Gender, Sentiment, and Market Power: Essays on market inefficiencies

Dissertation to obtain the doctoral degree of Economic Sciences (Dr. oec.)

Faculty of Business, Economics and Social Sciences

University of Hohenheim

Institute for Health Care and Public Management

submitted by

Kylie Ann Braegelmann

from *Alexandria, Minnesota, United States of America*

2024

Date of Oral Examination: 02.07.2024
Dean of the Faculty of Business, Economics, and Social Sciences: Prof. Dr. Jörg Schiller
First Reviewer: Prof. Dr. Jörg Schiller
Second Reviewer: Prof. Dr. Alfonso Sousa-Poza
Chair of the Examination Committee: Prof. Dr. Robert Jung

"Stay at the center of the circle, and let all things take their course."

Lao Tzu

UNIVERSITÄT HOHENHEIM

Abstract

Faculty of Business, Economics, and Social Sciences

Institute for Health Care and Public Management

Doctor of Philosophy

Gender, Sentiment, and Market Power: Essays on market inefficiencies

by Kylie Ann BRAEGELMANN

This dissertation contributes to the literature on financial market inefficiencies. In Chapter 2, my co-author and I investigate a two-pronged regulation in the German private health insurance market, revealing the mixed effectiveness of measures aimed at decreasing insurers' acquisition costs and increasing value of insurance for consumers in a market where brokers have market power. While the introduction of a minimum cancellation liability period appears effective, a commission cap shows limited success and unintended consequences for new business. In Chapter 3, my co-author and I provide an empirical test of social role theory, analyzing the relationship between gender equality and perception of income fairness in European countries. The results imply that higher gender equality is associated with individuals of all genders perceiving their income as fair, suggesting societal returns to gender equality; moreoever, the results also provide weak evidence that perception of income fairness converges with higher gender equality, which is in line with social role theory. In Chapter 4, my co-author and I explore gender bias in market reaction to CEO announcements, finding a negative reaction to female CEOs, with the bias diminishing over time and varying by firm size. In Chapter 5, I conduct sentiment analysis on new CEO announcements, showing an overall tendency for firms to use positive sentiment in new CEO announcements, and demonstrating that investors react to narrative sentiment, even in a setting where it is unverifiable. Overall, these studies enhance our understanding of market inefficiencies and their regulatory challenges, providing valuable insights for both academic research and practical implications in financial markets.

Acknowledgements

It takes a village. To mention just a few, I humbly thank:

Jörg Schiller, Alfonso Sousa-Poza, Casey Rothschild, and Joe Santos for being excellent mentors. I would not be here without them.

Steffen Otterbach and Nacasius Ujah for their role as co-authors and teachers.

Ann Moeller, for teaching me how to think by writing. Katherine Malone, for teaching me to see the world through a different lens.

Katharina, Markus, Tim, Andreas, Rebecca, and Felix for being a Lehrstuhl family.

My family and friends, for being cheerleaders, role-models, celebrators, and shoulders to cry on–for being there through it all.

My parents, Jay & Liz; my in-laws, Becky & Jeff; and the Hohenheimer Zwerge, without whose support I could never have completed my dissertation.

Jon, for keeping the ship on course.

Mathilde, for motivating me to be better than I ever thought I could be.

...

Contents

List of Figures

List of Tables

List of Abbreviations

AC	Acquisition costs
AR	Abnormal return
BaFin	Federal Financial Supervisory Authority (Bundesanstalt für Finanzdienstleistungsaufsicht)
CAR	Cumulative abnormal return
CEO	Chief executive officer
DD	Differences-in-differences
EIGE	European Institute for Gender Equality
ESS	European Social Survey
GE17	Overall gender equality index - 2017
GEI	Gender equality index
HI	Health insurance
JAEG	Substitutive health insurance eligibility threshold (Jahresarbeitsentgeltgrenze)
KPI	Key performance indicator
MBA	Master of Business Administration
MD	Standard & Poor's MidCap firms (Midcap 400)
MLR	Medical loss ratio
PHI	Private health insurance
RelAC	Relative acquisition costs
ROA	Return on assets
SHI	Statutory health insurance
S&P	Standard & Poor's
SM	Standard & Poor's SmallCap firms (SmallCap 600)
SP	Standard & Poor's LargeCap firms (S&P 500)
SubstHI	Substitutive health insurance
SuppHI	Supplemental health insurance
Total AC	Total acquisition costs

For Mathilde . . .

Chapter 1

General Introduction

This dissertation fits into the broad literature on market inefficiencies arising from information asymmetries, behavioral biases, and market power. In the worst case, market inefficiencies can lead to clear market failures requiring government intervention. However, market inefficiencies can also be difficult to measure and recognize, such as those arising from gender bias. Recognizing and addressing these inefficiencies is critical for the healthy functioning of financial and intermediary markets, which ultimately have ramifications for the overall welfare of society. In this dissertation, I focus on market power in insurance markets and possible behavioral biases in financial markets. The findings provide important empirical evidence for their relevant literatures, as well as practical implications for addressing inefficiencies in financial markets.

In Chapter 2 of this dissertation, my co-author and I examine the efficacy of a two-pronged reform to reduce total acquisition costs resulting from brokers' market power in the German market for private substitutive health insurance. In this market, new business is very profitable and very limited, resulting in market power for brokers, who represent the sole access point to new contracts. Prior to the reform, the market was characterized by very high commissions and strategic reshuffling of consumers in order to maximize brokers' provisions (Tica and Weißenberger 2022). These high acquisition costs were ultimately passed on to consumers via higher premiums; however, consumers gained no additional insurance for this additional cost. The two-pronged 2012 reform (Gesetz zur Novellierung des Finanzanlagenvermittler- und Vermögensanlagenrechts) consisted of a cancellation liability period of five years, whose goal was to reduce reshuffling; and a commission cap, which limited commission payments to overall 9.9 monthly premiums.

The results of Chapter 2 indicate that the introduction of a minimum cancellation liability period likely reduced reshuffling of customers, but the commission cap did not appear to have a significant impact on the level of total acquisition costs. Rather, it appears that the commission cap likely reduced the highest peak commissions and contributed to a decline in new business. Therefore, the results underscore

the difficulty of implementing cost regulation in insurance markets, but find support for the efficacy of minimum cancellation liability regulation.

In Chapter 3, my co-author and I provide an empirical test of the social role theory using gender equality and perception of income fairness in European countries. Our goal is to contribute to the ongoing debate surrounding the origins of gender differences: on the one hand, constructivist theories posit that gender differences arise from social constructs. On the other hand, theories of biological determinism imply that gender differences are innate. Focusing on social role theory, a prominent constructivist framework developed by Alice H. Eagly (1987), we extend our analysis beyond classical economic preferences to explore gender differences in the perception of income fairness. Using data from the European Social Survey (ESS) and gender equality indices from the European Institute for Gender Equality (EIGE), we analyze the relationship between gender equality and gender differences in perceived income fairness across European countries.

The results of Chapter 3 imply that higher gender equality is associated with a higher likelihood of all individuals—regardless of gender–perceiving their income as fair. This suggests that gender-equality policies, even if challenging in the short term, may yield long-term benefits for society as a whole. Additionally, our research provides initial evidence for the convergence of perception of income fairness in European countries with increasing gender equality. This is evidence in support of social role theory, which posits that gender differences should diminish as gender equality increases, because social roles become less binding. The nuanced relationship between gender equality and gender differences calls for further exploration in future research, especially in an international context.

In Chapter 4, my co-author and I examine whether and to what extent investors exhibit a gender bias in their reaction to the announcements of male and female CEOs. According to token theory (Kanter 1997), investors may have warped perceptions of female CEOs' abilities purely because of their token status. If such a gender bias does exist, this would lead to inefficient allocation of resources, as prices would no longer reflect an accurate assessment of future firm performance. In addition, we study whether gender bias changes over time (as female CEOs become more common and lose their token status) and we analyze whether gender bias is conditional on firm size (which may proxy for information).

The results of Chapter 4 imply that investors react negatively to female CEO announcements compared to male CEO announcements, which is evidence in favor of the existence of a gender bias in financial markets. Furthermore, we find evidence that the gender bias changes over time, with the bias appearing to shrink as female CEOs become more common, although this effect becomes less clear following the financial crisis of 2008. Finally, we find that market reaction to female CEO announcements varies by firm size, with gender bias appearing to be largest

for the smallest firms. These findings are an important contribution to the literature on gender bias in financial markets and provide empirical support for the ability of information to mitigate gender bias (Heilmann 2012, Tosi and Einbender 1985), as well as the implications of token theory, where gender bias should decrease as women make up a larger and larger share of all CEOs (Lee and James 2007).

In Chapter 5, I examine whether firms tend to use positive sentiment when writing new CEO announcements, and whether investors react to sentiment in these announcements. The setting of new CEO announcements is particularly interesting, because narrative sentiment in this context is uniquely unverifiable; that is, investors lack the typical verification mechanisms of financial figures or third-party supervision (Healy and Palepu 2001). Furthermore, the implications of voluntary disclosure theory (Dye 2001) indicate that firms are likely to emphasize positive information and deemphasize negative information, leading to positively-skewed sentiment in CEO announcements. Therefore, I am able to analyze whether investors react to narrative sentiment for its own sake in a setting where sentiment is likely used strategically.

First, I provide a sentiment analysis of the sample of CEO announcements and find that, on average, new CEO announcements have a positive sentiment. In other words: CEO announcements tend to use a generally optimistic tone, leaving readers with an overall positive impression. This represents support for the voluntary disclosure theory from Dye (2001) and the concept of impression management from the literature on corporate communications (Boudt and Thewissen 2019), which posit that firms emphasize positive information and deemphasize negative information in a strategic way. Finally, using an event study to measure market reaction, I find evidence that investors react to sentiment in new CEO announcements. This is an important finding, because it represents some of the first evidence that investors react to sentiment for its own sake. If firms know that investors react positively to positive sentiment, this creates an incentive to use sentiment strategically. This dynamic may lead to noisy stock prices and inefficient capital allocation. However, it remains unclear to what extent the significant reaction is attributable to a behavioral bias or whether unverifiable sentiment contains meaningful information.

In summary, this dissertation represents a contribution to the understanding of market inefficiencies. My research has four main implications. Chapter 2 provides a test of regulatory efforts to grapple with market power in a market for private health insurance. The findings of Chapter 3 indicate that increasing gender equality can provide societal benefits. The results of Chapter 4 indicate that investors display a gender bias; more importantly, the results imply that gender bias can be mitigated by information and by an increase in the proportion of female CEOs. Finally, Chapter 5 provides important initial evidence that investors react significantly to unverifiable sentiment, which is an important step to identifying a possible market inefficiency in which companies use information asymmetries to their advantage,

potentially using sentiment as a tool to influence investors. Based on the limitations of this research, future research should explore other tools to regulate market power, empirically assess whether female CEOs consistently impact firm performance, and robustly examine whether and how unverifiable sentiment could provide reliable information about future firm performance.

Chapter 2

Lowering Acquisition Costs with a Commission Cap? Evidence from the German private health insurance market

Abstract

When consumers are neither particularly financially literate nor price sensitive, insurers have a strong incentive to pay high commissions to intermediaries for profitable new business. As a part of cost reduction regulation in the German private substitutive health insurance market, a commission cap and a minimum cancellation liability period for insurance intermediaries were introduced in 2012. Despite the fact that the commission cap lowered commissions paid to intermediaries, we provide evidence that the reform was only partly effective, as it led to a decrease in reshuffling of new business in the substitutive market, but did not significantly reduce total acquisition costs of health insurers. Our findings confirm that cost regulation is tricky and can be easily circumvented by insurers, as commission payments are only a part of total acquisition costs.

Keywords: insurance regulation, health insurance, commission cap

Note: This chapter is based on joint work with Prof. Jörg Schiller from the University of Hohenheim. It is forthcoming in *Geneva Risk and Insurance Review* as of December 2023, DOI: https://doi.org/10.1057/s10713-023-00091-4. The candidate's individual contribution focused on the development of the empirical approach, data preparation and analysis, and drafting of the paper.

2.1 Introduction

Private insurance markets play a significant role in covering essential personal risks related to longevity and health care costs. As these risks are typically complex and the financial literacy of consumers is limited, a critical goal of insurance regulation is consumer protection, specifically, to ensure that insurance products meet minimum standards. The costs of insurance products are a crucial driver for insurance coverage being valuable to consumers. Due to regulatory market barriers to entry, insurance markets are imperfectly competitive, and insurers typically have significant market power. Hence, typical premium regulation in personal lines, like the minimum requirement on insurers' medical loss ratio (MLR) approach of the Affordable Care Act and similar rate of return regulation, introduce caps on insurers' profit margins but not on firms' costs.

In insurance markets, consumers typically rely on information and advice provided by intermediaries: independent contractors who serve as matchmakers between companies and consumers. When consumers are neither financially literate nor price sensitive, high commissions can be a very effective marketing instrument for insurance providers to attract profitable new business, especially for those insurers that utilize independent intermediaries, as they are more sensitive to changes in commission payments than are tied agents. Inderst and Ottaviani (2012) show that a commission cap can help restrict the exploitation of consumers by insurers. When commissions are insurers' only marketing instrument, a commission cap can be a reasonable means through which to regulate (commission-related) costs of insurance products. However, as insurers also use other marketing instruments, like sports sponsoring and TV or online ads, it is unclear whether total acquisition costs should decrease after the introduction of the commission cap. Nevertheless, it is straightforward that a commission cap negatively affects the efficiency of marketing instruments and consequently companies' ability to attract new business.

In Germany, a commission cap for private (substitutive) health insurance contracts was introduced in 2012 to limit commissions and insurers' acquisition costs. While the vast majority of the German population is covered by statutory health insurance (SHI), approximately 12 percent of the population is privately insured and covered by individual long-term insurance contracts. The latter is called substitutive health insurance, or SubstHI. Private health insurance companies also offer supplemental health insurance contracts (SuppHI) to close the coverage gaps of SHI, e.g., for outpatient or hospital treatments. SubstHI contracts are particularly complex and difficult for consumers to compare, as the offered health insurance coverage differs significantly between insurers. The comprehensive premium regulation allows insurers to increase premiums in the private health insurance market when overall claims increase by 10 percent or when mortality increases by 5 percent. Hence, health insurers' business risk is limited, and contracts are relatively profitable. As

access to the SubstHI market is limited to a few small consumer groups, like civil servants, self-employed individuals, or employees with salaries above an income threshold, there is fierce competition for new customers. Whereas the total number of SubstHI contracts in place between 2007 and 2018 ranged from 8.5 to almost 9 million, the yearly new entry from SHI was just 115,500 to 288,200 contracts per year in that period (German Association of Private Health Insurers 2020).

Insurers rely on intermediaries for new business. According to the overview in Tica and Weißenberger (2022), tied agents—those who represent a single insurer—receive lower commissions, independent brokers typically receive commissions of 6 or 7 monthly premiums (approximately 1,500 Euros to 1,800 Euros) for a new SubstHI contract. Some brokers have even successfully negotiated commissions up to 21 monthly premiums for one new SubstHI contract. As intermediaries were previously able to keep the whole commission after the end of a one-year cancellation liability period, some insurance intermediaries began reshuffling: steering consumers from one company to another, almost on a yearly basis, to maximize their commission income (see, i.e., Tica and Weißenberger 2022 for further details). The latter behavior led to intense competition for new customers and increasing acquisition costs. Due to intense discussions within the industry and the general public and as insurers cannot self-commit to lowering commissions because of anticompetition laws, health insurers themselves urged regulators to reform commissions (Schmitt 2010). This reform, after being agreed upon in 2011, came into force in April 2012 (Gesetz zur Novellierung des Finanzanlagenvermittler- und Vermögensanlagenrechts). The reform had two main features: a commission cap for new SubstHI contracts, which effectively limits commission payments to intermediaries in the year of contract signing to nine monthly premiums and overall to 9.9 monthly premiums, and a minimum cancellation liability period of five years, such that insurance intermediaries have to repay a portion of their commission if a contract is cancelled within this time period.

We study both parts of the reform. First, we descriptively analyze the overall market effect of the minimum cancellation liability period, which was binding for all health insurers in the market. Secondly, we estimate the additional effect of the commission cap, which was only binding for insurers that paid commissions above the commission cap threshold. We use survey data from Beenken (2011) to determine which firms paid commissions above the cap prior to the reform. This allows us to approximately determine which firms were directly affected by the commission cap (treated firms), and which were not. Insurers whose maximum reported commission was above the cap (above 10 monthly premiums) are considered treated, while insurers whose maximum reported commission was at or below the cap are considered untreated. We then use a differences-in-differences framework to estimate the effect of the commission cap on treated firms. Although our treatment assignment is imperfect, we believe that our estimators represent a fair approximation of the

overall effect of the commission cap on the treated health insurers.

The German Federal Ministry of Finance perceived this reform as a success because, in its view, insurers complied with the regulation by reducing their commissions and acquisition costs (Deutscher Bundestag 2018, p. 27). Indeed, the reform led to a market decrease of an important key performance indicator (KPI) for the insurance industry: relative acquisition costs (total acquisition costs for one year divided by total premiums of that year) decreased from 8.5 percent in 2010 to 6.8 percent in 2018. However, another measure of cost—the industry-wide total acquisition cost per year—only declined slightly, from 2.649 billion Euros in 2010 to 2.544 billion Euros in 2018.[1]

The aim of our study is to evaluate the efficacy of this regulatory reform; in particular, we focus on the impact of the commission cap on the total acquisition costs of insurers. In our analysis, we take a closer look at the apparent success story, which is particularly interesting, as insurance companies typically differ with respect to their distribution strategy and use different combinations of distribution channels. Hence, insurers are differently affected by the reform; for example, insurers that only use tied agents are typically not affected by the commission cap. We analyze the impact of the reform on companies' overall acquisition costs, relative acquisition costs, new business, and acquisitions costs per new contracts using a difference-in-differences event study approach.

In our analysis, we face four main empirical challenges. First, as all insurers operate in the same market, we lack a perfect control group for our differences-in-differences analysis.[2] We assume that the commission cap is binding for insurers who had previously paid commissions above the cap. However, even insurers for whom the cap is not binding may be indirectly affected, as these firms may also adjust their business strategy given the new market environment. This indicates that spillover effects may be present in our results. The second empirical challenge is that commissions do not have to be reported, so we are also unable to observe the actual commissions paid to intermediaries. As a proxy for commissions paid, we use a survey of insurance intermediaries, in which agents and brokers report the commissions they received from particular insurers before the reform (in 2011). Therefore, our treatment assignment is based on a noisy proxy. Third, the number of observations is quite small due to the limited number of firms operating in the SubstHI market. We observe only 30 insurers that offer SubstHI contracts, but as of 2018, these 30 insurers had a combined total market share of approximately 95 percent of total premiums in the German private health insurance market. Therefore, although our estimation strategy is limited by the small number of observations, the inference benefits from the fact that we are able to observe nearly the entire population

[1] The data on industry-wide relative and total acquisition costs are published on a yearly basis by the German Association of Private Health Insurers ("PKV-Zahlenportal").

[2] The SHI market cannot serve as a control group, as the insurance coverage offered by sickness funds (providers of SHI) is standardized, such that intermediaries play no role in the market.

of firms in the market. Finally, distribution strategies are part of insurers' business decision-making. We address this possible endogeneity in a robustness check where we use propensity score matching to generate a matched sample using pre-reform average total acquisition costs. Due to the small number of observations, our ability to execute rigorous matching with statistical power is limited. However, the results using our matched sample are in line with our main results.

