial-data/company-data/fundamentals-data/worldscope-fundamentals

Richardson, S. A., & Sloan, R. G. (2005). Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics*, *39*(3), 437–485. https://doi.org/10.1016/j.jacceco.2019.101235

Richardson, S. A., Teoh, S. H., & Wysocki, P. D. (1999). *Tracking analysts' forecasts over the annual earnings horizon: Are analysts' forecasts optimistic or pessimistic?* https://doi.org/10.2139/ssrn.168191

Rigby, R. A., Stasinopoulos, D. M., & Lane, P. W. (2005). Generalized additive models for location, scale and shape. *Applied Statistics*, *54*(3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Ronen, J., & Yaari, V. (2008). *Earnings Management*. Springer.

Roychowdhury, S. (2006). Earnings management through real activities manipulation. *Journal of Accounting and Economics*, *42*(3), 335–370. https://doi.org/10.1016/j.jacceco.2006.01.002

Schipper, K. (1989). Commentary on earnings management. *Accounting Horizons*, *3*(4), 91–102.

Scott, D. W. (1992). *Multivariate densitiy estimation: Theory, practice and visualization.* https://doi.org/10.1002/9780470316849

Shivakumar, L. (2000). Do firms mislead investors by overstating earnings before seasoned equity offerings? *Journal of Accounting and Economics*, *29*, 339–371. https://doi.org/10.1016/S0165-4101(00)00026-4

Soderstrom, N. S., & Sun, K. J. (2007). IFRS adoption and accounting quality: A review. *European Accounting Review*, *16*(4), 675–702. https://doi.org/10.1080/09638180701706732

Strong, J. S., & Meyer, J. R. (1987). Asset writedowns: Managerial incentives and security returns. *The Journal of Finance*, *42*(3), 643–661. https://doi.org/10.2307/2328376

Stubben, S. R. (2010). Discretionary revenues as a measure of earnings management. *The Accounting Review*, *85*(2), 695–717. https://doi.org/10.2308/accr.2010.85.2.695

Subramanyam, K. R., & Venkatachalam, M. (2007). Earnings, cash flows, and ex post intrinsic value of equity. *The Accounting Review*, *82*(2), 457–481. https://doi.org/10.2308/accr.2007.82.2.457

Teoh, S. H., Welch, I., & Wong, T. J. (1998). Earnings management and the long-run market performance of initial public offerings. *Journal of Finance*, *53*(6), 1935–1974. https://doi.org/10.1111/0022-1082.00079

Teoh, S. H., Welch, I., & Wong, T. (1998). Earnings management and the underperformance of seasoned equity offerings. *Journal of Financial Economics*, *50*, 63–99. https://doi.org/10.1016/S0304-405X(98)00032-4

The White House. (2004). Corporate Accountabilty Reform. https://georgewbush-whitehouse.archives.gov/infocus/achievement/print/chap9.html

Theil, H. (1966). *Applied economic forecasting* (3rd ed.). North-Holland Publishing Company.

Thomson Financial. (2007). Worldscope Database Datatype Definitions Guide.

Thomson-Reuters. (2010). I/B/E/S on Datastream.

Thomson-Reuters. (2013). Worldscope database data definitions guide. https://blogs.cul.columbia.edu/business/files/2014/02/Worldscope-Data-Definition-Guide-Issue-14.2.pdf

Trimble, M. (2018). A reinvestigation into accounting quality following global IFRS adoption: Evidence via earnings distributions. *Journal of International Accounting, Auditing and Taxation*, *33*, 18–39. https://doi.org/10.1016/j.intaccaudtax.2018.09.001

Ulbricht, N., & Weiner, C. (2005). Worldscope meets Compustat: a comparison of financial databases. https://doi.org/10.2139/ssrn.871169

van der Loo, M., van der Laan, J., Team, R. C., Logan, N., Muir, C., & Gruber, J. (2020). stringdist: Approximate string matching, fuzzy text search, and string distance functions. *R package version 0.9.6.3*. https://doi.org/10.32614/RJ-2014-011 van Tendeloo, B., & Vanstraelen, A. (2005). Earnings management under German GAAP versus IFRS. *European Accounting Review*, *14*(1), 155–180. https://doi.org/10.1080/0963818042000338988

Wallmeier, M. (1997). *Prognose von Aktienrenditen und -risiken mit Mehrfaktorenmodellen*. Uhlenbruch Verlag.

Wallmeier, M. (2021). *Cross-sectional relation of accruals, cash flow and profitability: A bayesian model for analyzing earnings management*. https://doi.org/http://dx.doi.org/10.2139/ssrn.3882906

Watts, R. L. (2003). Conservatism in accounting part I: Explanations and implications. *Accounting Horizons*, *17*(3), 207–221. https://doi.org/10.2308/acch.2003.17.3.207

Watts, R. L., & Zimmerman, J. L. (1978). Towards a positive theory of the determination of accounting standards. *The Accounting Review*, *53*(1), 112–134. http://www.jstor.org/stable/245729

Watts, R. L., & Zimmerman, J. L. (1990). Positive accounting theory: A ten year perspective. *The Accounting Review*, *65*(1), 131–156. https://doi.org/https://www.jstor.org/stable/247880

Wells, P. (2002). Earnings management surrounding CEO changes. *Accounting and Finance*, *42*(2), 169–193. https://doi.org/10.1111/1467-629X.00073

Wintoki, M. B. (2007). Corporate boards and regulation: The effect of the Sarbanes-Oxley Act and the exchange listing requirements on firm value. *Journal of Corporate Finance*, *13*, 229–250. https://doi.org/10.1016/j.jcorpfin.2007.03.001

Xue, Y. (2003). *Information content of earnings management: Evidence from managing earnings to exceed thresholds*. https://doi.org/10.2139/ssrn.582601

Ye, J. (2007). *Accounting accruals and tests of earnings management*, Baruch College. https://doi.org/10.2139/ssrn.1003101

Yu, D. (2014). Accounting standards and earnings management around the world. *Journal of International & Interdisciplinary Business Research*, *1*(5), 43–69. https://doi.org/scholars.fhsu.edu/jiibr/vol1/iss1/5

Zang, A. Y. (2012). Evidence on the trade-off between real activities manipulation and accrual-based earnings management. *Accounting Review*, *87*(2), 675–703. https://doi.org/10.2308/accr-10196

Zéghal, D., Chtourou, S., & Sellami, Y. M. (2011). An analysis of the effect of mandatory adoption of IAS/IFRS on earnings management. *Journal of International Accounting, Auditing and Taxation*, *20*(2), 61–72. https://doi.org/10.1016/j.intaccaudtax.2011.06.001

Appendices

Appendix A
THE DISAPPEARANCE OF THE ZERO-EARNINGS DISCONTINUITY: SOX, DOTCOM BOOM OR GRADUAL DECLINE?

		Interval 0.005					Interval 0.015				
Year	N	SD_{-1}	SD_1	SLD	SPD	MPL	SD_{-1}	SD_1	SLD	SPD	MPL
1987	4,119	-4.19	1.93	-0.57	0.25	1.06	-7.19	3.63	-0.31	0.17	0.52
1988	4,223	-2.92	1.79	-0.43	0.29	0.82	-7.94	4.48	-0.34	0.20	0.60
1989	4,075	-4.26	2.81	-0.54	0.51	1.19	-7.13	3.85	-0.36	0.20	0.62
1990	4,032	-2.59	2.29	-0.38	0.22	0.67	-3.69	2.36	-0.19	0.11	0.31
1991	4,034	-5.61	3.57	-0.68	0.60	1.61	-11.67	5.68	-0.47	0.27	0.88
1992	4,144	-2.74	1.08	-0.42	0.19	0.71	-9.38	5.08	-0.34	0.19	0.58
1993	4,395	-4.97	2.77	-0.58	0.28	1.11	-7.65	2.68	-0.30	0.09	0.44
1994	4,792	-5.16	3.39	-0.55	0.32	1.08	-6.26	1.24	-0.26	0.07	0.36
1995	5,087	-4.58	3.24	-0.45	0.31	0.88	-5.26	2.60	-0.20	0.11	0.32
1996	5,562	-2.87	0.39	-0.35	-0.01	0.42	-4.30	-0.61	-0.16	-0.001	0.17
1997	5,932	-4.49	4.32	-0.39	0.45	0.86	-10.18	4.46	-0.33	0.13	0.52
1998	5,812	-2.85	1.74	-0.28	0.16	0.48	-6.25	2.38	-0.20	0.06	0.28
1999	5,624	-1.45	0.57	-0.25	0.02	0.31	-3.11	0	-0.11	0.01	0.14
2000	5,381	-0.70	2.15	-0.07	0.26	0.30	-4.38	2.77	-0.15	0.09	0.26
2001	5,218	-2.20	0.96	-0.32	0.03	0.41	-2.88	1.13	-0.13	0.06	0.20
2002	4,921	-3.18	1.04	-0.41	0.07	0.59	-5.73	2.98	-0.23	0.11	0.37
2003	4,605	-2.21	1.18	-0.19	0.09	0.30	0.79	-1.93	0.01	-0.06	-0.07
2004	4,434	-0.76	1.47	-0.09	0.10	0.20	-2.52	-0.68	-0.11	-0.02	0.09
2005	4,271	-0.97	1.37	-0.11	0.14	0.26	-2.78	-0.43	-0.11	0.002	0.12
2006	4,133	-0.23	0.55	-0.05	0.12	0.17	-1.91	-0.21	-0.09	-0.01	0.09
2007	4,052	-2.06	0.92	-0.21	0.004	0.25	-0.42	-3.49	-0.03	-0.10	-0.07
2008	3,929	-0.96	1.19	-0.14	0.11	0.25	-3.84	1.80	-0.16	0.04	0.22
2009	3,780	-0.45	-0.18	-0.10	-0.03	0.07	-1.49	1.55	-0.08	0.05	0.13
2010	3,615	-3.02	0.20	-0.41	-0.11	0.40	-1.78	0.04	-0.08	0.002	0.09
2011	3,523	-1.28	0.40	-0.15	0.05	0.22	-2.92	1.25	-0.12	0.05	0.17
2012	3,439	-1.07	0.54	-0.14	0.05	0.21	-1.46	0.49	-0.08	0.02	0.10
2013	3,434	-1.77	1.54	-0.21	0.11	0.34	-0.75	-1.33	-0.06	-0.03	0.03
2014	3,440	-0.36	-0.50	-0.14	-0.11	0.04	-0.59	-2.87	-0.02	-0.08	-0.06
2015	3,397	-2.00	0.23	-0.36	-0.06	0.39	-3.38	-0.78	-0.15	-0.04	0.13
2016	3,277	-1.01	1.21	-0.18	0.09	0.29	-1.38	1.00	-0.06	0.05	0.11
2017	3,274	-0.50	1.41	0.01	0.15	0.13	-0.87	-1.64	-0.05	-0.05	-0.01
2018	3,529	1.12	-0.27	0.05	-0.07	-0.12	-2.61	-2.57	-0.11	-0.07	0.04
2019	3,124	2.87	-1.46	0.29	-0.11	-0.37	-1.15	-1.57	-0.05	-0.06	-0.01