We find evidence that the introduction of the minimum cancellation liability period may have had its intended effect of decreasing reshuffling in the market: untreated firms had, on average, 4,500 fewer new SubstHI contracts per year after the reform. This baseline effect of the minimum cancellation liability period implies a relative reduction of about 30 percent for untreated firms compared to the pre-reform level; however, the difference is only significant at the ten percent level. In addition, the commission cap did not appear to have a significant effect on total acquisition costs or acquisition costs per new contract. This finding is particularly surprising, as our analysis indicates that the commission cap contributed to a decrease of up to 48 percent in new SubstHI business for treated companies compared to 2010 averages. This substantial effect is in line with the industry-wide observation of a 46.9 percent decrease in new SubstHI-entry from SHI and a decrease of 42.2 percent of total new SubstHI contracts.

However, we cannot conclude that the commission cap is unambiguously responsible for the dramatic decrease in new SubstHI business for treated companies. Our results rely on the assumption that no other reform or trend affected the treated and untreated insurers differently. Because, for example, a 2009 reform that introduced the partial transferability of old-age provisions for SubstHI contracts may have had long-term effects on new business, it is possible that this reform confounds our results.[3] For contracts written before 2009, consumers completely lost their old-age provision when switching insurers within the SubstHI market. This non-transferability of old-age provisions led to a significant lock-in effect, as the savings of consumers for higher health care costs at later ages were lost when switching within the SubstHI market to another insurer. The transferability of old-age provisions only applies to new contracts established in 2009 or later. However, in examining the effects of the 2009 reform, Atal et al. (2019) find that the reform did not lead to a significant increase in insurer switching. Therefore, while the 2009 reform may have some confounding effect on our results, we believe the effects are likely minimal. In fact, to the extent that the 2009 reform should have increased switching behavior, the decline in new business after the 2012 reform is perhaps even more surprising.

[3]This transferability is partial, as only parts of the total old-age provisions are transferred when switching within the SubstHI market to another insurer. The remaining part is bequeathed to the old risk pool of consumers with the previous insurer.

We also want to call attention to increases in the income eligibility threshold for entrance into the SubstHI market (Jahresarbeitsentgeltgrenze, JAEG), as the JAEG was increased regularly over the time period in question. Adjusted for inflation, JAEG increased by a modest 4.41 percent over the total time period (German Association of Private Insurers, 2020). This is in line with the fact that the threshold is increased annually based on the change in average gross salary per employee for the previous calendar year (see § 6 Sozialgesetzbuch V). In our view, it is unlikely that regular adjustment of the JAEG to changes in gross salaries substantially affected new business nor do we have any indication that the increases have affected treated and untreated insurers differently.[4] Therefore, although we remain cautious in our interpretation, our results do provide evidence that the 2012 commission cap significantly and negatively affected new SubstHI business. However, our results should be interpreted with caution commensurate with the strong identifying assumptions for our empirical approach.

Generally, we conclude that the minimum cancelation liability period appeared to reduce reshuffling, while the commission cap was likely able to trim a small share of commission peaks but was unable to affect the overall level of acquisition costs. This finding implies that insurers may have adjusted their acquisition costs through other components, like marketing expenses, resulting in an additional significant reduction in new SubstHI contracts for treated firms.

The remainder of this paper proceeds as follows. The next section reviews the institutional background and provides an economic rationale for the observed changes in market behavior. In Section 3, we introduce our data and examine the descriptive evidence. In Section 4, we provide our empirical analysis using a difference-in-differences event study methodology. In Sections 5 and 6, we discuss the results of our analysis.

2.2 Institutional background and economic reasoning

2.2.1 Private health insurance in Germany and the reform

As mentioned in the introduction, the German private health insurance market is primarily comprised of SubstHI and SuppHI, where new business in the SubstHI market is limited by institutional factors, such that only employed individuals above an income threshold (JAEG in 2019: 60.750 Euros per year), civil servants and the self-employed are eligible for the SubstHI market. Only these consumer groups can opt out of SHI for their basic health insurance coverage. For 2019, the German Association of Private Health Insurers (PKV-Verband) reported that for their 50 member companies (17 mutual and 33 stock insurers), 8.7 million SubstHI contracts and 26.7 million SuppHI contracts were in place, which led to a yearly premium income of

[4] See Appendix A.1 for a graphical comparison of the development of inflation-adjusted JAEG and new SubstHI business.

	SubstHI			SuppHI		
Year	2009	2010	2011	2009	2010	2011
Tied agents	48.8%	48.9%	46.1%	40.4%	38.4%	40.8%
Nonexclusive agents	7.7%	6.6%	7.0%	4.7%	6.0%	3.6%
Brokers	39.6%	40.4%	42.7%	28.7%	30.6%	32.2%
Banks	3.3%	3.5%	3.6%	5.5%	5.1%	6.0%
Direct selling	-	-	-	10.9%	10.6%	9.8%
Sickness funds	-	-	-	6.1%	6.3%	6.6%
Others	0.6%	0.6%	0.6%	3.6%	3.0%	0.7%
Total	100.0%	100.0%	100.0%	99.9%	100.0%	99.7%

TABLE 2.1: Market share of distribution channels for new business between 2009 and 2011

Note: shows the market share of distribution channels for new businesses between 2009 and 2011. Data are from a market survey by Tower Watson (2012). Direct selling plays no role in the SubstHI market.

27.8 billion Euros (SubstHI) and 9.1 billion Euros (SuppHI). Average monthly premiums are consequently significantly higher for SubstHI contracts (266.28 Euros) than for SuppHI (28.40 Euros).[5] The market concentration in the German private health insurance market is moderate. Based on their yearly premium incomes in 2019, the top four (eight) health insurers had a joint market share of 43.78 percent (64.36 percent).[6]

Contracts in SubstHI and SuppHI are front loaded in the spirit of long-term guaranteed renewable contracts (Pauly et al. 1995). SubstHI contracts cover costs for outpatient, hospital, and dental treatment. SuppHI contracts can, for example, cover gaps in SHI for outpatient and hospital treatments or dental services. Insurers have no right to cancel contracts (one-sided commitment) but are able to adjust premiums to cost increases based on the whole risk pool.

As in many other insurance markets, German health insurers use different distribution channels to acquire new business. A market survey by Willis Towers Watson (2012) highlights the relative importance of different distribution channels for the new SubstHI and SuppHI business before the reform (Table 2.1).

In both SubstHI and SuppHI, intermediaries (tied agent, nonexclusive agents, and brokers) are the major distribution channels. Tied agents represent one insurer, nonexclusive agents offer products from a limited number of insurers, and brokers—the most independent type of intermediary—typically offer all products from all insurers. One important specialty in SubstHI is that direct selling (via internet or phone) plays virtually no role. This is mainly driven by the complexity of

[5] See the German Association of Private Health Insurers (2020).

[6] See the Federal Financial Supervisory Authority (2020).

SubstHI contracts and the very important decision to opt out of SHI, as there is almost no way back into SHI once opting out.[7] In SuppHI, contracts are sold directly or via sickness funds, which are SHI providers.

Premium regulation for SubstHI contracts is comprehensive (see, e.g., Hofmann and Browne 2013), the main objective of which is insurers' permanent ability to settle claims and therefore to prevent insolvency. Premiums are risk based only at the date of contract signing and are calculated to remain basically constant over time, such that in early years, parts of the premiums are accumulated in old-age provisions for higher future health care costs. Premiums are a function of the expected per capita health care claims (Kopfschäden), assumed guaranteed interest rate (Rechnungszins), probability to lapse (Stornowahrscheinlichkeit), life expectancy (Sterbewahrscheinlichkeit), and expected administrative and acquisition costs. Moreover, premiums include various safety margins (e.g., at least 5 percent of total premiums). Insurers have to share their profits with policyholders via premium rebates, which are calculated on a yearly basis. According to the Mindestzuführungsverordnung (minimum funding ordinance), minimum refund shares for interest rate and risk calculation profits are 90 percent, whereas the minimum refund shares for profits from other sources (including costs) are only 50 percent. In this respect, premium regulation entails incentives similar to those of the MLR approach since higher (calculated) costs result in higher absolute profits for insurers (Cicala et al. 2019.

Subsequently, we want to explicitly illustrate why insurers have weak incentives to keep their acquisition costs low. Commissions paid by insurers are direct AC, which are (together with indirect AC, like marketing expenses and administrative costs) ultimately financed by the premium payments of policyholders. Premiums paid by customers include an absolute surcharge for acquisition and administrative costs. [8]. For example, insurers can use surpluses from cost reductions or the old-age provisions via the Zillmer method by using a net premium reserve method (Asher 2006). Suppose that the net claim costs per month of a contract are 200 Euros. If an insurer charges 5 Euros per month for administrative costs and 10 Euros for acquisition costs, then the safety margin (5% of the total premium of 226.32 Euros) is 11.32 Euros. If the insurer increases the commission in a way that the acquisition costs per contract per month increases from 10 Euros to 15 Euros, then the new safety margin (5% of total premium) is 11.58 Euros, and the total premium rises to 231.58 Euros. As policyholders receive 90% of absolute profit (which, in our example, corresponds to the safety margin), the insurer's profit increases with higher costs (from 1.13 to 1.16 Euros per month). As SubstHI contracts are complex and market transparency is rather low, insurers have weak incentives to cut costs. However, insurers

[7] One exception is if a policyholder is employed and his or her yearly income drops below the JAEG, e.g., due to part-time employment or because the JAEG is increased and then exceeds the actual income of the insured.

[8] In fact, there are four other potential ways to finance direct AC (see Milbrodt and Röhrs 2016, pp. 204-243)

differ in their business strategies: some insurers try to keep costs low to become cost leaders, while other insurers try to realize competitive advantages by superior customer service or product leadership. However, as acquisition costs reduce the value of insurance for consumers, there is some competitive and public pressure toward lowering total costs, ultimately leading to commission cap regulation.

Due to premium regulation, contracts are relatively profitable, and private health insurers face limited business risk. Hence, some insurers have paid high commission for new SubstHI contracts. Typically, agents and brokers are predominantly remunerated at the date of contract signing (signing commission). In 2011, the average cancellation liability period for brokers was 1.5 years, such that brokers had to pay back parts of their signing commissions in the case of an early cancellation of a policyholder in the first 1.5 years (Beenken 2011, p. 28). The earlier the contract is cancelled by the policyholder, the higher the repayment. Given this setup, brokers and nonexclusive agents were engaging in reshuffling activities by placing their customers with different insurers almost on a yearly basis to maximize their commission income. Tica and Weißenberger (2022) present a detailed analysis of how this reshuffling practice led to the regulatory changes that we analyze. One of the major and very prominent players in this scandal was one broker (Mehmet E. Göker and his MEG Aktiengesellschaft) who was rumored to have received commissions of up to 21 monthly premiums for new SubstHI contracts. MEG typically sold SubstHI contracts with a short cancellation liability period between one and two years and were, thus, able to offer policyholders new contracts after one year, therefore generating a new broker commission for the same customer.

The reform was first initiated in April 2010 when a representative of the federal regulator (Bundesanstalt für Finanzdienstleistungsaufsicht, BaFin) criticized commission levels and announced an initiative to decrease them. Market experts estimated that approximately 30 to 40 percent of all new contracts were related to switching insurers within the SubstHI market.[9] Due to public pressure, health insurers also expressed their goal to reduce commission levels in November 2010 but urged the federal regulator to implement appropriate regulations (commission cap and a minimum cancellation liability period). The association of health insurers argued that they were not able to implement lower commissions based on voluntary self-commitment due to potential conflicts with anticompetition laws. [10] In February 2011, Parliament members proposed legislation to limit commissions to 12 monthly premiums and to introduce a minimum cancellation liability for agents of five years. At this point in time, due to discussions between regulators, legislators, and insurers, insurers had reliable information that the reform would pass. Finally, the reform was indeed passed with minimum cancellation period of five years and a commission cap of nine monthly premiums at contract signing and 9.9 monthly premiums

[9] See Lier (2010).
[10] See Ärzte Zeitung (2010), p. 4.

for overall commission payments to agents and brokers, effective on April 1, 2012.

2.2.2 Economic reasoning and related literature

This study is related to those focusing on the effectiveness of regulatory instruments, like the MLR in US health insurance, which aim to limit the profits of insurance companies. In general, there is evidence that price or premium regulation is tricky and may have unintended effects. The early work of Averch and Johnson (1962) highlights that firms that are constrained with respect to their rate of return adopt an inefficient production plan and accumulate an excessive amount of capital as a reaction to the regulation. In addition, Knittel and Stango (2003) find that price ceilings, which have the goal of lowering market prices, ultimately serve as a focal point, which leads to tacit collusive price setting by firms. Depending on the composition of the market in question, such a focal point could ultimately lead to higher rather than lower prices.

Related to insurance markets, there is substantial evidence that prior approval premium regulation in property and casualty markets, where state insurance commissioners review filed rates and then approve or deny the proposed rate change, may lead to unintended effects. This kind of rate regulation attempts to increase insurance affordability by denying or limiting rate increases. For example, Danzon and Harrington (2001) find a significant adverse impact of regulatory premium distortions on costs for workers' compensation insurance, showing that rate suppression is positively associated with higher loss cost growth, ultimately leading to higher premiums. Weiss et al. (2010) find that the presence of rate regulation in US auto insurance markets is associated with a positive and significant increase in average loss costs and insurance claim frequency. These results highlight that regulations that are designed to enhance insurance affordability by lowering premiums have no material effect on decreasing average premiums (Harrington 2004) and may instead lead to higher average costs.

Our study is also related to the literature on insurance distribution, financial advice and problems of commissions that are paid by product providers. As indicated above, health insurers need to utilize insurance intermediaries to sell particularly SubstHI contracts but can choose different distribution strategies. In the spirit of Trigo-Gamarra and Growitsch (2010), we distinguish three different strategies. Some insurers use a multichannel approach (M-type insurers) that combines at least two distribution channels. These firms are mainly comprised of exclusive and independent agents or insurance brokers. In contrast, firms can also use only one distribution channel. Independent agency insurers (I-type insurers) are distributed exclusively through independent agencies and insurance brokers, whereas other companies only use tied agents and employee sales representatives (T-type insurers). The distribution mix is a long-term strategy choice. To adopt new channels or to abandon major channels is costly and takes several years.

Some theoretical papers, like those of Posey and Yavas (1995), Posey and Tennyson (1998), Seog (1999), and Eckardt (2007), explain the coexistence of brokers and tied agents in the same market by search cost arguments. Consumers with lower search costs prefer searching sequentially for appropriate insurance products by obtaining offers from different tied agents, and consumers with high search costs prefer the costlier broker channel, where the independent intermediary provides multiple offers at one time. As independent intermediaries can offer a greater variety of contracts, they are—according to Regan (1997) and Regan and Tennyson (1996)—more specialized in assessing the risk of consumers and matching them with appropriate insurance products. Moreover, Inderst and Ottaviani (2012a) and Focht et al. (2013) show that independent intermediaries can execute their market power via commissions, as providers have an incentive to pay substantial commissions for profitable new business.

Some empirical studies find mixed evidence related to the performance of insurers using different distribution strategies or channels. Berger et al. (1997) find that US insurers using tied agents are more cost efficient but do not have higher profitability. Brockett et al. (2005) find that US property-liability insurers that use independent intermediaries are more revenue efficient than are those insurers using tied agents or selling directly to customers. Klumpes (2004) analyzes a sample of UK life insurance firms and finds that insurers using independent intermediaries are both less cost efficient and less profit efficient compared to those insurers using tied agents. Finally, Trigo-Gamarra and Growitsch (2010) analyze the German life insurance market, which is structurally similar to the health insurance market, for the years 1997-2005 and do not find any performance advantages of specialized insurers that use just one distribution channel. Their results can explain why different distribution strategies can coexist in one market.

In the specific context of SubstHI with complex and long-term contracts, insurers rely on intermediaries to attract new business. Obviously, intermediaries have to be compensated for their services related to risk assessment and matching. For insurers, commissions are investments in new long-run business. Commissions of independent intermediaries are significantly higher than those of tied agents, as their services are more complex, and they have the ability to steer the consumer toward different insurers. Therefore, M- and I-type insurers should be more likely to pay commission above the cap and have a higher probability of being treated. For treated insurers, the commission cap limits their optimal marketing-instrument mix consisting of commissions and other marketing instruments, like sports sponsorship or TV ads.

The reform has two parts: the increased minimum cancellation liability period, which affects all firms, and the commission cap, which only directly affects treated firms. The increased minimum cancellation liability period makes the reshuffling of

consumers within the SubstHI market less attractive since consumers can be contacted after five years at the earliest, instead of one or two years. Because the minimum cancellation liability period directly affects all insurers in the market, we make the following prediction:

Prediction 1a: Following the introduction of the minimum cancellation liability period, new SubstHI business decreases for all insurers in the market, as incentives for reshuffling decrease.

Over and above the market-level effect of the minimum cancellation liability period, the commission cap reduces intermediaries' compensation for acquiring new business in the newly-regulated line of business (related to both new entry from SHI and switching within the SubstHI market). Many findings in the literature suggests that intermediaries focus their efforts on products with the highest commissions (see for example Inderst and Ottaviani 2012a) in order to maximize their profit. Therefore, whether the commission cap acts as an incentive or a disincentive for selling SubstHI contracts depends on whether intermediaries are active in other lines of business; in the case of Germany, intermediaries sell across diverse lines of business, including workers compensation, life insurance, and supplemental health insurance. Following this line of reasoning, intermediaries would respond to a commission cap on new SubstHI contracts by focusing their selling efforts on other, unregulated lines of business in order to maximize their profit. Of course, insurers could use other marketing instruments to compensate for the restriction in commission payments. However, the new marketing-instrument mix can only be second-best efficient. Based on this economic reasoning, it is straightforward to see that the commission cap and the resulting second-best efficient marketing-mix reduces the new business of treated insurers, as intermediaries' compensation for acquiring new business, related to both new entry from SHI and switching within the SubstHI market, is reduced. Therefore, we develop the following prediction:

Prediction 1b: The commission cap should lead to an additional decrease in new SubstHI business for treated companies, as intermediaries' incentives for acquiring both new business from SHI and from reshuffling consumers within the SubstHI market decrease.

The idea of the commission cap was to lower commission expenses per new contract. If Prediction 1b is correct, then total acquisition costs should—ceteris paribus—significantly decrease for treated companies due to the resulting lower level of new business. Of course, treated companies are likely to readjust their optimal marketing-instrument mix given that they are constrained with respect to their commission payments after the reform. In the spirit of Averch and Johnson (1962), treated insurers can only reach a second-best situation. Given that insurers' marketing budgets are flexible, it is not clear how treated insurers optimally readjust their marketing-instrument mix. Under reasonable assumptions, insurers could cut, increase, or

leave their total AC unchanged after the reform. Related to Prediction 2, we are conservative and test regulators' initial idea that the reform should lead to a decrease in the total AC of treated insurers.

Prediction 2: Due to lower commission costs, the commission cap should lead to lower total acquisition costs for treated companies.

2.3 Data and descriptive evidence

2.3.1 Data

To examine the effect of the reform, we use a rich dataset that includes the majority of insurers in the German private health insurance market from 2007 to 2018. We only consider health insurers that offer SubstHI contracts. New business data come from the Zeitschrift für Versicherungswesen (ZfV), an academic journal for the German insurance industry (Surminski 2019). Data on acquisition costs come from federal regulators' (BaFin) annual statistics. All other company-specific data, such as annual premiums, number of insured persons, and loss ratios, are taken from ZfV, BaFin annual statistics, and insurers' financial statements. All monetary values are adjusted for inflation with the German Consumer Price Index (Federal Statistical Office 2023) using 2007 as a baseline year. Monetary values (JAEG, premiums, and AC) are therefore shown in 2007 prices.