Table A.1: Discontinuity Measures in the Subsample of Firms With Sales Greater Than 2 Million USD. Earnings are scaled by the market value of equity. N is the number of observations, SD is the standardized differences t-statistic, SLD is the small loss deviation, SPD is the small profit deviation and MPL is the modified profit-to-loss ratio. The measures are defined in Section 3.2.

Appendix B

KINKED ACCOUNTING? LOSS AVOIDANCE IN EUROPE AND (NOT) THE US

Year	US					Europe				
	N	SLD	SPD	$t(SD_{-1})$	$t(SD_1)$	N	SLD	SPD	$t(SD_{-1})$	$t(SD_1)$
1988	1,428	-0.09	0.03	-1.36	0.37	963	-0.41	0.08	-4.63	1.35
1989	1,455	-0.31	0.05	-4.08	0.63	1,403	-0.25	0.01	-3.32	0.65
1990	1,345	-0.38	0.20	-5.72	3.21	1,552	-0.39	0.10	-7.72	2.16
1991	1,423	-0.36	0.26	-6.52	4.57	1,608	-0.25	0.13	-5.02	2.75
1992	1,586	-0.34	0.17	-6.15	3.17	1,550	-0.28	0.22	-6.37	5.00
1993	1,732	-0.32	0.13	-6.14	2.46	1,671	-0.42	0.22	-10.83	5.38
1994	1,818	-0.36	0.03	-5.65	0.68	1,829	-0.40	0.22	-10.49	5.56
1995	2,293	-0.25	-0.003	-3.89	-0.29	1,330	-0.31	0.20	-5.12	3.13
1996	2,570	-0.03	0.07	-0.43	1.36	1,466	-0.20	0.18	-3.67	3.19
1997	2,832	-0.19	0.06	-3.10	1.19	1,610	-0.39	0.06	-6.74	1.48
1998	2,780	-0.18	0.04	-3.18	1.17	1,695	-0.40	0.12	-6.49	2.15
1999	2,654	-0.14	0.05	-2.48	0.98	1,704	-0.37	0.20	-7.29	3.69
2000	2,363	-0.17	0.08	-2.62	1.24	1,794	-0.28	0.15	-5.45	2.81
2001	2,460	-0.07	0.09	-1.50	1.57	1,849	-0.17	0.17	-3.77	3.76
2002	2,320	-0.13	0.08	-1.99	0.96	1,699	-0.22	0.18	-4.93	3.95
2003	2,414	-0.08	-0.03	-1.15	-0.73	1,816	-0.31	0.17	-8.09	4.28
2004	2,436	-0.06	0.003	-0.78	0.35	1,899	-0.29	0.09	-5.94	2.00
2005	2,436	-0.09	0.03	-1.45	0.83	2,066	-0.34	0.16	-6.68	3.47
2006	2,390	-0.16	0.06	-3.19	1.38	2,292	-0.24	0.10	-4.25	2.20
2007	2,280	0.04	-0.05	0.90	-1.06	2,501	-0.30	0.07	-5.78	1.22
2008	1,981	-0.10	0.02	-1.30	0.41	2,106	-0.24	0.16	-4.85	3.73
2009	2,000	-0.10	0.07	-1.89	0.60	2,074	-0.23	0.15	-5.00	3.73
2010	2,032	-0.09	-0.02	-1.52	-0.34	2,045	-0.34	0.17	-7.83	3.73
2011	1,907	-0.03	0.02	-0.43	0.52	1,913	-0.26	0.14	-5.42	2.56
2012	1,865	-0.10	0.01	-1.51	0.15	1,879	-0.16	0.07	-2.95	1.28
2013	1,840	0.05	-0.05	0.28	-0.99	1,918	-0.31	0.15	-6.88	3.90
2014	1,888	-0.05	-0.005	-0.95	-0.10	1,821	-0.23	0.15	-4.72	2.89
2015	1,814	0.04	-0.12	0.99	-2.61	1,730	-0.25	0.11	-4.64	2.20
2016	1,748	-0.09	0.02	-0.84	-0.14	1,705	-0.32	0.18	-6.58	3.93
2017	1,694	-0.08	-0.0005	-1.00	-0.20	1,804	-0.13	0.08	-2.97	1.20
2018	1,633	0.02	-0.06	0.84	-1.18	1,801	-0.29	0.08	-5.66	1.53
2019	1,635	-0.19	-0.03	-2.61	-0.22	1,762	-0.20	0.11	-4.15	1.79

Table B.1: Discontinuity Measures in the U.S. and European Samples. Earnings are scaled by total assets. N: number of observations; SLD: small loss deviation; SPD: small profit deviation; $t(SD_{-1})$ and $t(SD_1)$: standardized differences t-statistic for the first loss and profit interval, respectively. The measures are defined in Section 4.3. The interval width is 0.015.