As insurers are not required to report commissions, there is a distinct lack of publicly available data on annual commissions and commissions per contract. To approximate the level of commissions paid by each insurer, we utilize a survey of insurance intermediaries, in which agents and brokers report the commissions they received from particular insurers in 2011, 2015, and 2017 (Beenken 2011; Beenken and Radtke 2015; Beenken and Radtke 2017). We use the results of the survey in 2011 to determine whether companies should be included in the treatment group. Insurers whose maximum reported commission was above the cap (at or above 10 monthly premiums) are considered treated, while insurers whose maximum reported commission was at or below the cap are considered untreated. For companies that have SubstHI contracts but are not included in the survey, we additionally use publicly available information about distribution strategies (PKV-Wiki 2022). Of these four companies, none use brokers (T-type distribution strategy) and are therefore considered untreated.[11]

In all, we observe rich panel data from 2007 to 2018 for 30 insurers. As of 2018, these 30 insurers have a combined total market share of approximately 95 percent of total premiums in the German private health insurance market. Furthermore, while some of these insurers do not report new business to ZfV, we are able to observe the

[11] Our results are robust to different specifications of the treatment threshold. Additional specifications are available upon request.

majority of new contracts in the market (see Appendix A.1). On average, we observe over 70 percent of all new contracts. The sample, excluding those insurers who do not report new business, represents 85 percent of the market in 2018. In addition, to ensure that the voluntary reporting of new business does not lead to selection bias in our treatment group, we analyze the characteristics of reporting and nonreporting firms. We find that there are no substantive differences with respect to assignment to the treatment group. For the related analysis, see Appendix A.2.

2.3.2 Descriptive analysis

We specifically focus on total acquisition cost (total AC) and relative acquisition costs (RelAC). Total AC includes a number of different expenses that insurers incur directly or indirectly related to new business. Direct costs are, for example, commissions for intermediaries, costs for risk assessment of new customers or costs for processing new contracts. Indirect costs are marketing expenses, general costs for the handling of new contracts that are not directly related to specific contracts or training costs for intermediaries. Apart from total AC, which measures the absolute cost for the new business of a company in a given year, both insurers and regulators also use RelAC as a relative cost measure, which is defined for company i in year t as follows:

$$RelAC_{i,t} = \frac{AC_{i,t}}{TotalPremiums_{i,t}} \tag{2.1}$$

One well-known problem of RelAC is that it relates the cost of new business in a year to the total premiums of an existing business in the same year. New companies with increasing new and relatively low existing business therefore typically have relatively large RelAC values, whereas larger companies with stable new and typically large existing business have low RelAC.[12]

In the first step of our analysis, we want to analyze the reform's impact at the market level. Table 2.2 highlights the impact of the reform on the market for SubstHI. The yearly number of new policyholders that opt out of SHI (Column 2) into SubstHI has an absolute decrease of 115,043, which is significantly lower after the reform. While some of this decrease could be explained by the regular increases in the annual eligibility threshold (JAEG, Column 1), the sharply decreasing trend after 2011 is much stronger than the slow and steady rate of annual increases in the JAEG.

In contrast, the yearly exit from SubstHI to SHI (Column 3) is just slightly reduced, by 8,886. As there is no voluntary option for policyholders to exit SubstHI, these policyholders are mainly employed individuals who fall below the income

[12]Some companies relate total AC to the (expected) premium volume of new business to use a more accurate relative cost measure that directly relates the cost for new business to the generate revenue of the related new business. However, this measure heavily relies on accurate expectations about contract duration.

threshold. Column 4 shows the total number of SubstHI contracts, and Column 5 displays the total change compared to the previous year. As Column 6 indicates, the total number of new SubstHI per year is -219,669, which is also notably lower after the reform. While more than half of this decline is due to fewer new entrants from SHI (Column 2), the other half is hence of this decline is caused by less switching within the SubstHI market. This finding is supported by the development of the annual early cancellation rate (Column 7), which represents the percentage of contracts which were cancelled in that year within 24 months of signing. This measure, which gives an indication of the extent of reshuffling, declines dramatically (-23.14 percent) after the introduction of the reform. Taken together, these developments provide support for Prediction 1a, indicating that the minimum cancellation liability period decreased reshuffling.

The market-level data indicate that the reform had a significant effect on new SubstHI business for insurers. However, total AC (Column 9) decrease only moderately after the reform by 12.1 percent (adjusted for inflation). Compared with the substantial decrease in total new SubstHI contracts of -219,669 (-42.2 percent), this modest decrease is surprising, implying that companies adjusted to the reform by increasing the utilization of other marketing instruments. For RelAC (Column 10), there was a relative reduction of 19.51 percent after the reform. However, this decrease was clearly driven both by the substantial premium increase for existing business of health insurers of 9.1 percent and the decrease in total AC of -12.1 percent. This substantial change in RelAC highlights why this KPI can be misleading, as it relates the total AC of a new business to the total premiums of the existing business. Furthermore, we note that the effects of the 2009 reform of old-age provisions are noticeable in Table 2.2. As the new regulation introduced more flexibility for consumers, the highest number of new SubstHI contracts and the highest increase in net new SubstHI contracts (Column 5) occurred in 2009. As Table 2.2 indicates, although the years before and after 2009 are quite comparable, 2009 is an exceptional year, though the 2009 reform predates and is unrelated to the reform examined in this paper.

Year	(1) JAEG	(2) New from SHI	(3) Exit to SHI	(4) Total SubstHI contracts	(5) Net change in SubstHI contracts	(6) New SubstHI contracts	(7) Early cancellation rate	(8) Total premiums	(9) Total AC	(10) Rel AC
2007	47,700	233,700	154,700	8,549,000	59,900	517,200	0.17	27,578.40	2,383.40	0.086
2008	46,930	244,900	151,000	8,639,300	90,300	534,600	0.18	27,641.62	2,464.52	0.089
2009	47,230	288,200	146,500	8,810,900	171,600	566,100	0.15	28,565.31	2,593.29	0.091
2010	48,075	227,700	153,200	8,895,500	84,600	498,000	0.20	30,004.14	2,549.57	0.085
2011	46,654	232,000	157,600	8,976,400	80,900	488,300	0.19	30,690.20	2,597.74	0.085
2012	47,083	159,900	162,400	8,956,300	-20,100	413,200	0.21	31,126.76	2,396.57	0.077
2013	47,671	123,900	161,200	8,890,100	-66,200	316,400	0.18	31,040.00	2,229.50	0.072
2014	48,462	115,500	145,700	8,834,400	-55,700	267,800	0.13	31,049.59	2,132.58	0.069
2015	49,459	120,400	140,200	8,787,300	-47,100	265,400	0.12	31,194.86	2,149.73	0.069
2016	50,448	129,100	130,600	8,772,700	-14,600	280,200	0.10	31,473.72	2,188.70	0.070
2017	50,973	129,300	133,000	8,753,400	-19,300	280,600	0.11	32,260.71	2,183.10	0.068
2018	51,742	133,700	132,900	8,736,300	-17,100	284,600	0.12	32,444.60	2,216.38	0.068
Mean 2007-11	47,318	245,300	152,600	8,774,220	97,460	520,840	0.18	28,895.93	2,517.70	0.087
Mean 2012-18	49,406	130,257	143,714	8,818,643	-34,300	301,171	0.14	31,512.89	2,213.79	0.070
Absolute diff.	2,088	-115,043	-8,886	44,423		-219,669	0.04	2,616.96	-303.91	-0.017
Relative change	4.41%	-46.9%	-5.8%	0.5%		-42.2%	-23.14%	9.1%	-12.1%	-19.5%

TABLE 2.2: Market share of distribution channels for new business between 2009 and 2011

Note: reports the development of new business, total premiums, and total AC in the German SubstHI market between 2007 and 2018. Data are from the website of the German Association of Private Insurers ("PKV Zahlenportal"). JAEG are reported in Euros. Total premiums, and total AC are reported in million Euros and adjusted for inflation using the German CPI with 2007 as a baseline year. Table is based on data for the entire PKV market.

	Untreated					Treated				
	Mean	Std. Dev.	Min	Max	n	Mean	Std. Dev.	Min	Max	n
Total premiums*	805.34	1173.81	5.83	4580.15	21	1296.91	846.66	218.02	3123.13	9
Loss ratio	0.6	0.12	0.31	0.84	21	0.64	0.08	0.53	0.77	9
Total AC*	54.72	74.31	0.6	338.83	21	122.24	65.46	26.45	285.18	9
RelAC	0.09	0.05	0.02	0.23	21	0.11	0.04	0.06	0.21	9
Total SubstHI$^+$	252.58	459.39	0.79	2148.96	21	337.94	227.36	41.36	737.57	9
Total SuppHI$^+$	506.56	652.78	0.08	3449.22	21	811.08	470.04	227.27	1717.6	9
New SubstHI$^+$	14.79	20.84	0.61	87.24	13	26.15	15.51	6.55	69.95	7
New SuppHI$^+$	43.07	33.65	3.17	174.63	13	67.95	46.44	7.17	209.37	7
Stock	0.71	0.45	0	1	21	0.78	0.42	0	1	9
I-Type	0.29	0.45	0	1	21	0.44	0.5	0	1	9
M-Type	0.24	0.43	0	1	21	0.56	0.5	0	1	9
T-Type	0.45	0.5	0	1	21	0	0	0	0	9

* in million Euros.
$^+$ in thousands.

TABLE 2.3: Summary statistics – Prior to the reform (2007 – 2010)
Note: Shows summary statistics prior to the reform. Data are from the website of the German Association of Private Insurers ("PKV Zahlenportal") as well as Beenken (2011). Total premiums and total AC are adjusted for inflation using the German CPI with 2007 as a baseline year. Table is based on the full sample.

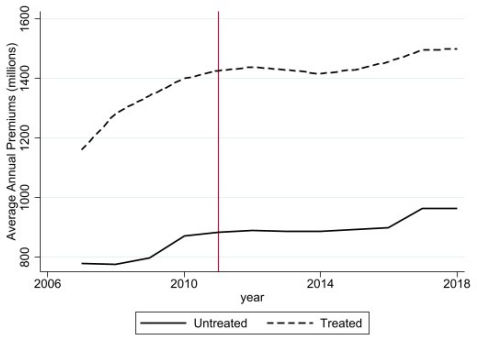
Table 2.3 displays summary statistics for the entire sample in the four sample years (2007-2010) before the announcement of the commission cap (2011).[13] On average, untreated firms are smaller in terms of total premiums, with 805.34 million in average total premiums compared with 1,296.91 million in average total premiums for treated companies. Similarly, total AC is, on average, lower for firms in the control group. Loss ratios (ratio of loss expenses to total premiums) are somewhat lower for untreated firms. However, the treated and untreated groups have similar RelAC before the commission cap, with ratios of 11 and 9 percent, respectively.

In terms of absolute enrollment, the treatment group has somewhat more total insured persons in the SubstHI line (337,940 compared to 252,580 for untreated firms). The treatment group has a higher number of insured persons in the SuppHI line, with 811,080, on average, compared to 506,560 for untreated insurers. In addition, on average, treated firms' business mix includes a somewhat higher proportion of SuppHI (71 percent SuppHI) compared to that of untreated firms (67 percent). New SubstHI contracts are somewhat higher in the treatment group, with an average of 26,150 annual new contracts compared to 14,790 in the untreated group. The number of new SuppHI contracts are also higher for firms in the treatment group. The legal form between the two groups is comparable, where treated firms are slightly more likely to be stock insurers. With regard to business strategy, untreated firms are mostly of T-type (45 percent of firms) with smaller proportions of I-type (29 percent) and M-type (24 percent) firms. Treated firms are either M-type (56 percent) or I-type (44 percent). None of the treated firms exclusively used tied agents before the introduction of the reform. For the summary statistics by business strategy, see Appendix A.3.

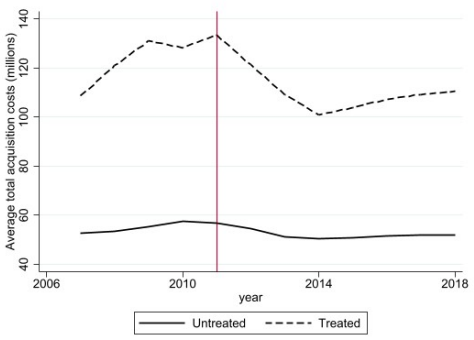
Figures 2.1A and 2.1B confirm the conclusions from the analysis of market-wide data (Table 2.2). Both treated and untreated firms have increasing annual premiums, on which the reform did not have a major impact. For untreated firms, total AC is quite stable over the whole observation period. However, total AC for treated firms increased substantially before the reform and appeared to decrease following the reform up until 2014 and increase thereafter. For RelAC (Figure 2.1C), there was a brief decreasing trend after the cap; this effect appears to diminish after 2014. This development is in line with the general consensus of regulators that the reform has been successful in dampening commission costs.

Figure 2.2A illustrates the effect of the reform on the new SubstHI business. The substantial decrease for all firms—including untreated firms—can be attributed to the minimum cancellation liability period, which may have resulted in less short-term reshuffling of contracts by intermediaries. We test Prediction 1a with a t-test of the means for the untreated group before and after the introduction of the minimum cancellation liability cancellation period. The results are shown in Table 2.4.

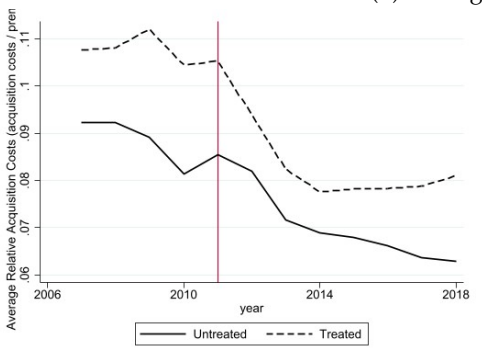
[13] We use the announcement date (2011), as insurers had reliable information and could act accordingly on that date.



(A) Average Annual Premiums



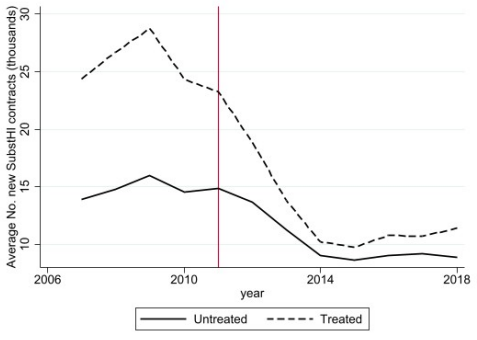
(B) Average TotalAC



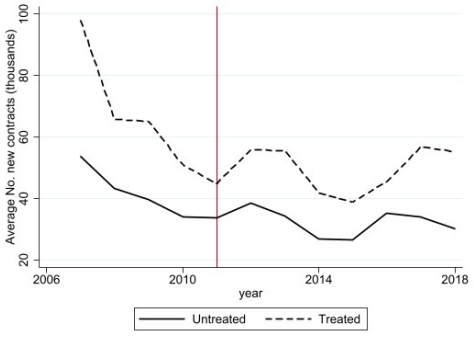
(C) Average RelAC

FIGURE 2.1: Descriptive statistics – Development of variables of interest

Note: shows the trends for the variables of interest between 2007 and 2018. Treated firms are shown with dashed lines, and untreated firms are shown with solid lines. The reform is indicated with the vertical line in 2011. All monetary values are inflation adjusted with the German CPI using 2007 as a baseline year.



(A) Average No. New SubstHI contracts



(B) Average No. New SuppHI contracts

FIGURE 2.2: Descriptive statistics – Development of new business

Note: shows the trends for the variables of interest between 2007 and 2018. Treated firms are shown with dashed lines, and untreated firms are shown with solid lines. The reform is indicated with the vertical line in 2011.

	Obs.	Mean	Std. Err.	Std. Dev.	95% Conf. Interval	
Pre-reform	50	14.794	2.947	20.836	8.873	20.716
Post-reform	111	10.243	1.924	20.267	6.431	14.056
Difference		4.5510*	3.519		-2.438	11.54

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE 2.4: Mean difference comparison in new SubstHI for untreated firms

Note: shows summary statistics prior to the reform. Data are from the website of the German Association of Private Insurers ("PKV Zahlenportal") as well as Beenken (2011). Total premiums and total AC are adjusted for inflation using the German CPI with 2007 as a baseline year. Table is based on the full sample.

On average, untreated firms have about 4,500 fewer new SubstHI contracts following the reform, which implies a reduction of 30.8 percent of new SubstHI business for untreated insurers. We consider this a rough estimation of the baseline effect of the minimum cancellation liability period. However, this difference is only significant at the 10 percent level. As can be seen in Figure 2.2A, the treated firms experience a more dramatic decline in SubstHI compared to untreated firms. This can be explained by the commission cap, which only directly affects treated firms, in addition to the effect from the minimum cancellation liability period. The apparent convergence of the new SubstHI business could be triggered by the smaller difference in commissions paid by insurers. New SuppHI business (Figure 2.2B) already decreased in the years prior to the reform. After the reform, the development is similar, but treated firms still have, on average, a higher number of new contracts per year than do untreated firms.

Based on this descriptive analysis, it appears that the reform dramatically reduced new SubstHI business overall and for treated firms in particular. However, total AC only decreased in the first years after the reform before it increased again to nearly pre-reform levels. RelAC decreased both for treated and untreated firms, but this reduction was driven by increasing premiums for existing business. Figures 2.1 and 2.2 highlight that untreated firms are also affected by the minimum cancellation liability period, but for the whole observation period, the total AC of untreated firms remains more or less unchanged. Furthermore, treated companies' new SubstHI business decreases much more dramatically than that of untreated companies. Hence, this much larger decrease is likely due to the commission cap, which only directly affects treated firms, while the overall market decrease in the new SubstHI can be explained by the fact that all firms are directly affected by the minimum cancellation liability period.

2.4 Empirical analysis

2.4.1 Difference-in-differences event study approach

To examine the effect of the commission cap on new SubstHI business (Prediction 1b) and total AC (Prediction 2), we use an event study-style approach within a difference-in-differences framework. This approach allows us to control for the initial differences between the treatment and control groups as well as the effect of time. As a result, this empirical approach should allow us to isolate the effect of the commission cap on treated insurers in each year following the reform.

To measure the average effect of the reform on treated insurers, we estimate the following equation:

$$y_{it} = \alpha_i(i = treated) + \sum_{j=2007, j\neq 2010}^{2018} \beta_j I(t = j) * I(i = treated) + FE_{year*company} + \epsilon_{it} \tag{2.2}$$

where $I(t = j)$ is the year indicator and $I(i = treated)$ is the treatment indicator. The reference year is 2010. The interaction of these indicators, $\beta_j I(t = j) * I(i = treated)$ gives us an estimate of the average effect of the commission cap on treated firms for each year in the sample. Therefore, β_j is a yearly difference-in-differences estimator. Our fully specified model includes two-way fixed effects on year and company.

2.4.2 Notes on empirical limitations

Before presenting our results, we wish to address the possible limitations of this empirical approach that are specific to our case. First, we must address the possible issue of the endogeneity of treatment. We identify distribution channels as the most likely cause of endogeneity, as each insurer develops its own distribution strategy based on individual firm objectives. Therefore, we analyze data from intermediary surveys (Beenken and Radtke 2015; Beenken and Radtke 2017) to determine whether distribution channels represent a variable that varies either by group or time but not by both.

As seen in Table 2.5, the group of untreated firms had higher concentrations of T-type distribution strategies in all waves of the survey, while the treatment group had higher concentrations of M-type distribution strategies in all waves. This finding can explain some of the heterogeneity in total AC at the outset, as treated firms, compared to untreated firms, have, on average, higher total AC, likely due to the use of more costly distribution channels. Furthermore, although both groups appear to shift away from M- and I-type strategies and toward T-type strategies, the

	I-type			M-type			T-type		
	2011	2015	2017	2011	2015	2017	2011	2015	2017
Untreated	6	2	3	5	5	3	10	7	8
%	29%	14%	21%	24%	36%	21%	48%	50%	57%
N	21	14	14	21	14	14	21	14	14
Treated	4	0	1	5	7	4	0	1	3
%	44%	0%	13%	56%	88%	50%	0%	13%	38%
N	9	8	8	9	8	8	9	8	8

TABLE 2.5: Distribution strategy
Note: shows the percentage of firms using the I-, M-, and T-type distribution strategies based on the results of the 2011, 2015, and 2017 surveys from Beenken 2011; Beenken and Radtke 2015; and Beenken and Radtke 2017. Table is based on the full sample.

underlying structural differences across the groups does not change. Therefore, distribution strategies represent a variable that is group invariant, which indicates that the difference-in-differences approach can adequately control for this effect.

Furthermore, it is likely that spillover effects exist in this analysis: because all firms are active in the SubstHI market, both treated and untreated firms are, technically, subject to the reform. We have already addressed the direct effect of the reform on untreated companies in our analysis of Prediction 1a, where the minimum cancellation liability period led to a decrease in the new SubstHI for all firms (the overall decrease shown in Figure 2.2A). However, the commission cap may also have an indirect effect on untreated firms, resulting from the changing business strategies of treated firms, which may lead to strategic changes in untreated firms. For firms whose pre-reform commissions were already below the commission cap (in our analysis, untreated firms), the cap is simply not binding. However, this does not mean that the cap had no effect on the business decisions of these firms. Unfortunately, due to the small number of firms in the market, our data do not allow us to robustly analyze spillover effects or even the intensity of treatment because when the data are grouped by different levels of maximum reported commissions, the number of observations quickly becomes small. For example, only one firm fell into the category of maximum commissions lower than 6 monthly premiums. Therefore, our analysis is unable to control for the spillover effects of the commission cap, which presents a limitation that must be taken into account when interpreting our results. However, as shown in Figures 2.1A-2.1C and 2.2B, the trends for untreated firms exhibit relatively stable development.

2.4.3 Results

In Table 2.6 and Figure 2.3, we display the results of the full specification of our model, which includes time and company fixed effects.[14] The pre-reform difference-in-differences (DD) estimators represent an explicit test of the parallel trend assumption; if the estimators are significant, then the pre-reform slopes for treated and untreated firms are significantly different from one another in that year. The fact that the pre-reform estimators are not significant indicates that the parallel trends assumption holds for our variables of interest. We consider the estimator for new SuppHI in 2007 to be noncritical for the validity of the estimation, as the effect occurred long before the introduction of the commission cap and the other estimators are statistically insignificant.

The DD estimators for new SubstHI contracts (Column 4 and Figure 2.3D) show significant and economically substantial average annual decreases in new SubstHI business following the introduction of the commission cap. The DD estimators are large and highly significant beginning in 2013 and remain significant until the end of the period of measurement. The effect of the cap is largest in 2015, where the coefficients correspond to a decrease of 11,730 contracts for treated firms compared to that in 2011. Compared to the pre-reform average annual new SubstHI business of treated firms of 24,370 contracts per year in 2010 (see Appendix A.5), this figure represents an average decrease in new business for treated companies of up to 48 percent. Although this effect is large, it is in line with the reported industry-wide reduction of 42.2 percent, which we show in Table 2.2 (Column 6). It is important to note that the estimators measure the difference in new SubstHI due to the commission cap, which only directly affected treated firms. Therefore, we find evidence in support of Prediction 1b; we conclude that the commission cap has substantially contributed to an additional decrease in new SubstHI business for treated firms. We find no significant effect on new SuppHI business. Hence, we do not see that intermediaries switched towards selling more SuppHI.

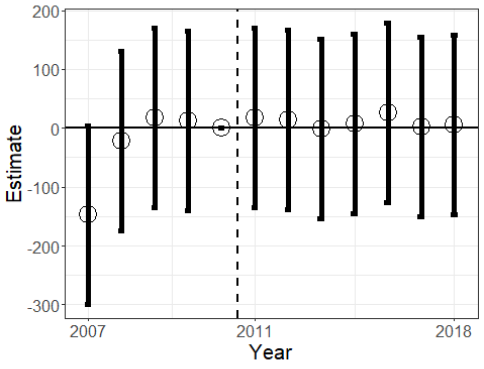
However, Prediction 2 is rejected, as the DD estimators in Table 2.6 for total AC (Column 2 and Figure 2.3B) are never strongly significantly different from zero. The weakly significant results in 2014 and 2015 are interesting; we find that the coefficients of 20.1 and 17.6 million Euros are large in economic terms but that this effect is only present briefly and is never significant above the 5 percent level. The results of a matched sample DD regression (Appendix A.6), where we match based on pre-reform average total AC, also lead to a nonsignificant effect on total AC. The results in Table 2.6 show nonsignificant effects for premiums (Column 1 and Figure 2.3A) and RelAC (Column 3 and Figure 2.3C). Finally, it is worth mentioning that these results are robust to other assignments of the treatment and control groups.

[14] The results without fixed effects can be found in Appendix A.4.

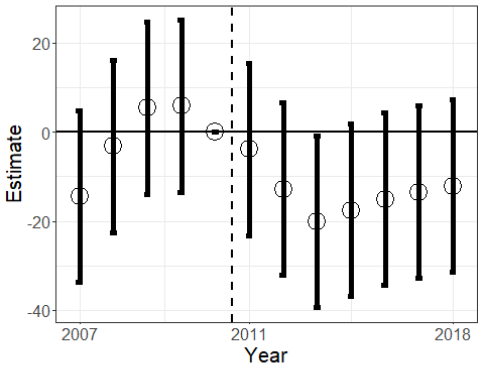
	(1)	(2)	(3)	(4)	(5)
	Annual premiums (millions)	Total AC (millions)	RelAC	No. new SubstHI contracts (thousands)	No. new SuppHI contracts (thousands)
DD 2007	-147.49*	-14.46	-0.00777	-1.277	40.48***
	(76.11)	(9.65)	(0.00839)	(3.907)	(13.85)
DD 2008	-21.46	-3.25	-0.00729	1.326	18.20
	(76.11)	(9.65)	(0.00839)	(3.768)	(13.37)
DD 2009	17.30	5.3492	-0.000123	2.704	13.09
	(76.11)	(9.65)	(0.00839)	(3.664)	(13.02)
DD 2011	12.35	5.81	-0.00314	-1.431	-6.101
	(76.11)	(9.65)	(0.00839)	(3.701)	(13.17)
DD 2012	18.05	-3.89	-0.0110	-4.654	2.225
	(76.11)	(9.65)	(0.00839)	(3.701)	(13.33)
DD 2013	14.35	-12.78	-0.0123	-10.20***	-2.552
	(76.11)	(9.65)	(0.00839)	(3.779)	(13.72)
DD 2014	-0.76	-20.10**	-0.0144*	-11.11***	-7.072
	(76.11)	(9.65)	(0.00839)	(3.638)	(12.91)
DD 2015	7.01	-17.57*	-0.0128	-11.73***	-10.05
	(76.11)	(9.65)	(0.00839)	(3.624)	(12.85)
DD 2016	26.33	-15.05	-0.0109	-10.93***	-4.462
	(76.11)	(9.65)	(0.00839)	(3.613)	(13.17)
DD 2017	2.032	-13.44	-0.00789	-11.23***	6.940
	(76.11)	(9.65)	(0.00839)	(3.613)	(12.81)
DD 2018	5.86	-12.10	-0.00484	-10.18***	9.294
	(76.11)	(9.65)	(0.00839)	(3.613)	(12.81)
Obs.	360	360	360	241	225
R-squared	0.258	0.057	0.420	0.426	0.267
No. Comp.	30	30	30	25	24
Controls	No	No	No	No	No
Fixed effects	Yes	Yes	Yes	Yes	Yes

(Standard errors shown in parentheses) $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

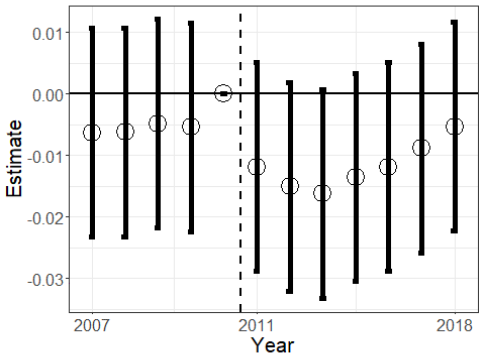
TABLE 2.6: Fully specified model
Note: displays the results of the fully specified differences-in-differences model.



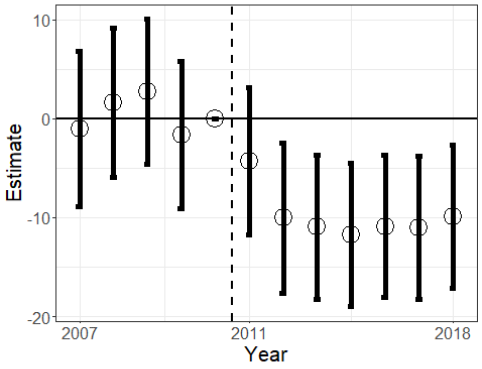
(A) Premiums



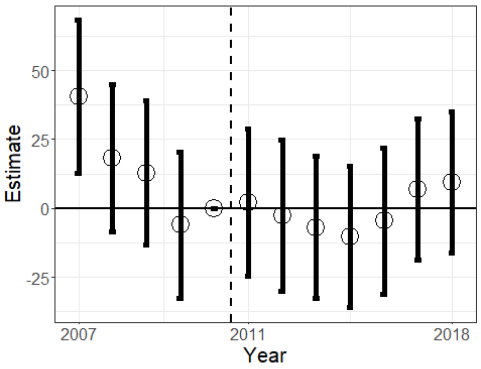
(B) Total AC



(C) RelAC



(D) New SubstHI



(E) New SuppHI

FIGURE 2.3: Results - Event study figures

Note: depicts the event-study figures of our full specification, where 2010 is the baseline year. The circles represent the coefficient of the DD estimator in each time period. The black bars indicate 99 percent confidence intervals. Confidence intervals which include 0 in pre-event periods indicate parallel trends.

Year	< 10	10 - 12	12 - 14	14-16	16-18	Comp.
2011	0%	67%	11%	11%	11%	9
2015	100%	0%	0%	0%	0%	8
2017	100%	0%	0%	0%	0%	8

TABLE 2.7: Monthly premiums by year
Note: shows the percentage of firms using the I-, M-, and T-type distribution strategies based on the results of the 2011, 2015 and 2017 surveys from Beenken, 2011; Beenken and Radtke, 2015; and Beenken and Radtke, 2017. Table is based on the full sample.

The event study analysis with respect to new business and total AC only indirectly allows for limited conclusions with respect to the change in commissions and the marketing-instrument mix of insurers. In the next section, we try to infer how commissions per new contract changed based on the observed total AC and new SubstHI and SuppHI business.

2.4.4 Commission cap - a success?

To provide more insight into the effect of the success of the commission cap on commissions per contract, we analyze the results of the 2015 and 2017 follow-up surveys of commissions from before the cap (Beenken 2011; Beenken and Radtke 2015; Beenken and Radtke 2017). Table 2.7 shows treated companies' maximum reported commissions before and after the cap.[15]

As seen in Table 2.7, the commission cap seemingly had the effect of decreasing the maximum commissions paid; in 2011, the highest maximum reported commission for a treated firm was between 16 and 18 monthly premiums. In contrast, the maximum reported commission in 2015 and 2017 was 9 monthly premiums, which corresponds to the maximum allowable commission after the reform. Indeed, eight of the nine treated firms included in both surveys reported a maximum commission of 9 monthly premiums in 2017. Therefore, unsurprisingly, the level of commissions appears unchanged, as the average reported commission decreased only slightly, from 7.79 monthly premiums in 2011 to 7.53 monthly premiums in 2017. Eight out of nine reporting firms reported a decrease in average commission, with one firm reporting no change in average commission. Of those firms whose reported commissions decreased, the average decrease was 0.25 monthly premiums. Ultimately, it appears that the commission cap did not affect the level of commissions per contract but, rather, simply trimmed away the few high commissions.[16]

Based on this analysis and the results of the regression, we find that a deeper analysis of the commission cap is necessary. RelAC is a good measure if the business environment is stable, in terms of both the business mix and the number of new contracts. However, this is not the case in the German PHI market. If the goal is

[15] The company that was included in 2011 but not in 2015 or 2017 had a maximum reported commission of 10–12 monthly premiums.

[16] These data are available upon request.

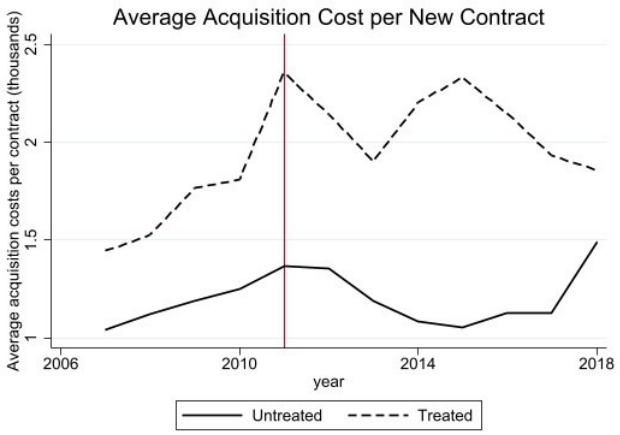


FIGURE 2.4: Descriptive statistics – Alternative measure of acquisition costs

Note: shows the trends in average AC per new HI contract between 2007 and 2018. Treated firms are shown with dashed lines, and untreated firms are shown with solid lines. The reform is indicated with the vertical line in 2011. AC are inflation adjusted with the German CPI using 2007 as a baseline year.

the accurate measurement of acquisition cost per contract, then one must look at acquisition cost per premium volume of new business, and lines should be reported separately. However, such data are not publicly reported.

To attempt to imperfectly measure acquisition costs per contract, we analyze how AC per new contract changed with the reform. This measure is defined as the total AC of insurer i in year t divided by the sum of the new SubstHI and SuppHI contracts of insurer i in year t:

$$AC\ per\ new\ HI_{i,t} = \frac{Total\,AC_{i,t}}{No.\ New\ SubstHI_{i,t} + No.\ New\ SuppHI_{i,t}} \tag{2.3}$$

This approach has the benefit of tying acquisition costs to the new business in that period. However, there is no way to determine which lines of business drive acquisition costs, as only total acquisition costs are reported. This measure, although imperfect, is better equipped to assess the efficacy of the commission cap. As shown in Figure 2.4, AC per new HI appears to have a generally increasing trend over time for both groups. Furthermore, as shown in Table 2.8, AC per new HI contract are notably higher for the treatment group than for the control group.

	Untreated					Treated				
	Mean	Std. Dev.	Min.	Max	n	Mean	Std. Dev.	Min.	Max	n
AC per new HI[1]	1.14	0.81	0.35	3.21	13	1.76	0.95	0.43	4.43	7

[1] in thousand Euros

TABLE 2.8: Summary statistics – Prior to the reform (2007 – 2010)

Note: Displays pre-reform summary statistics for the variable AC per new HI. Data are calculated based on data from the website of the German Association of Private Insurers ("PKV Zahlenportal"). AC is inflation-adjusted using 2007 prices according to the German CPI. The table is based on the sample of firms who report new SubstHI before the reform

To analyze the effect of the commission cap on our new measure of AC per new contract, we apply the same empirical approach as that outlined in Section 3, where we conduct a difference-in-differences analysis using an event-study framework. The results of the regression are shown in Table 2.9 (Column 6) and Figure 2.5.

The nonsignificant DD estimators before 2011 indicate that the parallel trends assumption holds. If the commission cap had its intended effect, then AC per new HI should decrease significantly. However, the effect is weak and ambiguous, as there is a weakly positive effect in 2016 and a weakly negative effect in 2018. Moreover, the estimators are, in general, both economically and statistically insignificant, indicating that the commission cap did not have a meaningful or lasting effect on the average AC per new HI. Based on this additional analysis, we conclude that the commission cap was not entirely effective, as there is no evidence that the cap effectively lowered commissions with respect to new business.

2.5 Discussion

In insurance markets, consumers, consumer protection agencies, and regulators are particularly concerned about the cost of insurance products, as these are a crucial driver for the value of insurance. However, another important goal of insurance regulation is the financial stability of insurance companies. One problem specific to the German market for SubstHI is that the general premium regulation, which is outlined in Section 2, is aimed at reducing the insolvency probability of insurance companies by mandating cautious premium calculation with sufficient safety margins. This regulatory approach leads almost certainly to profits, which in turn have to be shared with policyholders through a premium rebate system. The considered reform with the commission cap and a minimum cancellation liability period aims, among other things, at reducing the cost of insurance. However, as our analysis indicates, this reform is only partly effective, as such cost-based premium regulation only induces weak or even no incentives for insurance companies to keep their cost level down. This result is in line with the MLR approach in US health insurance; for example, higher costs lead to higher absolute profits for health insurance companies (Cicala et al. 2019).

The reform had its predicted effect with respect to new SubstHI business, as it contributed to a significantly negative effect on new SubstHI business for all firms, and for treated firms in particular. The relative reduction of over 40 percent is substantial—though this should be interpreted cautiously, as the identifying assumptions for the study are strong. As the descriptive analysis indicates, all companies experienced a baseline reduction in new SubstHI business due to the introduction of the minimum cancellation liability period and its reduction of reshuffling. From an economic perspective, the minimum cancellation liability period can help reduce

	(1)	(2)	(3)	(4)	(5)	(6)
	Annual premiums (millions)	Total AC (millions)	RelAC	No. new SubstHI contracts (thousands)	No. new SuppHI contracts (thousands)	AC per new HI (thousands)
DD 2007	-147.49*	-14.46	-0.00777	-1.277	40.48***	-0.970
	(76.11)	(9.65)	(0.00839)	(3.907)	(13.85)	(4.410)
DD 2008	-21.46	-3.25	-0.00729	1.326	18.20	-0.749
	(76.11)	(9.65)	(0.00839)	(3.768)	(13.37)	(1.360)
DD 2009	17.30	5.3492	-0.000123	2.704	13.09	-0.224
	(76.11)	(9.65)	(0.00839)	(3.664)	(13.02)	(1.322)
DD 2011	12.35	5.81	-0.00314	-1.431	-6.101	-0.221
	(76.11)	(9.65)	(0.00839)	(3.701)	(13.17)	(1.336)
DD 2012	18.05	-3.89	-0.0110	-4.654	2.225	-0.597
	(76.11)	(9.65)	(0.00839)	(3.701)	(13.33)	(1.336)
DD 2013	14.35	-12.78	-0.0123	-10.20***	-2.552	-1.815
	(76.11)	(9.65)	(0.00839)	(3.779)	(13.72)	(1.364)
DD 2014	-0.76	-20.10**	-0.0144*	-11.11***	-7.072	-0.432
	(76.11)	(9.65)	(0.00839)	(3.638)	(12.91)	(1.313)
DD 2015	7.01	-17.57*	-0.0128	-11.73***	-10.05	-0.102
	(76.11)	(9.65)	(0.00839)	(3.624)	(12.85)	(1.308)
DD 2016	26.33	-15.05	-0.0109	-10.93***	-4.4626	2.445*
	(76.11)	(9.65)	(0.00839)	(3.613)	(13.17)	(1.304)
DD 2017	2.032	-13.44	-0.00789	-11.23***	6.940	-0.697
	(76.11)	(9.65)	(0.00839)	(3.613)	(12.81)6	(1.304)
DD 2018	5.86	-12.10	-0.00484	-10.18***	9.294	-2.197*
	(76.11)	(9.65)	(0.00839)	(3.613)	(12.81)	(1.304)
Obs.	360	360	360	241	225	241
R-squared	0.258	0.057	0.420	0.426	0.267	0.08
No. Comp.	30	30	30	25	24	25
Controls	No	No	No	No	No	No
Fixed effects	Yes	Yes	Yes	Yes	Yes	

(Standard errors shown in parentheses) $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE 2.9: Fully specified model

Note: displays the results of the fully specified differences-in-differences model including the variable AC per new HI.