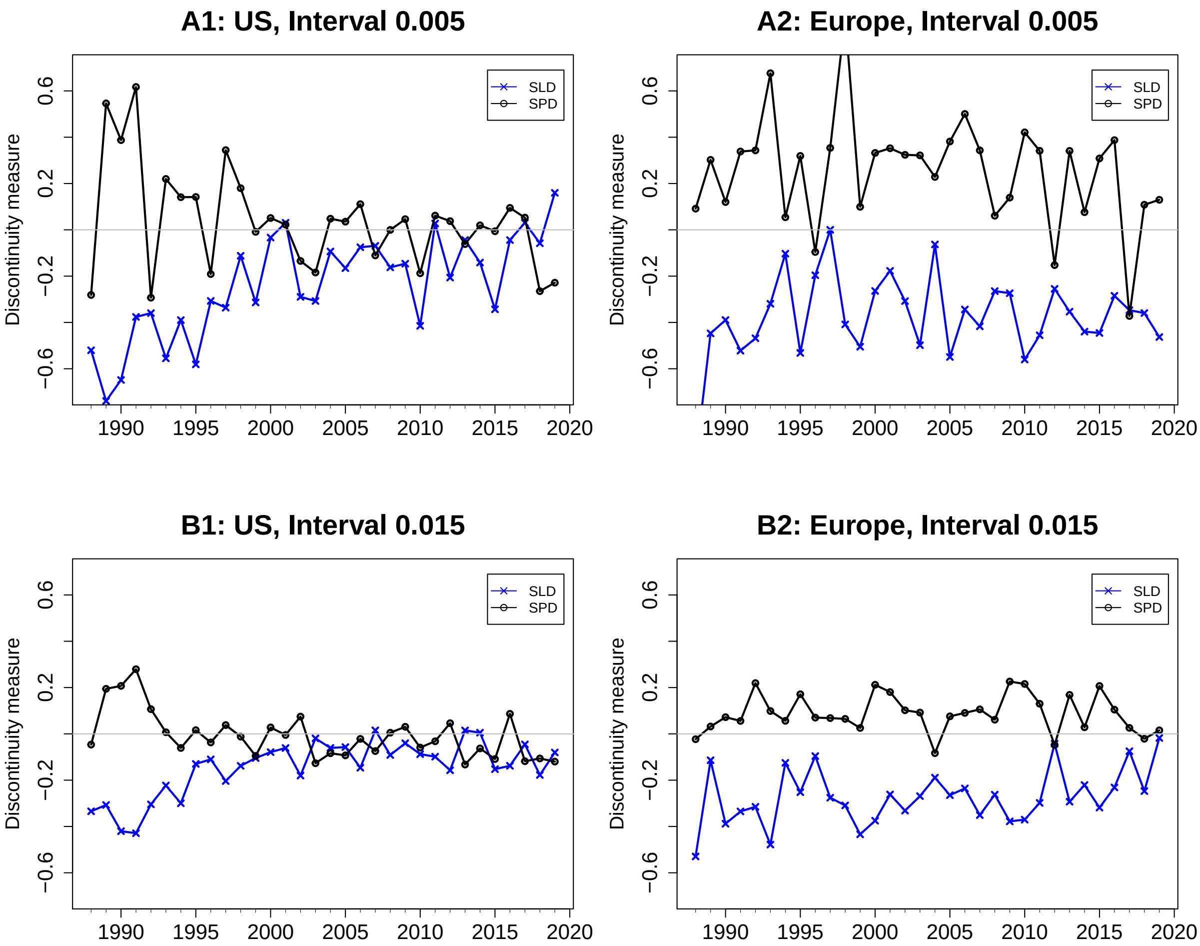

Figure B.1: Discontinuity Measures for the U.S. and Europe for Scaling With the Market Value of Equity. The discontinuity measures capture the share of excess observations (pos. sign) or missing observations (neg. sign) in the intervals of scaled earnings directly below and above the zero threshold. *SLD*: small loss deviation; *SPD*: small profit deviation. Earnings are scaled with the market value of equity at the beginning of the year.

Country	Code	UAI	1988–2003			2004–2019		
			N	SLD	SPD	N	SLD	SPD
United States	USA	46	49,951	-0.16	0.08	52,935	-0.10	0.05
United Kingdom	GBR	35	7,990	-0.15	0.08	6,265	-0.14	0.02
France	FRA	86	3,191	-0.29	0.12	4,013	-0.31	0.10
Germany	GER	65	2,988	-0.28	0.11	3,904	-0.24	0.17
Sweden	SWE	29	928	-0.17	0.16	1,867	-0.20	0.02
Italy	ITA	75	1,176	-0.23	0.15	1,542	-0.28	0.14
Switzerland	SWI	58	1,184	-0.37	0.17	1,427	-0.30	0.15
Netherlands	NET	53	1,112	-0.36	0.08	922	-0.51	0.18
Finland	FIN	59	595	-0.39	0.39	885	-0.24	0.04
Norway	NOR	50	509	-0.17	0.13	969	-0.05	0.17
Denmark	DEN	23	819	-0.20	0.09	520	-0.19	0.09
Belgium	BEL	94	604	-0.40	0.31	623	-0.15	-0.10
Spain	SPA	86	336	-0.45	-0.003	762	-0.40	0.29
Greece	GRE	112	229	-0.36	0.49	851	-0.13	0.21
Austria	AUT	70	290	-0.52	0.22	367	-0.45	0.18
Portugal	POR	104	323	-0.44	0.29	249	-0.42	0.51

Table B.2: Discontinuity Measures in the First and Second Half of the Sample Period. Earnings are scaled by total assets. UAI is the Uncertainty Avoidance Index; N the number of observations; SLD the small loss deviation and SPD the small profit deviation. The measures are defined in Section 4.3. The interval width is 0.015. The rows are sorted by the number of observations for the entire period.

Appendix C

EARNINGS MANAGEMENT MODELS AND ACCOUNTING FRAUD PREDICTION

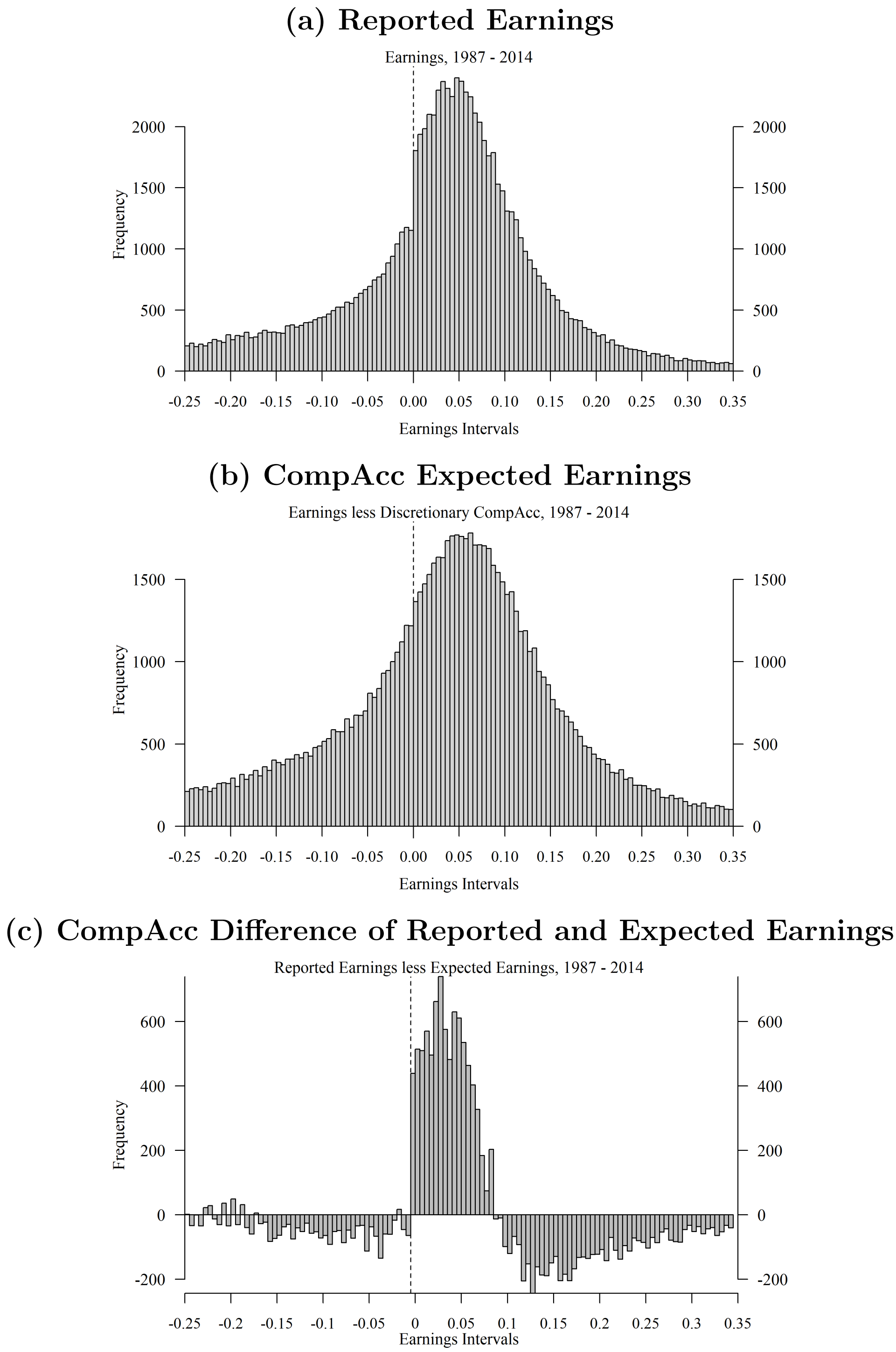

Figure C.1: Reported and Expected Earnings and the Difference of Both for CompAcc. The figure shows the observed and expected earnings distribution as well as the difference between the two. The observed distribution consists of Earnings before Extraordinary Items (WC #1551) scaled by lagged Total Assets (WC #2999). The expected distribution is the observed distribution less discretionary WCA computed from the modified Jones model. The third plot shows the difference between the observed and expected distribution. The interval width is 0.005 with 120 intervals. The dashed line indicates the zero earnings threshold. The frequency represents the number of observations in each interval.