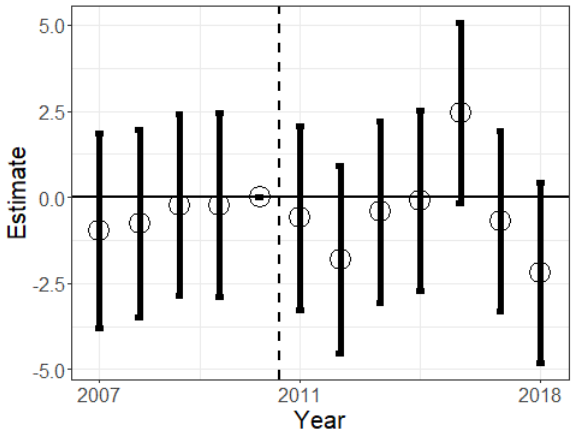


FIGURE 2.5: Results - Event study figures - Alternative measure of acquisition costs

Note: depicts the event-study figures of our full specification for total AC per new HI, with time=0 on the x-axis representing 2011. The circles represent the coefficient of the DD estimator in each time period. The black bars indicate 99 percent confidence intervals. Confidence intervals including 0 pre-event periods indicate parallel trends.

those unnecessary costs from reshuffling of business by intermediaries that want to maximize their commission income. Of course, switching insurers can be beneficial for consumers, but it is unlikely that such a switch is objectively necessary in the first years of a SubstHI contract. We conclude that the minimum cancellation period was able to limit the reshuffling of business, which was tantamount to a market failure.

However, whether or not the commission cap is economically reasonable is more difficult to assess. Our results highlight that the commission cap may negatively affect intermediaries' incentive to search for new business, which is important because it is unclear whether the high commissions before the reform represented a market failure. If insurers optimally trade off the cost for a new business (AC) and the future profit generated by that new business, then if future profits are large, high commissions are reasonable. However, as the SubstHI market is not very transparent, market discipline is quite low, such that insurers can pass on higher costs via premiums to consumers. Therefore, consumers ultimately bear higher costs through higher premiums. Our results show that the reform did not have the intended effect of reducing total AC, as these costs decreased only immediately after the reform until 2014 and then increased again. Considering the whole post-reform period, treated insurers did not significantly reduce their total AC. Instead, insurers may have reacted to the reform by readjusting their optimal marketing-instrument mix. As the result of the reform was stable total AC paired with dramatically fewer new contracts, our results imply that the resulting mix of marketing instruments may be less efficient.

2.6 Conclusion

We find evidence that the reform was only partially successful. The introduction of a minimum cancellation liability period of five years led to a decrease in reshuffling in the SubstHI market: we estimate a baseline reduction of approximately 4,500 new SubstHI contracts per year after the reform, although we note that this effect is only weakly significant. We find that the commission cap was successful in the sense that treated insurers seemingly reduced their commission below the cap threshold, but neither the minimum cancellation liability period nor the commission cap appeared to significantly reduce total AC. Furthermore, we find evidence that the commission cap contributed to a substantial decrease of over 40 percent in new SubstHI business. As treated insurers had significantly fewer new SubstHI contracts but little change in total AC, this indicates that resulting mix of marketing instruments of insurers may be less efficient.

Our analysis of the commission cap in the German PHI market indicates that directly regulating commission costs is not a very effective tool for increasing value of insurance for consumers. These findings are consistent with those of Cicala et al. (2019), Danzon and Harrington (2001), and others who find that effective cost regulation is difficult to implement. A potential tool could be regulating loss ratios in the spirit of MLR regulation in the United States. However, Cicala et al. (2019) show that this approach also has imperfections, and, in the German PHI market, regulation of loss ratios would conflict with existing premium regulation. Further research should examine the tradeoffs involved with other potential tools for increasing value of insurance for consumers.

Chapter 3

Testing the Social Role Theory in Context: Does gender equality attenuate gender differences in perceived income fairness?

Abstract

We test the implications of social role theory in the context of gender differences in perception of income fairness. Using data from the 2018 European Social Survey, we find that higher gender equality is significantly and positively correlated with the likelihood of all individuals perceiving their income as fair. Furthermore, the gender difference in income fairness perception is weakly smaller in countries with higher gender equality, which is in line with the predictions of social role theory. We conclude by discussing the practical implications of our results and introducing a more nuanced interpretation of the theoretical framework, which underscores the need for a formal synthesis of resources into the social role theory.

Keywords: social role theory, gender equality, income fairness, European Social Survey, gender differences

Note: This chapter is based on joint work with Dr. Steffen Otterbach of Universität Hohenheim. The research is yet unpublished. I wish to particularly thank Jörg Schiller, Alfonso Sousa-Poza, and the participants of the 2020 MRIC Winter Workshop for their feedback.

3.1 Introduction

The origins of gender differences have long been a source of debate in the social sciences. Social role theory and other constructivist theories posit that gender differences (in preferences, in perceptions, in decision-making, and so on) are a result of social constructs. These theories suggest that gender differences attenuate when gender equality eases the prescriptive nature of social constructs. On the other hand, others argue that gender differences are biologically determined, and, transitively, gender differences increase as gender equality increases. Moreover, recent work has pointed out the insidious nature of confirmation bias and social constructs in researching gender differences (Nelson 2014). In light of the ongoing theoretical debate and mixed empirical findings, we believe that further empirical evidence is needed to understand these conflicting theoretical perspectives.

In this paper, we take up the question of gender differences in perception of income fairness. Our measure of income fairness includes wages, social benefits, and pension benefits. As such, it is a measure of income fairness, but it is also closely tied to earned wages and social benefits, which are determined by former wages. While much of the existing literature on gender differences focuses on classical economic preferences (risk-taking, positive and negative reciprocity, patience, altruism, and trust), perception of income fairness represents a natural extension of these preferences. In particular, there is evidence that reciprocity preferences are closely tied to perceptions of wage fairness (Fehr and Gächter 2000; Falk et al. 2006). Therefore, extending this framework to the question of income fairness perception will provide a valuable test of social role theory; the results provide practical implications for governments and institutions as well as theoretical implications for scholars.

In addition, perception of income fairness is a particularly interesting test of this framework, as there is a consensus that perception of wage fairness is conditional on sex (Paul 2006; D'ambrosio et al. 2018, Pfeifer and Stephan 2019; Tyrowicz et al. 2018) to that point that sexes are often analysed separately in the literature (Mayraz et al. 2009; Schneck 2014). The consensus in the wage fairness literature echoes the related job satisfaction literature, where women are more likely to be satisfied with their jobs (Clark 1996; Clark and Oswald 1997; Clark 1996; Hauret and Williams 2017) despite being relatively disadvantaged in the labour market (Clark 1997; Sousa-Poza and Sousa-Poza 2000). Sousa-Poza and Sousa-Poza call this phenomenon the "gender/job-satisfaction paradox" (2000, p. 137).

Perception of income fairness has its own significant implications, as perceived unfairness can lead to decreased well-being and increased quitting behaviour (Clark and Oswald, 1998; D'ambrosio et al. 2018) and lower life satisfaction, especially for men (Mayraz et al. 2009). However, our primary interest in perception of income fairness is its relationship with developing gender equality. Therefore, a rigorous decomposition of the gender gap for its own sake is not the goal of this paper.

In our study, we use data from the European Social Survey (ESS) and gender equality indices from the European Institute for Gender Equality (EIGE). The ESS collects data on perceived income fairness which, when combined with the detailed gender equality indices, allows us to analyse the association between gender equality and gender differences in perceived income fairness. Using European data, we are able to analyse countries with historically high gender equality, such as the Scandinavian countries, as well as countries were gender equality movements are in earlier stages, such as Hungary. Therefore, our analysis will provide a unique and valuable insight into the underlying mechanisms proposed in social role theory.

The test of our first hypothesis indicates that gender equality is significantly positively correlated with the likelihood of perceiving income as fair. The practical implication of this finding is that, while gender-equality policies like board quotas may be challenging in the short-term, they may be worth the investment for the long-term gain of all individuals in a society. Our data imply that gender equality is a rising tide that lifts all boats; therefore, with respect to perceptions of income fairness, these policies could be worth the effort.

Furthermore, we find weak evidence that increasing gender equality coincides with converging perception of wage fairness in European countries, which supports social role theory and other constructivist theories of gender differences. That is, the gender difference in perception of income fairness is generally smaller in countries with higher gender equality; however, these differences are generally not statistically significant. This is in line with social role theory, which predicts that increasing gender equality attenuates gender differences by interrupting the self-perpetuation of gender roles. Based on a synthesis of the theory and empirical evidence, we posit that a more nuanced relationship exists between gender equality and gender differences, such as differences in perceived income fairness. This relationship, as well as the validity of the results in an international context, should be addressed in future research.

3.2 Background and theoretical framework

3.2.1 Motivation - Debating the origin of gender differences

In recent and widely cited empirical research, Falk and Hermle (2018) posit that living in a society with high measures of gender equality is analogous to possessing socioeconomic resources: when men and women have similar social and economic power, both sexes have the prerequisites that allow them to focus on understanding, developing, and expressing their individual preferences. This process of self-actualization, facilitated by sufficient socioeconomic resources, leads individuals to differentiate themselves based on their unique, underlying preferences. The authors argue that the process of self-actualization on the individual level will result in stronger differences between genders. In essence, the authors argue that gender

differences are innate, and gaining resources allows for these underlying, fundamental preferences to be discovered and displayed.

This interpretation presents a stark contrast to constructivist theories of preference formation, where preferences are formed endogenously as individuals interact with societies and institutions. For example, in the Empowerment Model, van Staveren (2013) enumerates how agency, access to resources, and gendered institutions drive the empowerment of women. Similarly, Sent and van Staveren (2019) underscore the importance of understanding how institutions and social norms influence gender differences in preferences. These constructivist theories posit that the co-evolution of institutions, power, and society create binding gender norms, wherein gender differences are a function of socioeconomic institutions (and vice versa) rather than fundamental differences based on innate or biological factors. This mismatch between recent empirical evidence in light of the increasingly accepted constructivist framework motivates this research.

3.2.2 Social role theory

As the foremost constructivist theory in the literature, social role theory posits that gender differences arise in response to social norms, which Cialdini and Trost define as, "rules and standards that are understood by members of a group, and that guide and/or constrain social behaviour without the force of laws" (1998, p. 152). These social norms arise out of interacting with and observing others. Norms can be descriptive; they describe how people generally do behave (Cialdini et al. 1990). They can also be prescriptive; they describe how people should behave (Cialdini et al. 1991, Cialdini and Trost 1998).

According to this theory, gender differences arise when individuals conform themselves to existing gender-specific social norms, called gender roles (Eagly 1987). They serve as an unwritten rulebook for preferences, perception, and behaviour in a particular society. These gender roles, "are derived from observations of the role performances of men and women and thus reflect the sexual division of labour and gender hierarchy of the society" (Eagly et al. 2000).

According to social role theory, this division of labour and gender hierarchy is not caused by innate psychological differences between the sexes, rather, from a historical perspective, "[P]hysical gender differences, in interaction with social and ecological conditions, influence the roles held by men and women because certain activities are more efficiently accomplished by one sex" (Eagly and Wood 1999, p. 412). Before modern times, these certain activities generally involved gestation, breastfeeding, and childrearing for females, while the relative physical strength of

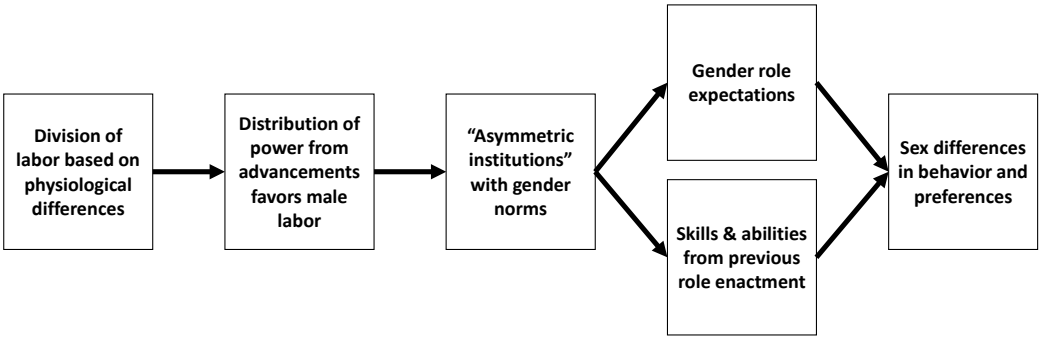


FIGURE 3.1: Gender difference formation in social role theory
Note: Developed from Eagly and Wood (1990) and Sent and Staveren (2019)

males, coupled with their lack of biological nurturing constraints, made them efficient hunters, herders, and warriors. As technological and socioeconomic developments favoured male-performed activities, socioeconomic systems became patriarchal (Eagly and Wood 1999), creating self-perpetuating gender roles from the initial distribution of labour.

Although the comparative advantages derived from a sex-based division of labour was historically advantageous, such efficiency gains are questionable in modern times (see Rendall 2017). Despite this, the gender roles that arose from the initial division perpetuate themselves. According to social role theory, this is a result of the fundamental attribution error, also referred to as correspondent inference (Jones and Harris 1967; Eagly et al. 2000). As members of society, individuals observe how others enact their role. In doing so, they underweight societal and situational influences, attributing the behaviour to the desire of the individual. For example, in observing a stay-at-home mother, one may assume that her role primarily reflects her personal preferences rather than societal or familial expectations. Figure 3.1 illustrates the underlying mechanisms of social role theory.

Social role theory implies that gender equality frees individuals from the obligation to fit into gender roles. Gender equality weakens the prescriptive nature of gender roles; it becomes acceptable for individuals to exhibit non-typical characteristics. Furthermore, because gender roles are no longer prescriptive, individuals are less likely to acquire specialized skill sets based on previous gender role enactment. Therefore, increasing gender equality has a two-fold effect on the self-perpetuation of gender roles, leading to the attenuation of gender differences. This leads to the testable prediction that, in the context of this study, increasing gender

equality should lead to a convergence in perception of income fairness by men and women.

3.2.3 Hypothesis development

We analyse how the gender difference in perception of income fairness responds to increasing gender equality. In social role theory, gender equality disrupts the self-perpetuation of gender differences by simultaneously weakening gendered expectations and decreasing the likelihood of sex-specific skills from previous role enactment. Therefore, gender equality should significantly affect the likelihood of perceiving income as fair. Based on this analysis, we formulate the following hypothesis:

H1: Likelihood of perceiving income as fair is conditional on measures of gender equality.

To formally test the predictive power of social role theory for perception of income fairness, we formulate our second hypothesis as follows:

H2: The gender difference in perceiving income as fair will decrease with increasing gender equality.

3.3 Data and methods

Our analysis is based on data from the ESS. The ESS is a cross-national longitudinal survey of private households and their living conditions and has been conducted biannually since 2002. Currently, representative data are collected from 36 countries. While ESS is mainly known as a longitudinal survey and is used as such, our income fairness and control variables originate from ESS Round 9 (2018). The main variable of interest, income fairness, is generated from the following question, where individuals are asked to characterize the fairness of their income:

"Your net [pay/pensions/social benefits] is unfairly low, fair, or unfairly high"

We generate a binary dependent variable by recoding the values, where income that is perceived as fair or more than fair is equal to 1 and income that is less than fair is equal to 0. Results do not change significantly when ordered logistic regressions are run with the original variable encoding. We choose to recode the variable of interest to make the interpretation of results more intuitive within the social role theory framework. Other results are available upon request to the authors.

We develop our sample as follows: We begin with all individuals, of which we observe 16,727 men and 19,094 women with non-missing values in our main variable of interest. After including our covariates, there are 13,932 men and 15,538 women from a total of 24 different European countries. As can be seen in Table 3.1, 44 percent of men and 35 percent of women consider their incomes fair; this difference is statistically significant. Men and women are of comparable age (approximately

	Men		Women	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Income perceived as fair [0=no; 1=yes]	0.44	0.50	0.35	0.48
Age [years]	52.92	17.27	53.86	17.56
Age^2	3098.40	1837.69	3209.80	1897.80
Partnered [0=no; 1=yes]	0.57	0.50	0.47	0.50
Number of children	1.52	1.29	1.67	1.27
Education [years]	13.30	4.22	13.20	4.19
$Education^2$	194.79	134.76	191.70	124.54
Foreign born [1=no; 2=yes]	1.09	0.29	1.09	0.29
Income quintile in country				
1. quintile	0.16	0.37	0.25	0.43
2. quintile	0.21	0.41	0.22	0.41
3. quintile	0.22	0.41	0.20	0.40
4. quintile	0.22	0.41	0.19	0.39
5. quintile	0.19	0.39	0.14	0.35

TABLE 3.1: Summary statistics - Covariate means and standard deviations by gender

Note: Summary statistics are based on the regression model of the overall gender index with 13,392 and 15,538 observations for men and women, respectively. Source: European Social Survey Round 9.

53 and 54 years, respectively). Men are slightly more likely to be in a committed partnership (partnered) at 57 percent compared to 47 percent of women; however, the difference is not significant. Both men and women have comparable number of children, on average, and years of education are similar with an average of approximately 13 years. Women and men are equally likely to be foreign-born in this sample. On average, women are slightly more likely to be in the first income quintile, with 25 percent of the sample compared to 16 percent for men, however, the difference is not statistically significant.

Similarly, Table 3.2 indicates that our sample includes large variation in perception of income fairness and large variation in country-level gender differences. Considerable gender differences exist within some countries, for example, the largest gender differences are in Croatia and Cyprus, with statistically significant gaps of 13 percent between men and women. In contrast, the smallest gaps can be found in the Netherlands, where 61 percent of men and 59 percent of women perceive their wage as fair, and Belgium, where 42 percent of men and 44 percent of women perceive their wage as fair. Interestingly, Hungary, which has the second-lowest overall perceived income fairness, also has a small 2 percent gap, with only 17 percent of men and 15 percent of women perceiving their income as fair. See Table 3.2 for more detail.

For our measure of gender equality, we use the 2017 gender equality index published by the European Institute for Gender Equality (EIGE). The overall gender equality index (GE17), which the EIGE considers the overall measure of gender

	Men		Women	
Country	Mean	Std. Dev.	Mean	Std. Dev.
Austria (AT)	0.55	0.50	0.49	0.50
Belgium (BE)	0.42	0.49	0.44	0.50
Bulgaria (BG)	0.15	0.36	0.09	0.28
Cyprus (CY)	0.35	0.48	0.22	0.41
Czech Republic (CZ)	0.32	0.47	0.25	0.43
Germany (DE)	0.42	0.49	0.38	0.48
Denmark (DK)	0.64	0.48	0.60	0.49
Estonia (EE)	0.30	0.46	0.26	0.44
Spain (ES)	0.31	0.46	0.24	0.43
Finland (FI)	0.46	0.50	0.35	0.48
France (FR)	0.33	0.47	0.29	0.45
United Kingdom (GB)	0.60	0.49	0.53	0.50
Croatia (HR)	0.29	0.45	0.16	0.37
Hungary (HU)	0.17	0.38	0.15	0.35
Ireland (IE)	0.53	0.50	0.49	0.50
Italy (IT)	0.32	0.47	0.29	0.45
Lithuania (LT)	0.18	0.38	0.12	0.33
Latvia (LV)	0.28	0.45	0.25	0.43
Netherlands (NL)	0.61	0.49	0.59	0.49
Poland (PL)	0.28	0.45	0.17	0.37
Portugal (PT)	0.26	0.44	0.18	0.39
Sweden (SE)	0.53	0.50	0.43	0.50
Slovenia (SI)	0.29	0.45	0.32	0.47
Slovakia (SK)	0.20	0.40	0.14	0.34

TABLE 3.2: Summary statistics - Perception of income fairness by country

Note: Summary statistics are based on the regression model of the overall gender index with 13,932 and 15,538 observations for men and women, respectively. Source: European Social Survey Round 9 and European Institute for Gender Equality 2017 Edition.