(a) WCA Components

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
WCA	18	0.206	0.248	-0.137	-0.019	0.464	0.535
Net Working Capital	18	0.340	0.379	-0.218	0.083	0.570	1.134
Current Assets	18	1.411	1.610	0.199	0.552	1.481	7.387
Cash	18	0.743	1.541	0.001	0.063	0.813	6.743
Current Liabilities	18	0.404	0.188	0.170	0.253	0.527	0.857
Short Term Debt	18	0.076	0.121	0.000	0.000	0.070	0.390

(b) CompAcc Components

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
CompAcc	18	0.545	0.660	-0.319	0.089	1.407	1.407
Common Equity	18	2.075	2.371	0.358	0.759	2.344	9.375
Cash	18	0.743	1.541	0.001	0.063	0.813	6.743

(c) Jones-Type Model Components

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Sales	18	1.377	0.750	0.446	0.725	2.021	2.777
Receivables	18	0.360	0.250	0.065	0.169	0.540	0.842
PPE	18	0.232	0.209	0.035	0.088	0.262	0.841
Earnings	18	-0.001	0.201	-0.451	-0.100	0.116	0.342

Table C.1: Summary Statistics for Accruals and Their Components. The table shows the summary statistics for the variables required to compute accruals.

Panel (a) shows the summary statistics of the change in annual WCA:
$WCA_{i,t} = \Delta\ Current\ Assets\ (WC\ \#02201)_{i,t} - \Delta\ Cash\ \&\ Short\ Term\ Investments\ (WC\ \#02001)_{i,t} - (\Delta\ Current\ Liabilities\ (WC\ \#03101)_{i,t} - \Delta\ Short\text{-}Term\ Debt\ (WC\ \#03051)_{i,t})$.

Net working capital represents the total share of total assets computed as follows:
$Net\ Working\ Capital_{i,t} = Current\ Assets\ (WC\ \#02201)_{i,t} - Cash\ \&\ Short\ Term\ Investments\ (WC\ \#02001)_{i,t} - (Current\ Liabilities\ (WC\ \#03101)_{i,t} - Short\text{-}Term\ Debt\ (WC\ \#03051)_{i,t})$.

Panel (b) shows the CompAcc, which are calculated as annual changes:
$CompAcc_{i,t} = \Delta\ Common\ Equity\ (WC\ \#03501)_{i,t} - \Delta\ Cash\ \&\ Short\ Term\ Investments\ (WC\ \#02001)_{i,t}$.

Panel (c) shows the summary statistics for the explanatory variables of the modified Jones model, consisting of $Sales\ (WC\ \#01001)_{i,t}$, $Receivables\ (WC\ \#02051)_{i,t}$, and $PPE\ (WC\ \#02301)_{i,t}$.

All variables are scaled by lagged $Total\ Assets\ (WC\ \#02999)$.

Appendix D
PREDICTING ACCRUALS

Variable name
AU 08141 Net Sales Perc Working Capital
AU 08401 Total Asset Turnover
AU 08406 Assets per Employee
AU 08411 Capital Expenditure Perc Gross Fixed Assets
AU 08416 Capital Expenditure Perc Total Assets
AU 08421 Capital Expenditure Perc Sales
AU 08426 Accumulated Depreciation Perc Gross Fixed Assets
AU 08431 Net Sales to Gross Fixed Asset
BS A 02001 Cash Short Term Investments
BS A 02003 Cash
BS A 02008 Short Term Investments
BS A 02051 Receivables Net
BS A 02097 Raw Materials
BS A 02098 Work in Process
BS A 02099 Finished Goods
BS A 02100 Progress Payments Other
BS A 02101 Inventories Total
BS A 02140 Prepaid Expenses
BS A 02149 Other Current Assets
BS A 02201 Current Assets Total
BS A 02250 Other Investments
BS A 02256 Investment in Associated Companies
BS A 02258 Long Term Receivables
BS A 02301 Property Plant and Equipment Gross
BS A 02401 Accumulated Depreciation
BS A 02501 Property Plant and Equipment Net
BS A 02647 Deferred Charges
BS A 02648 Tangible Other Assets
BS A 02649 Total Intangible Other Assets Net
BS A 02652 Other Assets
BS A 18375 Land
BS A 18376 Buildings
BS A 18377 Machinery Equipment
BS A 18379 Property Plant Equipment Other
BS A 18380 Transportation Equipment
BS A 18381 Property Plant Equipment under Capitalized Leases
BS A 18390 Construction Work in Progress
BS A S 02300 Total Assets As Reported
BS A S 02502 Goodwill Gross
BS A S 02507 Brands Patents Net
BS A S 02508 Brands Patents Accumulated Amortization
BS A S 02509 Brands Patents Gross
BS A S 02510 Licenses Net
BS A S 02511 Licenses Gross
BS A S 02512 Licenses Accumulated Amortization
BS A S 02513 Other Intangibles Net

BS A S 02514 Other Intangible Assets Gross
BS A S 02515 Other Intangible Assets Accumulated Amortization
BS A S 02516 Computer Software Gross
BS A S 02517 Computer Software Accumulated Amortization
BS A S 02654 Total Intangible Other Assets Gross
BS A S 02655 Total Intangible Other Assets Accumulated Amortization
BS A S 18165 Deferred Tax Asset Current
BS A S 18280 Goodwill Cost in Excess of Assets Purchased
BS A S 18293 Restricted Cash Current
BS A S 18297 Trade Receivables Net
BS A S 18298 Provision for Bad Debt
BS A S 18299 Computer Software Net
BS A S 18382 Computer Software and Equipment
BS A S 18408 Derivative Assets Non Current
BS A S 18409 Derivative Assets Current
BS E 03426 Minority Interest
BS E 03451 Preferred Stock
BS E 03501 Common Equity
BS L 03040 Accounts Payable
BS L 03051 Short Term Debt Current Portion of Long Term Debt
BS L 03054 Accrued Payroll
BS L 03061 Dividends Payable
BS L 03063 Income Taxes Payable
BS L 03066 Other Current Liabilities
BS L 03101 Current Liabilities Total
BS L 03245 Long Term Debt Excluding Capitalized Leases
BS L 03249 Capitalized Lease Obligations
BS L 03251 Long Term Debt
BS L 03260 Provision for Risks and Charges
BS L 03262 Deferred Income
BS L 03263 Deferred Taxes
BS L 03273 Other Liabilities
BS L 03351 Total Liabilities
BS L 18183 Deferred Taxes Credit
BS L 18184 Deferred Taxes Debit
BS L 18281 Non Convertible Debt
BS L 18282 Convertible Debt
BS L S 03062 Interest Payable
BS L S 03069 Other Accrued Expenses
BS L S 03255 Total Debt
BS L S 03261 Pension Postretirement Benefits
BS L S 18141 Lease Commitments Year 1
BS L S 18142 Lease Commitments Year 2
BS L S 18143 Lease Commitments Year 3
BS L S 18144 Lease Commitments Year 4
BS L S 18145 Lease Commitments Year 5
BS L S 18146 Lease Commitments Over 5 Years

BS L S 18166 Deferred Tax Liability Current
BS L S 18232 Current Portion of Long Term Debt
BS L S 18286 Derivative Liabilities Non Current
BS L S 18287 Derivative Liabilities Current
BS L S 18289 Non Redeemable Preferred Stock
BS L S 18290 Redeemable Preferred Stock
BS L S 18352 Unfunded Pension Liabilities
BS L S 18851 Comprehensive Income Pension Liability
BS L S 18852 Comprehensive Income Hedging Gain Loss
BS L S 18854 Comprehensive Income Other
CF 04001 Net Income Starting Line
CF 04049 Depreciation and Depletion
CF 04050 Amortization of Intangible Assets
CF 04051 Depreciation Depletion Amortization
CF 04101 Deferred Income Taxes Investment Tax Credit
CF 04151 Total Other Cash Flow
CF 04199 Deferred Income Taxes
CF 04201 Funds from Operations
CF 04251 Net Proceeds from Sale Issue of Common Preferred
CF 04301 Proceeds from Stock Options
CF 04302 Other Proceeds from Sale Issuance of Stock
CF 04351 Disposal of Fixed Assets
CF 04355 Net Assets from Acquisitions
CF 04401 Long Term Borrowings
CF 04440 Decrease in Investments
CF 04446 Other Sources Financing
CF 04447 Other Uses Financing
CF 04448 Other Sources Uses Financing
CF 04551 Cash Dividends Paid Total
CF 04601 Capital Expenditures Additions to Fixed Assets
CF 04651 Additions to Other Assets
CF 04701 Reduction in Long Term Debt
CF 04751 Com Pfd Purchased Retired Converted Redeemed
CF 04760 Increase in Investments
CF 04795 Other Uses Investing
CF 04796 Other Sources Investing
CF 04797 Other Uses Sources Investing
CF 04821 Increase Decrease in Short Term Borrowings
CF 04825 Decrease Increase in Receivables
CF 04826 Decrease Increase in Inventories
CF 04827 Increase Decrease in Accounts Payable
CF 04828 Increase Decrease in Income Taxes Payable
CF 04829 Increase Decrease in Other Accruals
CF 04830 Decrease Increase in Other Assets Liabilities
CF 04831 Funds from for Other Operating Activities
CF 04840 Effect of Exchange Rate on Cash
CF 04851 Increase Decrease in Cash Short Term Investments