	Mean	Std. Dev.	Min	Max
Overall gender equality index	65.92	8.34	51.90	83.60
Domain money	79.51	8.20	61.80	88.30
Domain work	72.98	4.44	63.10	83.00
Domain power	50.37	16.42	20.60	83.40
Domain time	65.12	12.47	42.70	90.10
Domain health	87.38	4.53	77.10	94.70
Domain knowledge	61.67	6.92	49.70	73.80

TABLE 3.3: Summary statistics - Gender equality index and domains
Note: Summary statistics are based on the regression model of the overall gender index with 13,932 and 15,538 observations for men and women, respectively. Source: Gender Indices from the European Institute for Gender Equality 2017 Edition.

equality in a country, is calculated by weighting the six domains of gender equality: work, money, power, health, education, and time use. Index calculation and weighting is based on the unified Gender Equality Index methodology from the EIGE. See Figure 3.2 for a visual illustration of the GE17 values for each of the 24 countries in this analysis. Lighter shades indicate higher gender equality. As shown in Figure 3.2, the Scandinavian countries Sweden (SE) and Denmark (DK) have the highest gender equality, with overall scores of 83.6 and 77.5, respectively. Slovakia (SK) and Hungary (HU) have the lowest gender equality scores, with 54.1 and 51.9, respectively. France (FR) has a score of 74.6, United Kingdom (GB) has a score of 72.2, Spain (ES) has a score of 70.1, and Germany (DE) has an average score of 66.9.

Table 3.3 displays the summary statistics for the gender equality index and the six domains, where a higher value indicates higher gender equality. The money domain, which captures gender equality in monthly income, poverty risk, and income distributions, has an overall average of 79.51. The work domain, which captures gender equality in workforce participation, job security, and career prospects, has an average score of 72.98. The power domain, which captures gender equality in representation in political and corporate decision-making bodies, has the lowest average score, with a score of 50.37. The time domain, which captures gender equality in household, caregiving, and social activities, has an average score of 65.12. The health domain, which captures gender equality in perceived health status, behaviour, and access, has the highest average score, with 87.38. Finally, the knowledge domain, which captures gender equality in education attainment, participation, and segregation, has an average score of 61.67.

3.3.1 Estimation approach

To test whether gender equality affects the likelihood of perceiving income as fair, we jointly estimate logistic regressions in a full interaction model in which we allow for different slopes and intercepts for men and women. Sample weights are applied

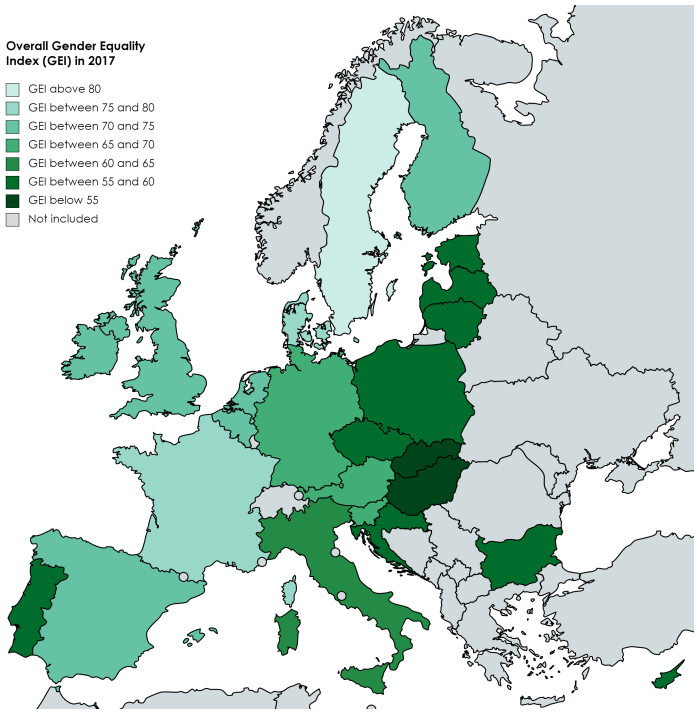


FIGURE 3.2: Overall 2017 gender equality index by country
Note: Darker (lighter) shades indicate lower (higher) overall gender equality. Source: Gender Indices from the European Institute for Gender Equality 2017 Edition.

based on the ESS standard procedure. Our binary logistic model takes the following form:

$$log(odds(y_{ij}) = \alpha + \alpha_f G_f + \beta X_i + \beta_f G_f X_i + \gamma Z_j + \gamma_f G_f Z_j \qquad (3.1)$$

where y_{ij} is a binary outcome variable indicating whether individual i in country j perceives their income as fair, and G_f indicates gender, where male = 0 and female = 1. X_i is a set of individual characteristic variables. The variable of interest is Z_j, the country-level gender equality index. Due to the full interaction of the model the slopes of interest regarding the country-level gender equality index Z_j are γ and $\gamma + \gamma_f$ for males and females, respectively. A separate regression is run for the overall gender equality index, as well as for each of the domain indices.

The choice of covariates is largely driven by the existing literature on income fairness (see D'ambrosio et al. 2018; Mayraz et al. 2009, and Pfeifer and Stephan 2019) and by data availability. In our specification, X_i includes age, age squared, a partnered binary variable (equal to 1 if in a marriage or partnership, equal to 0 otherwise), number of children, and years of education, years of education squared and a foreign-born binary variable (equal to 1 if born in of country of residence, equal to 2 otherwise). Additionally, we control for relative household income level by including income quintile within country of residence.

3.4 Results

We test whether the likelihood of perceiving income as fair is conditional on gender equality. We run full interaction regressions for the overall index as well as the index for each of the five domains. In Table 3.4 we report the results for the overall gender index regressions.

	Men		Women	
Variable	Coef	SE	Coef	SE
Overall gender index	0.045***	0.004	0.052***	0.004
Partnered [0=no; 1=yes]	-0.117	0.076	0.032	0.071
Number of children	-0.004	0.030	-0.009	0.028
Age [years]	-0.061***	0.012	-0.067***	0.010
Age^2	0.001***	0.000	0.001***	0.000
Education [years]	0.085***	0.021	0.186***	0.032
$Education^2$	-0.001**	0.001	-0.005***	0.001
Foreign born [0=no; 1=yes]	0.011	0.104	0.031	0.109
HH income [ref.: 1st quintile]				
2nd quintile	0.412***	0.116	0.409***	0.096
3rd quintile	0.672***	0.114	0.479***	0.096
4th quintile	0.885***	0.111	0.685***	0.106
5th quintile	1.275***	0.126	1.011***	0.115
Constant / Female	-3.721***	0.431	-0.925	0.572
F[a] (adjusted Wald test)	1.975			
p[a]	0.160			
N	29,470			
F	28.927			
p	0.000			

$***p < 0.01, **p < 0.05, *p < 0.1$

TABLE 3.4: Results - Overall gender index

Note: Logistic regressions were estimated jointly for men and women in a full interaction model and include sample weights. Regressions also include a constant and an intercept for being female (not reported here).

[a] Adjusted Wald test to test whether the gender index of men and women is equal.

3.4.1 Hypothesis 1 - Overall gender equality index

As shown in Table 3.4, we find that the likelihood of perceiving income as fair is conditional on gender equality for both men and women; this effect is significant at the 1 percent level for both men and women. Moreover, this effect is positive for both sexes: as gender equality increases, both men and women are more likely to perceive their income as fair.

The binary variable partnered is not significant for men or women; similarly, the number of children is not statistically or economically significant for either group. In contrast, we find that age is negative and significant for both sexes, while age-squared is positive and significant. Years of education is positive and significant for

both men and women, with the coefficient for women being notably larger than that for men. Therefore, additional education makes both men and women more likely to perceive their income as fair; however, the effect is stronger for women. Furthermore, as years of education squared is negative for both groups, this indicates that individuals with the highest levels of education may be less likely to perceive their income as fair. However, the small coefficients indicate that this effect may not be economically significant.

Foreign born status does not have a significant effect for either gender. However, relative household income, with the first quintile as a reference value, has a positive and significant effect for men and women. Unsurprisingly, the coefficients increase in economic significance as the relative household income increases. For example, a household income in the fifth quintile strongly increases the likelihood that an individual perceives their income as fair, with coefficients of 1.275 for men and 1.011 for women.

3.4.2 Hypothesis 1 - Gender equality domains

Table 3.5 reports the coefficients of the six gender index domains.

	Men		Women		
Domain	Coef	SE	Coef	SE	F^a, p
Overall gender index	0.045***	0.004	0.052***	0.004	1.975, 0.160
Work domain	0.087***	0.008	0.088***	0.007	0.016, 0.900
Money domain	0.049***	0.005	0.063***	0.005	5.209, 0.022
Knowledge domain	0.039***	0.005	0.043***	0.004	0.320, 0.572
Power domain	0.007***	0.002	0.010***	0.002	1.593, 0.207
Time domain	0.043***	0.003	0.047***	0.002	1.528, 0.216
Health domain	0.116***	0.009	0.129***	0.008	1.315, 0.252

$***p < 0.01, **p < 0.05, *p < 0.1$

TABLE 3.5: Results - Gender index domains

Note: Logistic regressions were estimated jointly for men and women in a full interaction model and include $N = 29,470$ observations and the same set of control variables as in Table 3.4. Sample weights are applied. Regressions also include a constant and an intercept for being female (not reported here).
[a] Adjusted Wald test to test whether the gender index of men and women is equal.

We find that higher gender equality in all domains has a positive and significant relationship to the likelihood of both men and women perceiving their income as fair. Of particular note, we find that perception of income fairness is conditional on the work domain of gender equality. The effect is positive and significant at the 1 percent level for both men and women; furthermore, the relatively large size of the coefficients indicate that increasing gender equality in the work domain has a particularly strong relationship to the likelihood of perceiving income as fair. Additionally, the effects in the money, knowledge, and time domains are, as expected, also economically significant. These domains are closed related to work and income, as money includes economic situation and access to financial resources; knowledge

includes educational attainment, participation, and segregation; and time includes measures of time spent on care activities.

The interpretation of the relatively small coefficient of the power domain and the relatively large coefficient of the health domain is not immediately clear. As power includes economic, political, and social representation, it is possible that the distribution of power at the societal level does not coincide with meaningful improvements at the individual income level. Moreover, it is possible that health and income are complements, such that equitable access to healthcare has positive knock-on effects for perception of income fairness.

In all, the above analysis confirms Hypothesis 1. Gender equality is significantly and positively correlated with income fairness perception, as the overall index and all of the domains exhibit a positive relationship to the likelihood of men and women perceiving their income as fair. That is, as a general tendency in this data, higher gender equality is positively correlated with the probability of perceiving income as fair for all individuals.

3.4.3 Hypothesis 2

To test Hypothesis 2, we examine whether perception of income fairness by men and women converges with increasing gender equality, as predicted by social role theory. To do so, we calculate the probabilities of perceiving one's income as fair for both men and women, as predicted by our logistic regressions. In Figure 3.3 we plot these predicted probabilities of perceiving income as fair for both men (shown with a solid blue line) and women (shown with adotted red line) as well as the share of men (shown with green diamonds) and women (shown with orange squares) who perceive their income as fair in the 24 European countries.

In Figure 3.3, there is a clear trend that increasing gender equality affects the probability of perceiving income as fair positively for both men and women. Furthermore, it appears that there is a larger gap in perception of income fairness between men and women at lower levels of gender equality and that this difference attenuates as gender equality increases. However, based on the results of the adjusted Wald test (see Table 4), we cannot say that these differences are statistically significant. In all, there is no evidence of divergence of gender differences in perception of income fairness; rather, the tendency is toward convergence, although the effect is not significant. For further analysis, we present the results of the six gender equality domains in Figure 3.4.

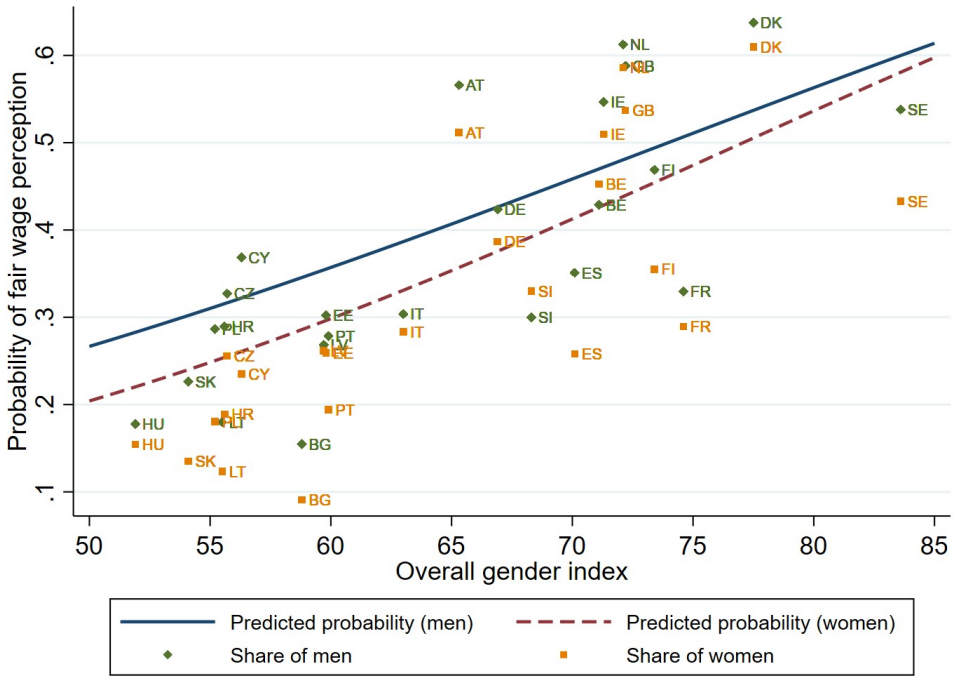


FIGURE 3.3: Predicted probabilities of perceiving income as fair by sex - Overall gender index

Note: shows the predicted probability of perceiving income as fair based on the regression results. The male (female) share indicates the actual proportion of men (women) who reported their income as fair in the particular country.

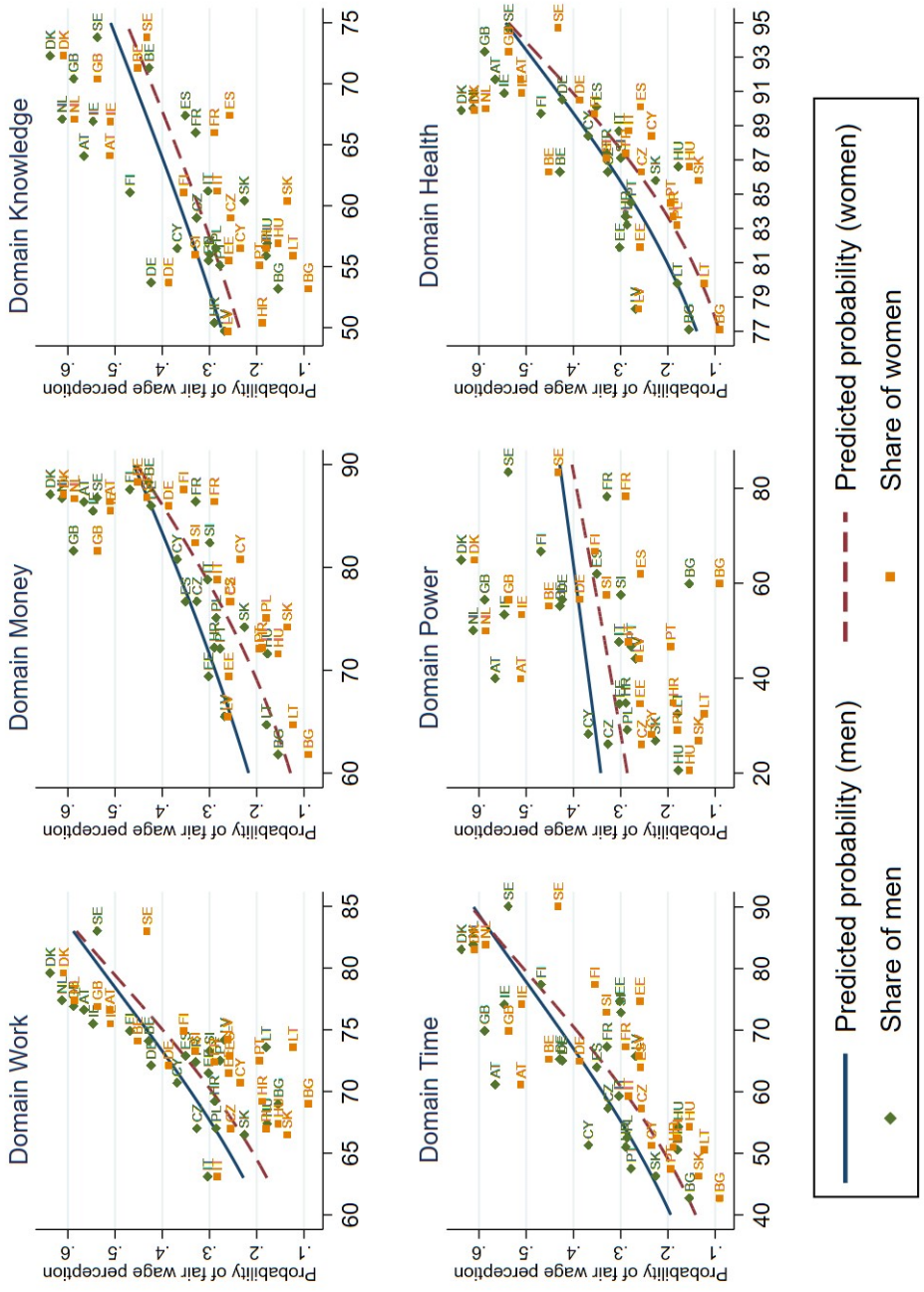


FIGURE 3.4: Predicted probabilities of perceiving income as fair by sex - Gender equality domains

Note: shows the predicted probability of perceiving income as fair based on the parsimonious logistic regressions. The male (female) share indicates the actual proportion of men (women) who reported their income as fair in the particular country.

As discussed previously, the likelihood of perceiving income as fair for both genders is positively related to gender equality in all domains of gender equality, with the strongest effects in the domains of health, work, money, and time. Furthermore, upon visual inspection of Figure 3.4, it appears that the gender difference in perception of income fairness shows a generally attenuating trend in all domains.

However, based on the predicted probabilities by gender and the results of the Wald test shown in Table 5, the convergence is only significant in the money domain, where the results are significant at the five percent level. Therefore, only in this domain can we reject the null hypothesis that the slopes are equal for men and women, and conclude that gender gap in perception of income fairness attenuates at higher levels of gender equality in the money domain. Based on this evidence, we find only limited support for Hypothesis 2.

3.5 Discussion

We find evidence of a strong, positive relationship between gender equality and the likelihood that all individuals perceive their income as fair. This trend holds for the overall gender equality index developed by EIGE as well as every sub-domain of gender equality, with the relationship being strongest in the work, health, and money domains. Put simply: Increasing gender equality (in general and in all domains) is correlated with increasing perceptions of income fairness for both men and women. However, we find only mixed evidence that the gender difference in perception of income fairness attenuates as gender equality increases. Although Figures 3.3 and 3.4 imply that these perceptions are generally converging, the result is only statistically and economically significant in the money domain. However, as the money domains is largely a measure of gender equality in income, this result is an intuitive one. With it, we can tentatively conclude that increasing economic gender equality leads to attenuation in the gender gap in perception of income fairness.

In our analysis of Hypothesis 1, we also find that domestic partnership status and number of children is not statistically significant for men or women. This result is somewhat surprising, as Pfeifer and Stephan (2019) found that the differences in perceiving wage as fair are most significant for married women. Indeed, this result is somewhat surprising in light of sticky household labour division (Kan et al. 2011), where women continue to bear the largest share; and the negative relationship between paid labour and share of household labour (South and Spitze 1994). Our findings imply that domestic status and number of children do not significantly affect likelihood of perceiving income as fair for men or women.