CF 04860 Net Cash Flow Operating Activities
CF 04870 Net Cash Flow Investing
CF 04890 Net Cash Flow Financing
CF 05376 Common Dividends Cash
CF 05401 Preferred Dividends Cash
CF S 04053 Asset Disposal
CF S 04056 Equity In Earnings
CF S 04057 Other Cash Flow
CF S 04148 Interest Paid
CF S 04150 Taxation
CF S 04500 External Financing
CF S 04900 Increase Decrease in Working Capital
G 06010 General Industry Classification
G 06011 Industry Group
G 07021 SIC Code 1
G 07800 Parent Auditor 1
G 07011 Employees
G 08001 Market Capitalization
G 08006 Trading Volume Amount
G 08021 Closely Held Shares Perc
GR 08616 Equity Growth
GR 08621 Total Assets Growth
GR 08631 Net Sales Revenues Growth
IS 01001 Net Sales or Revenues
IS 01051 Cost of Goods Sold
IS 01100 Gross Income
IS 01101 Selling General Administrative Expenses
IS 01148 Depreciation
IS 01149 Amortization of Intangibles
IS 01150 Amortization of Deferred Charges
IS 01151 Depreciation Depletion Amortization
IS 01230 Other Operating Expenses
IS 01249 Operating Expenses Total
IS 01250 Operating Income
IS 01251 Interest Expense on Debt
IS 01253 Extraordinary Credit Pre tax
IS 01254 Extraordinary Charge Pre tax
IS 01255 Interest Capitalized
IS 01262 Other Income Expense Net
IS 01266 Non Operating Interest Income
IS 01401 Pre tax Income
IS 01451 Income Taxes
IS 01501 Minority Interest
IS 01503 Equity in Earnings
IS 01505 Discontinued Operations
IS 01551 Net Income before Extraordinary Items Preferred Dividends
IS 01601 Extraordinary Items Gain Loss Sale of Assets

IS 01701 Preferred Dividend Requirements
IS 01706 Net Income after Preferred Dividends Basic EPS
IS 18185 Income Tax Credits
IS 18186 Current Domestic Income Tax
IS 18187 Current Foreign Income Tax
IS 18188 Deferred Domestic Income Tax
IS 18189 Deferred Foreign Income Tax
IS S 01084 Staff Costs
IS S 01155 Amortization of Other Intangibles
IS S 01201 Research Development Expense
IS S 01204 Taxes Other than Income Taxes
IS S 01306 Gain Loss on Disposal of Assets
IS S 01352 Foreign Exchange Transactions
IS S 01651 Net Income Bottom Line
IS S 01705 Net Income Used to Calculate Fully Diluted Earnings per Share
IS S 01751 Net Income Used to Calculate Earnings per Share
IS S 18140 Rental Operating Lease Expense
IS S 18150 Income from Continuing Operations
IS S 18155 Operating Income before Depreciation Amortization Operating EBITDA
IS S 18191 Earnings before Interest and Taxes EBIT
IS S 18198 Earnings before Interest Taxes Depreciation Amortization EBITDA
IS S 18200 Discontinued Operations Total
IS S 18218 Cumulative Effect of Accounting Change
IS S 18227 Restructuring Expense
IS S 18274 Impairment of Property Plant Equipment
IS S 18275 Impairment of Financial Fixed Assets
IS S 18324 Unrealized Valuation Gains Losses Total
IS S 18574 Unrealized Valuation Gains Losses Hedges Derivatives
Lev 08201 Equity Perc Total Capital
Lev 08241 Common Equity Perc Total Assets
Lev 08266 Fixed Assets Perc Common Equity
Lev 08271 Working Capital Perc Total Capital
Lev 08287 Total Assets Common Equity Ratio
Lev 15121 Total Capital Perc Assets
Liq 08101 Quick Ratio
Liq 08106 Current Ratio
Liq 08111 Cash Equivalents Perc Total Current Assets
Liq 08121 Receivables Perc Total Current Assets
Liq 08131 Accounts Receivables Days
PR 08306 Gross Profit Margin
PR 08311 Cash Flow Sales
PR 08316 Operating Profit Margin
PR 08321 Pre tax Margin
PR 08326 Return on Assets
PR 08331 Cost Goods Sold Sales
PR 08351 Sales per Employee
PR 08361 Operating Income Total Capital

PR	08366	Net Margin
PR	08376	Return on Invested Capital
PR	08381	Cash Earnings Return on Equity
PR	08656	Reinvestment Rate Total
SP	08801	Total Investment Return
SP	09100	Price Earnings Ratio High
SP	09101	Price Earnings Ratio Low
SP	09104	Price Earnings Ratio Close
SP	09304	Price Book Value Ratio Close
SP	09604	Price Cash Flow Ratio
SP	09904	Price Sales per Share Ratio

Table D.1: All Variable Names. The table lists all names of the used variables. The characters before the digit refer to the Refinitiv Worldscope section as described below. The second element stands for the five-digit code from Refinitiv Worldscope. The last element is the Refinitiv Worldscope name according to the industrial template (Thomson-Reuters, 2013).
AU = Asset Utilization
BS A = Balance Sheet Assets
BS A S = Balance Sheet Assets Supplementary fields
BS E = Balance Sheet Equity
BS L = Balance Sheet Liabilities
BS L S = Balance Sheet Liabilities Supplementary fields
CF = Cash Flow
CF S = Cash flow Supplementary fields
G = General
GR = Growth
IS = Income Statement
IS S = Income Statement Supplementary fields
Lev = Leverage
Liq = Liquidity
PR = Profitability
SP = Stock Performance