However, we find that age has the expected significant effect, where age generally has a non-linear relationship with perception of wage fairness (see Paul 2006, among others). This implies that, as individuals age into their prime earning years, they are more likely to perceive their income as unfair. After a certain inflection point

later in life, including nearing mandatory retirement age (Lazear 1979), individuals are once again more likely to perceive their income as fair. We find a similar effect in education, where more education increases the likelihood of perceiving income as fair until a point, where those with the highest levels of education are less likely to perceive their income as fair. And, as expected, individuals in higher quintiles of the income distribution are much more likely to perceive their income as fair.

Our weak evidence for Hypothesis 2 indicates that the perception of income fairness by men and women may converge with increasing gender equality. This echoes findings such as Fors Connolly et al. (2020), who, based on a similar theoretical framework, found evidence of convergence within personality traits. Our results are therefore in line with the implications of social role theory and other tests of constructivist theories.

Even where evidence to the contrary (i.e. that gender differences are innate or biologically determined) may be found, such as in Falk and Hermle (2018), there are other explanations for an increase in gender differences with increasing gender equality. First, as gender equality increases, individuals may begin to compare their situation across genders rather than within their own gender (Wood and Eagly 2012, Mayraz et al. 2009). This echoes the reference-point argument in the literature on perceived income fairness (for example Schneck 2014), where women may be more likely to perceive their wage as fair simply because they use other women as their reference point. By this logic, an increase in gender equality could lead women to use men as a reference point, making them less likely to perceive their wage as fair. Or, it is possible that socioeconomic resources give an individual the power to exhibit non-normative behaviours or preferences. Such an explanation could tie in with the existing frameworks that underscore the importance of historically asymmetric institutions, such as those of Sent and van Staveren (2019) and the Empowerment Model in van Staveren (2013). Finally, it is entirely possible that becoming aware of socially constructed norms, like those in social role theory, is part of the self-actualization mechanism.

The most robust and important finding of this paper is that gender equality has a large and significant positive relationship to perception of income fairness for both men and women. This has important practical implications. This finding implies that formal gender equality policies at the national, institutional, or company level could have a direct positive effect for the group in question and indirect positive effects for society at large. As gender equality policies are often difficult to implement, our findings imply that temporary inefficiencies may be outweighed by long-term benefits for all individuals and society as a whole. Indeed, this implication is critical knowledge for policymakers, especially those who argue for the implementation of gender equality measures. As the literature on well-being and income fairness clearly indicates, individuals are more satisfied and happier with a higher perception of income fairness (Clark and Oswald 1997, D'Ambrosio et al. 2018). Therefore,

according to our findings, increasing gender equality is a situation where the “rising tide lifts all boats" and policies to improve gender equality are likely worth the effort.

Our findings underscore the need for future research to synthesize the role of socioeconomic resources and self-actualization within social role theory. We find that a nuanced extension of the theory is likely in order; however, this will require further empirical testing. Although our empirical findings pertain to Europe in particular, the implications are important from an international perspective, as the underlying mechanisms may be at play across countries and cultures. Future research should undertake to research the relationship between gender equality and perception of income fairness across all countries. This would enable a more robust test of the theoretical framework, as an international analysis would provide more meaningful variation in social norms and gender equality policies.

3.6 Conclusion

Based on evidence from European countries, higher gender equality is positively correlated with the likelihood that individuals of all genders perceive their income as fair. In addition, the gender gap in perception of income fairness may attenuate with higher gender equality, at least in the domain of money. This evidence provides support for the constructivist social role theory of gender differences, which implies that gender equality interrupts the self-perpetuation of gender roles, leading to a convergence in preferences and perceptions as gender equality increases. However, we believe that future research should look at international data and formalize the role of resources within social role theory. In addition, our findings have important practical implications, indicating that the social returns to gender equality measures are positive to people of all genders. Therefore, policymakers, companies, and social institutions should continue their efforts to implement policies and practices that encourage gender equality.

Chapter 4

Gender Matters: Market perception of future performance

Abstract

This paper reexamines gender bias in financial markets, focusing on market reactions to CEO announcements. We examine the cumulative abnormal returns around CEO announcements from 1992 through 2016 using an event study. By extending the data and modifying the event window, we contribute to the literature by studying bias over time and firm size. Results indicate that financial markets react more favorably to male CEO announcements, with a cumulative abnormal return (CAR) of 49 basis points above the reaction to their female counterparts. Moreover, we find that market reaction varies over time, which may be due to the increasing proportion of female CEOs; and by firm size, which may be due to differences in new information available to investors. Limitations include a small sample size of female CEO announcements and controls for industry, detailed CEO characteristics, and announcement content.

Keywords: financial markets, event study, female executives, gender bias, token theory

Note: This chapter is based on joint work with Nacasius U. Ujah from South Dakota State University. It was published in 2020 in *Managerial Finance*, Vol. 46 Issue 10, p. 1247 - 1262. The candidate's contribution comprised the project conception, literature review, data and empirical analysis, intepretation of results, and drafting of the paper and revisions. The candidate is the lead author of the published article. The paper is reprinted with the permission of Emerald Publishing.

4.1 Introduction

Female CEOs remain tokens. In the SP 1500, only 0.25 percent of newly announced chief executive officers were female in 1992. Over 20 years later, in 2016, female CEO announcements were still a novel event, comprising just 4.74 percent of newly announced executives. Reasons for this underrepresentation are longstanding and complicated, and even actions to foster equal representation can have negative consequences (Heilman and Welle, 2006). This research synthesizes theories from sociology and finance to determine whether the market reaction to CEO announcement is dependent on CEO gender. According to the efficient-markets hypothesis (EMH), return patterns around a specific event reflect market expectations of future performance given the new information. Token theory (Kanter 1977) demonstrates why market expectations may be biased: the gender of a female CEO may be salient for investors, even if it does not affect future performance. If, as predicted by token theory, gender matters, the market will price the effect of CEO gender in terms of returns.

To examine tokenism, we study stock returns around CEO announcements using event study methodology. In particular, this paper provides novel contributions by performing the first firm size and time event studies in the literature and by instituting a longer event window. This paper consists of three studies. Study 1, a classical event study, implements our novel extended event window. Study 2 is comprised of a five-part event study of bias development over time. In the Study 3, the effect of information is examined through a four-part study of the effect of firm size on market bias. In each study, we analyze a sample of 100 female CEO announcements, matched by firm size and announcement year of 102 male CEO announcements. Data were sourced from EXECUCOMP, COMPUSTAT, CRSP, and LexisNexis for the years 1992 through 2016.

Despite no clear evidence that CEO gender affects firm performance, market reaction to CEO announcement was conditional on gender in Study 1, indicating a biased reaction. Furthermore, in studies two and three, this bias in the market varies over time and by firm size. These findings support the validity of the population-proportion mechanism of token theory; bias should attenuate as the token population grows. Furthermore, the results underscore the importance of the quality and availability of information in mitigating investor bias.

4.2 Literature review

4.2.1 The theory of gender bias

Token theory provides the central theoretical pillar of this research. Kanter (1977) frames the problem of gender bias in the context of minority-majority group dynamics or tokenism. When females are in a token position—one that is rare or novel,

nearly alone—they likely face a variety of double standards, including "role entrapment" which is "the distortion of the characteristics of tokens to fit preexisting generalizations about their category" based on their proportional rarity (Kanter 1977, p. 980). In the token theory framework, gender bias is a function of the proportion of the minority population; therefore, the heavily skewed gender distribution at the CEO level makes investors particularly susceptible to gender bias.

While investors make decisions based on expectations of future earnings, their perception of a female CEO, due only to her paucity, may be warped. In this context, investors may fall back on gender archetypes as heuristics, mentally fitting female CEOs with stereotypical sensibilities, and framing performance expectations upon these biased assumptions. According to token theory, if investors are prone to bias, their perception of female CEO suitability could be divorced from actual performance.

An interdisciplinary body of research examines this phenomonen. Gender bias has been shown to affect the perception of career suitability (Heilman et al. 1989; Gaucher et al. 2011; Uhlmann and Cohen 2005) and performance and behavior (Jamieson 1995; Eagly and Karau 2002; Oakley 2000). Oakley (2000) notes a clear example of Kanter's (1977) role entrapment, finding that male executives "associate the female executives with other women in their lives, such as their wives, daughters, and secretaries" (p. 330). Furthermore, research shows that affirmative action policies do not always align perceptions with reality (Heilman and Welle 2006; Hekman et al. 2017), regardless of the individual's proven ability (Heilman et al.1997).

In the face of heightened visibility and warped perceptions, information is key to mitigating bias (Tosi and Einbender 1985; Swim et al. 1989; Davison and Burke 2000; Heilman 2012). According to Heilman et al. (2012), "The quantity of information that is available to evaluators has been consistently shown to be related to bias ... with limited information facilitating the use of stereotype-based expectations" (p. 118). Within this framework, the amount and quality of information that investors rely on are particularly relevant, as ambiguity exacerbates bias. As the amount of information available to investors is positively correlated with firm size (Shores 1990; Ataise 1985; Christensen et al. 2004), it is likely that investors bias will vary by firm size.

4.2.2 CEOs: Performance and gender

Before moving forward with the study, the effect of CEO gender on future firm performance must be addressed. If CEO gender is a legitimate predictor of future firm performance, then perhaps a market bias is simply an accurate evaluation of future earnings. However, if there is no significant difference in performance, the result underscores the importance of bias measurement.

The effect of gender on CEO ability and firm performance remains an open question (see also Mohan 2014). Powell and Butterfield (1994) find that risk propensity and financial strategy are dependent on gender, and some research finds a negative correlation between female CEOs and corporate risk-taking (Huang and Kisgen 2013; Faccio et al. 2016), but this may simply indicate that male CEO are relatively overconfident (Huang and Kisgen 2013). Despite Amore and Garofalo (2016) finding that female leaders may underperform in highly competitive environments, there is evidence that female leadership is positively related to firm performance (Erhardt et al. 2003; Krishnan and Parsons 2008; Khan and Vieto 2013), and that innovative firms are likely to have more female leaders (Deszo and Ross 2008). Other research finds ambiguous relationships between higher proportions of female leadership and firm performance (Lam et al. 2013; Peni and Vähämaa 2010; Smith et al. 2006; Shrader et al. 1997) or no significant difference in financial decision-making when controlling for other constraints (Atkinson et al. 2003). An additional summary of this literature can be found in Appendix B.1.

Endogeneity problems further complicate the relationship between CEO gender and firm performance. There is mixed evidence for the existence of a "Glass Cliff" phenomenon, which points to the idea that female CEOs are more likely to be appointed in precarious situations (Ryan and Haslam 2005; Haslam and Ryan 2008; Adams et al. 2009; Cook and Glass 2014; Elsaid and Ursel 2018). In contrast, Elsaid and Ursel (2018) find that avoidance of the negative publicity that would accompany the termination of female CEOs may actually increase female CEO tenure.

Based on the synthesis of the interdisciplinary literature, this research uses the working assumption that any perceived difference of suitability based on gender is a social construct at best (see also Eagly and Wood 1991 and Gneezy et al. 2009); in any case, there is insufficient evidence to conclude that CEO gender has predictive power for future firm performance. Therefore, a systematic difference in investor preferences (for male- or female-led firms) may be the result of bias.

4.2.3 Measuring gender bias

As CEO gender has no reliable power to predict future firm performance, it follows that significant pricing differences would result from exogenous factors, like bias. However, measuring bias poses several econometric challenges. Unlike studies of returns over time (see Wolfers 2006 and Kolev 2012), event study methodology overcomes many of these challenges. Introduced by Fama et al. (1969), event studies allow researchers to control for general market returns and risk factors in order to isolate the incremental effect of a locally unique event, like a CEO announcement (Beatty and Zajac 1987). The EMH underpins this methodology: assuming that markets make use of all available information, efficient markets imply that there are no undervalued stocks; therefore, investors cannot beat the market. The return patterns around an event reflect collective market reaction to incremental amounts of

new information; that is, returns indicate whether markets perceive the information as good news or bad news for future earnings.

In an event study also based on token theory, Lee and James (2007) show that while the market reacts negatively to female CEO announcements, the rarity of female announcement events makes it difficult for the authors to generalize conclusions. They also indicate that future research should address bias development as the female CEO population grows over time (2007). Lucy and Carron (2011) apply the methodology from Lee and James 2007 to the FTSE 100, and in doing so, find that the "stock market reacts against females appointed to the male-dominated roles of executive director and CEO, but that the role of nonexecutive director is regarded as suitable for either gender" (p. 1229). Cook and Glass (2011), however, find conservative evidence that markets react slightly positively to female executive announcements; they also find that the effect is stronger in female-dominated industries. Similarly, in line with token theory, Cook and Glass (2011) conclude that market reaction to female executive announcements is adverse in male-dominated industries, stating, "the strength of the market's reaction to women leaders diminishes as the industry becomes more male-dominated" (p. 513). These findings support the token theory prediction that bias should diminish (grow) as the token population proportionally increases (decreases).

However, Gondhalekar and Dalmia (2007) and Martin et al. (2009) find no statistically significant difference. Using a 40-day window to measure abnormal returns, Coxbill et al. (2009) also find no difference between male and female CEO announcements. See Appendix B.2 for a summary of the event study literature.

4.2.4 Contribution

Contribution Studying implicit bias is inherently complex; measuring collective bias in financial markets is even more so. Both the general body of literature and, more specifically, event studies, are currently inconclusive about the interaction between CEO gender, CEO ability, and market reaction. This paper includes three contributions in determining whether the market perception of future performance is dependent on CEO gender.

This paper makes two contributions the existing empirical literature. At the conclusion of their study, Lee and James (2007) indicate that measuring bias development over time (as the proportion of female CEOs increases) is essential for future research. To the authors knowledge, this is the first paper that studies bias development over time, providing a critically needed test of the population-proportion mechanism from token theory. Similarly, the effect of firm size, which generally correlates to the amount and quality of bias-mitigating information, was previously unexplored.

In addition to these contributions, this paper posits that the traditional three-day window used for cumulative abnormal returns is insufficient (see also Coxbill et al. 2009). It is likely that the market hears word of the CEO change before the announcement: in this sample, the mean time between CEO announcement and CEO appointment is significantly (at the 99 percent level) longer for female CEOs (77 days) than for male CEOs (54 days). This indicates that a longer event window is appropriate, as the market will absorb more information during this waiting period; therefore, the event window comprises a 26-day period in study one. In order to better compare new results with the existing evidence, the time and firm size studies utilize the traditional 3-day event window. With these contributions in mind, this research studies whether financial market reaction to CEO announcement is conditional on CEO gender.

4.3 Methodology and data

4.3.1 Event studies

Conceptually, event studies rely on measures of firm-level expected and actual returns. Expected returns, also known as normal returns, describe a counterfactual: the returns a firm would have achieved had the analyzed event not taken place; thus, a systematic difference between actual returns and expected returns reflects the event. This difference is, by assumption, abnormal (because CEO gender should not matter in this way). Aggregated across firms and around CEO-announcement events, significantly positive or negative abnormal returns identify systematic biases favoring male or female CEOs, respectively.

More formally, to estimate expected return and, thus, abnormal return—the difference between actual and expected return—this paper estimates a Fama-French three-factor model. Equation 4.1 is the sample regression function associated with this model.

$$R_{RiskModel,i,t} = R_{f,t} + \hat{\alpha}_i + \hat{\beta}_{1t}\left(R_{m,t} - R_{f,t}\right) + \hat{\beta}_{2t}SMB_t + \hat{\beta}_{3t}HML_t + \hat{\epsilon}_{i,t} \tag{4.1}$$

where $R_{RiskModel,i,t}$ is the actual daily return of the firm, $R_{f,t}$ is the risk-free rate in the market, $\hat{\alpha}$ is the intercept term, $\hat{\beta}_{1t}$ is the part of the systematic risk that captures the market premium, $\hat{\beta}_{2t}$ is the part of the systematic risk that captures the excess return of small over big stocks, $\hat{\beta}_{3t}$ is the part of the systematic risk that captures the excess return of stocks with high market-to-book ratios over stocks with a low market-to-book ratio. The estimated residual is represented by $\hat{\epsilon}_{i,t}$. A notational hat denotes an estimated value. Equation 4.1 is used to estimate Equation 4.2, which specifies the expected return $\hat{R}_{RiskModel,i,t}$ for firm i.

$$\hat{R}_{RiskModel,i,t} = R_{f,t} + \hat{\alpha}_i + \hat{\beta}_{1t}\left(R_{m,t} - R_{f,t}\right) + \hat{\beta}_{2t}SMB_t + \hat{\beta}_{3t}HML_t \quad (4.2)$$

The abnormal return for firm i on day t within the event window is the difference between the actual stock return on that day $R_{RiskModel,i,t}$ and the corresponding expected (or normal) return $\hat{R}_{RiskModel,i,t}$. As such, the abnormal return is specified as Equation 4.3.

$$AR_{i,t} = R_{RiskModel,i,t} - \hat{R}_{RiskModel,i,t} \quad (4.3)$$

To estimate the risk model surrounding the announcement dates, the study specifies the length of the estimation period in trading days over which the risk model is estimated (Equation 4.1). The estimation period is 250 calendar days, with at least 200 days of trading activity. There is also a gap period of 20 days before the event window, for a total of 270 days. To measure the total impact of an event over a particular period of time—the event window—the abnormal returns for the 26 (3) days in the event window are summed. Cumulative abnormal returns as are measured as specified in Equation 4.4.

$$CAR_i = \sum_{t=1}^{T} AR_{i,t} \quad (4.4)$$

The cumulative abnormal return (CAR) is computed for each CEO announcement by either year, firm size, or one pooled group. Finally, using a t-test, mean CARs are compared to determine whether the mean CAR for male-CEO announcements is significantly different than the mean CAR for female-CEO announcements in each case. For this mean test, CAR is measured for each group as specified in Equation 4.5 and the mean CAR is calculated over the event window.

$$CAR = \frac{\sum_{i=1}^{N} AR_i}{N} \quad (4.5)$$

4.3.2 Data

All executive-related data comes from EXECUCOMP (2017). Securities information is from the Center for Research in Security Pricing (CRSP 2020), as administered through EVNTSTUDY software, and firm financials come from COMPUSTAT (2020). Also, CEO announcement dates were gathered through searches of LexisNexis.

Data comes from the full EXECUCOMP database of 273,478 observations. In order to study firms with relatively similar market structures, this study removes

Panel 1. Mean Firm Financial Information	Male	Female
Total Assets	6231.39	7364.531
Return on Assets	0.029	0.029
Tobin's Q	2.147	2.309
Leverage Ratio	0.224	0.199
Number of CEOs	102	100
Panel 2. Mean Executive Information	**Male**	**Female**
Age	54.10	53.60
Salary	778.47	772.58
Bonus	303.88	234.48
Total compensation	5269.16	5259.46

TABLE 4.1: Male and female characteristics
Note: reports summary statistics of firms led by male and female CEOs for years 1992 through 2016.

utility firms (SIC codes 4900-4999) and financial firms (SIC codes 6000-6999), resulting in 228,048 executive observations. As EXECUCOMP includes data for many top management positions, irrelevant non-CEO observations are removed, leaving 36,249 executive observations.

In constructing the female sample, this study removes all male executives and their respective observations. Annual observations are filtered so that each executive-company combination is represented once. This results in a list of 156 female CEOs and their respective data. Finally, CEOs announced before 1992 are removed because of incomplete or inadequate information. Of the 152 female CEOs used for the event study, 52 had inadequate data to complete the study, resulting in a final female sample size of 100 female CEO announcements.

Similar to Cook and Glass (2011), this paper uses a random matching methodology to construct the male sample. From the pool of EXECUCOMP CEO data, the 869 female observations are removed, resulting in 35,560 male CEOs observations. As the analysis showed there is no significant difference in results between propensity score matching methodology and random matching methodology, this study used a randomly selected male CEO appointed in the same year for each female CEO announcement. Doing so resulted in a random matched sample based on appointment year. Table 4.1 demonstrates the similarity between observations in the matched sample. While female-led firms are slightly larger in this sample, return on assets (ROA), Tobin's Q, and leverage ratio are very similar for both gender samples. Furthermore, executive data shows that male and female executives have similar age, salary, and total compensation. As with the female group, some male observations did not have adequate information to complete the event study. The final male matched sample size is 102 CEO announcements.

4.4 Results

4.4.1 Study 1 – Full sample event study

Study 1 uses classical event study techniques with the innovation of an extended event window. To test whether market reaction is conditional on CEO announcement for the sample as a whole, all female and matched male CEO announcements for the period of 1992 through 2016 were analyzed. Results indicate that financial market reaction to CEO announcement is conditional on CEO gender in Study 1. As can be seen in Table 4.2, the CAR for female CEO announcements shows the market reacted negatively, here, a decrease by about 15 basis points within the 26 days' event window. However, matched male CEO announcement garnered positive sentiment in the market, as the market reacted with a 34 basis point increase in CAR. As seen in Table 4.2, a mean test of the CAR for both genders indicates that the difference is statistically significant at the 99 percent level with a difference of male CEO to female CEO announcement of almost 50 basis points. See Appendix B.4 for a graphical representation of return patterns for the entire window.

Female		Male			Difference	
mean CAR	mean AR	mean CAR	mean AR	n =	CAR	AR
-0.0015	0.0000	0.0034	0.0000	202	0.0049 ***	0.0004
					(0.0016)	(0.0008)

(Standard errors shown in parentheses) $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE 4.2: Study 1 results - Pooled study

Note: shows the CARs for female CEOs vs their counterpart male CEOs using a pooled event study methodology. A sample of 100 female CEO announcements is compared with a matched sample of 102 male CEO announcement for the period of 1992 through 2016. CARs are estimated using the event study methodology by applying the Fama–French three-factor model and a 26-day window. We perform a mean test and estimating the Welch degrees of freedom. Standard errors are in parentheses. ***, ** and * denote significantly different from zero at the 1, 5 and 10 percent level, respectively.

These results indicate that the market reacts differently to female CEO announcements; moreover, the market reacts negatively to females relative to males. Although the extant literature indicates that CEO gender is not a reliable predictor of future performance, markets react conclusively and differently to male and female CEO appointments in Study 1.

4.4.2 Study 2 - Event study over time

In Study 2, we examine market reaction over time. To do so, time periods with notable economic or social characteristics were generated, as these could affect the likelihood of female CEO announcement and general market sentiment at the time.

The time periods are: vanguard announcements (1992-1995); pre-millennium (1996-1999) and post-millennium (2000-2005) announcements; announcements in the financial crisis years (2006-2010); and announcements in the post-crisis growth years (2011-2016).

Similar to Study 1, the findings documented in Table 2.3 indicate that overall market reaction is conditional on CEO gender. The difference between male mean CAR and female mean CAR is significant during all periods but one. However, the direction of the difference varies. See Table 2.3 and Figure 4.1. Results are robust to different time period specifications; these results, as well as detailed returns (such as Appendix B.4 for Study 1), are available upon request.

	Female		Male			Difference	
Years	mean CAR	mean AR	mean CAR	mean AR	n =	CAR	AR
1992-1995	-0.0033	0. 0050	-0.0302	0.0153	11	0.0269**	0.0203
						(0.0121)	(0.0139)
1996-1999	-0.0043	-0.0016	0.0091	0.0045	16	-0.0134***	-0.0060
						(0.0042)	(0.0059)
2000-2005	-0.0011	-0.0021	0.0112	0.0049	44	-0.0123***	-0.0070
						(0.0036)	(0.0045)
2006-2010	-0.0038	-0.0001	-0.0091	-0.0043	51	0. 0053	0.0042
						(0.0040)	(0.0048)
2011-2016	-0.0085	-0.0023	-0.0022	-0.0005	52	-0.0062***	0.0019
						(0.0025)	(0.0046)

(Standard errors shown in parentheses) $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE 4.3: Study 2 results - Year effect

Note: This reports event-study test female CEO announcements compared with a matched sample for male CEO announcement for five grouped periods between 1992 and 2016. CARs are estimated using the event study methodology by applying the Fama–French three-factor model and a three-day window. We perform a mean test and estimating the Welch degrees of freedom. Standard errors are in parentheses. ***, ** and * denote significantly different from zero at the 1, 5 and 10 percent level, respectively

The market reacts negatively to female CEO announcements in all periods. In period 1, 1992 to 1995, both male and female CEO announcements result in negative cumulative abnormal returns over the 26-day window. The female mean CAR of negative 33 basis points shows a smaller reactionary effect than male mean CAR (equaling negative 302 basis points). In period 2, 1996 to 1999, the market reacts positively to male announcements and negatively to female announcements. In absolute terms, the positive change in male mean CAR from the first to the second period is 393 basis points, while the negative reaction to female announcements remains relatively unchanged. In the first and second periods, male CEO announcements result in large abnormal movements in stock price, while female CEO announcements have smaller effects. This trend continues in the third period, where the negative reaction to female CEO announcements attenuates by approximately 30 basis points.

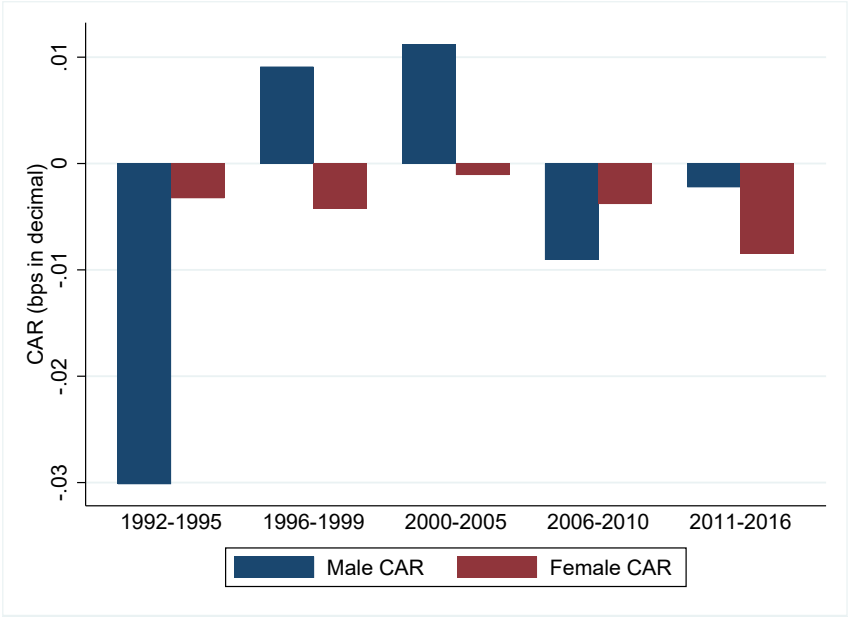


FIGURE 4.1: Market reaction to CEO gender over time

In the fourth period, 2006 to 2010, there is no significant difference between CEO genders. This is not an unexpected result, as the financial crisis during this period may have overshadowed the effect of CEO gender on market reaction. In the fifth period, 2011-2016, female CEO announcements again result in negative CAR relative to male announcements.

4.4.3 Study 3 - Event study by firm size

Empirically, the extant literature documents that firm size does matter for firm performance in general: see Baker and Hall (2004), Moeller et al. (2004), and Lee (2009). However, to our knowledge, none has specifically considered this effect on CEO announcements. In using a modified market model like Fama-French 3 factor model, the presumption may be that the model intrinsically controls for size. This research goes further, decomposing the female and matched male CEO samples into four groups based on capitalization. First, the largest firms are those firms that belong within the S&P 500 (SP) as coded by EXECUCOMP. The second group contains firms that belong to the Mid-Cap group (MD). The third is made up of firms that are in the Small-Cap group (SM). Lastly, firms that do not belong to any of the groups mentioned above are grouped as (EX). Based on these groups, results indicate that Market reaction to CEO gender varies with firm size. See Table 4.4 and Figure 4.2. Detailed returns (such as Appendix B.1 for Study 1) are available upon request.

Similar to the analysis of bias over time, the effect of gender bias varies with firm size. The difference in mean CAR between genders is significant only for the Small-Cap (SM) 600 firms. For these relatively small firms, male CEO announcements are greeted with a 88 basis point mean increase, while female CEO announcements result in a mean decrease of 80 basis points. Reactions to firms that are not part of

	Female		Male			Difference	
Index	mean CAR	mean AR	mean CAR	mean AR	n =	CAR	AR
SP	-0.0064	-0.0013	-0.0028	0.0001	47	-0.0036	-0.0014
						(0.0024)	(0.0050)
MD	0.0020	0.0025	0.0011	-0.0014	16	0.0009	0.0039
						(0.0039)	(0.0043)
SM	-0.0080	-0.0010	0.0088	0.0030	44	-0.0169***	-0.0040
						(0.0039)	(0.0088)
EX	-0.0017	-0.0012	-0.0002	0.0006	77	-0.0015	-0.0019
						(0.0042)	(0.0064)

(Standard errors shown in parentheses) $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE 4.4: Study 3 results - Size effect

Note: reports female CEO announcements compared with a matched sample for male CEO announcement for index between 1992 and 2016. CARs are estimated using the event study methodology by applying the Fama–French three-factor model and a three-day window.Weperform a mean test and estimating the Welch degrees of freedom. Standard errors are in parentheses. ***, ** and * denote significantly different from zero at the 1, 5 and 10 percent level, respectively.

the S&P 1500 index (EX firms) are close to zero for both male and female announcements. The market reaction to both male and female announcements for MidCap 400 (MD) is weakly positive, while the reaction to both in the S&P 500 (SP) is negative. However, these differences in market reaction are not significant.

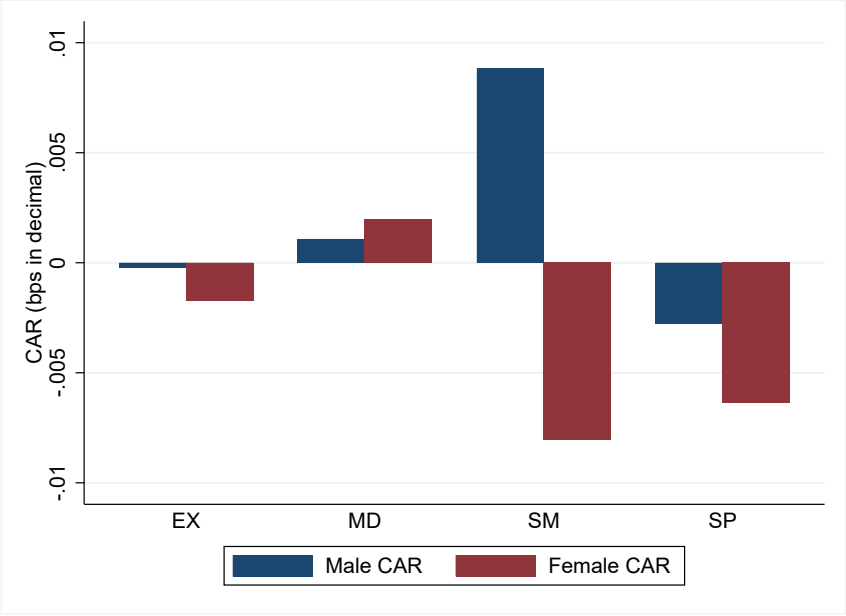


FIGURE 4.2: Market reaction to CEO gender by firm size

4.5 Discussion and conclusion

The findings of Study 1 indicate that financial markets react negatively to female CEO announcements compared to male CEO announcements, in line with the findings of Lee and James (2007) and Lucey and Carron (2011), and in contradiction to Gondhalekar and Dalmia (2007), Martin et al. (2009), and Coxbill et al. (2009). The extended event window and relatively large sample size could play a role in explaining this result. Although the extended event window captures the waiting period effects (as discussed in the contribution section), it also increases the likelihood of capturing other events that affect future firm performance, such as mergers or earnings announcements, that may occur within the extended window. This limits the interpretation of the results as measurement of gender bias, as such confounding events are not controlled for in the study.

The results of Studies 2 and 3 provide some new evidence for the application of token theory to bias in financial markets. Lee and James (2007) note that studies over time are crucial to understanding the bias mechanism in financial markets, and conclude that "stock market reactions or other firm performance indicators, as the number of women in senior leadership positions increases" remain regrettably unexamined (p. 239). To the authors' knowledge, Study 2 is the first explicit study of this development. The findings shed light on the bias mechanism: as female CEOs grow as a proportion of the CEO population, the difference in market reaction to male and female CEO announcements shrinks over time. This trend continues up to the financial crisis years; thereafter, the trend is unclear. These results appear to support the predictions of token theory; as the token population grows, the bias should diminish (Kanter 1977; Karsten 1994; Lee and James 2007). The results of Study 2 provide the first notable test of the population-proportion mechanism over time; however, due to the current paucity of female CEO announcements, future research is critically needed to determine a reliable trend over time.

Furthermore, the finding that firm size affects market reaction to CEO gender has implications for the information effect literature (Tosi and Einbender 1985; Swim et al. 1989; Davison and Burke 2000; and Heilmann 2012). As firm size is positively correlated with information (Shores 1990; Ataise 1985; Christensen et al. 2004), larger firms likely have highly publicized corporate announcements and more information is available to investors. For small firms, announcements may take the form of one-line press releases with little more than the female name. With so little information, the market is likely to rely on stereotypes (Heilman 2012), which could explain the adverse reaction to female CEO announcements for SmallCap 600 firms. However, the reaction to other firm size categories are not significant. Further research is necessary to determine the effect of both content and length of announcements on market reaction.

Limitations of this work include those of previous research: the scarcity of female CEOs results in small sample sizes of female CEO announcements. Furthermore, there are a variety of factors that should be explicitly examined in future research, including industry, CEO education, differing leadership styles, announcement content, and possible Glass Cliff scenarios. In addition, the extended event window utilized here should be improved in the future by controlling for confounding events. Although gender bias remains inherently challenging to study, there remain promising opportunities for future research.

Findings indicate that markets react differently to male and female CEO announcements; that is, market expectations of firm performance are dependent on CEO gender, and this bias varies over time and by firm size. Token theory explains why this phenomenon exists; the market reacts differently to female CEOs based on the salience of their minority, and the availability of information affects investor perception. However, token theory also offers possible solutions. Karsten suggests that to shed this token status, a minority group must grow to at least 35-40 percent of the total population (1994). Further research on this topic could elucidate how best to approach the challenge.

Chapter 5

The Impact of Narrative Sentiment on Market Reaction to New CEO Announcements

Abstract

This study investigates whether investors react to sentiment in new CEO announcements. Using sentiment analysis, I find that, on average, firms craft a positive sentiment in CEO announcements, even following an forced turnover. This supports the voluntary disclosure theory from Dye (2001), which implies that firms tend to emphasize favorable information. Moreover, regression results indicate that investors react to positive sentiment with an average 354 basis point higher abnormal return compared to neutral sentiment. In the specific context of CEO announcements, where the two mechanisms for verifying sentiment are notably absent, the results suggest that investors may react to sentiment for its own sake. These findings have implications for market efficiency, as companies could employ sentiment strategically, influencing investors and generating noise in financial markets. However, this research is not able to determine to what extent the significant reaction is attributable to a behavioral bias, or whether sentiment can be verified in other ways. Other notable limitations include the inablity to observe market dynamics such as strategic noise and timing of announcements.

Keywords: CEO announcements, financial markets, sentiment analysis, voluntary disclosures

Note: This chapter is based on independent (single-author) work. The research is yet unpublished. I wish to particularly thank Jörg Schiller, Casey Rothschild, Alfonso Sousa-Poza, Joe Santos, Tim Philippi, Andreas Blickle, and participants of the Hohenheim Finance and Economics Brown Bag Workshops for their helpful comments. I also wish to thank Felix Sandern and Daniel Kastl for their data collection efforts; the 2021 Hohenheim DALAHO Data Grant for providing the funds for a subscription to EXECUCOMP; and Prof. Daniel Hoang for his help in making BoardEx available.

5.1 Introduction

There is a wealth of empirical research indicating that investors update their expectations of future firm performance when a new CEO is announced. At the same time, sentiment analysis of other types of corporate narratives, such as earnings press releases (Huang at al. 2014, Henry 2008) and annual reports (Yekini et al. 2016), provide initial evidence that investors react to the overall emotional impression, or sentiment, of the narrative. However, there is a notable gap in the literature: to my knowledge, the sentiment of new CEO announcements—and whether this sentiment matters to investors—remains unexamined. Researching this gap is important because existing studies focus on narratives in the presence of at least one verification mechanism: the simultaneous release of financial figures or formal third-party scrutiny (Healy and Palepu 2001). Neither mechanism is present in the case of CEO announcements. Rather, CEO announcements represent purely qualitative and arguably speculative narratives about future firm performance under new leadership.

By definition, a new CEO announcement is the formal, public announcement of new top leadership in a company. By extension, a CEO announcement is also a signal of the company's future direction. According to Park and Berger (2004), "CEOs may literally and symbolically *be* the organization to stakeholders" (p. 93, see also Pincus et al. 1991). As CEOs are tantamount to company image, the announcement of a new CEO is a major public relations event, providing an opportunity for the company to publish a cohesive narrative of future leadership and performance.

Indeed, the new CEO announcements examined in this research are *voluntary* disclosures. Management exercises discretion over how and when to make a formal announcement to the public. As a result, announcements are crafted with great care and attention, and content and sentiment varies depending on the company and the situation. The briefest announcements include company name and the names of the incoming and outgoing CEOs, while longer announcements also contain biographical information and interviews with executives–which often provide optimistic speculation about future firm performance.

Given their nature as voluntary disclosures and given the lack of verification mechanisms in the institutional setting, new CEO announcements present a unique opportunity to examine whether investors react to unverifiable sentiment—or, in other words, whether investors react to sentiment for its own sake. My research is motivated by the following research question:

Do markets react to the sentiment of new CEO announcements?

Answering this research question makes an important contribution, both theoretically and empirically. First, this research provides the first evidence on whether the voluntary disclosure theory from Dye (2001) and the impression management literature (Boudt and Thewissen 2019) hold for CEO announcements. These theories posit that firms disclose favorable information and omit unfavorable information. Second, it provides the first empirical results on the extent to which investors react to sentiment in CEO announcements, which is particularly important because sentiment is unverifiable in this setting.

The scope of this research is limited to examining whether investors react to sentiment in CEO announcements. Therefore, the primary empirical analysis focuses on variables gleaned from the announcement texts. Although the related literature on CEO succession indicates that firms engage in other strategic activities around CEO announcements, such

as releasing other information to generate noise around announcements (see Graffin et al. 2011) and timing announcements around the activities of rival companies (see Burchard et al. 2021), addressing these dynamics remain out of the scope of this paper.

I am, however, able to control for the firm-specific reason for the CEO turnover event. Using the CEO turnover dataset from Gentry et al. (2021), I capture the background situation surrounding the new CEO announcement. For example, my data differentiates between a standard retirement of the previous CEO and an involuntary turnover due to poor firm performance. Therefore, although I am not able to control for the larger market dynamics mentioned above, I am able to control for the critical underlying variable of why the firm is announcing a new CEO to begin with, which is the backdrop against which investors update their expectations.

Based on a sample of 738 CEO announcements in the S&P 1500 between 2010 and 2020, I find that, on average, firms craft a positive sentiment in their CEO announcements. This finding holds even for announcements following an involuntary turnover due to poor performance or personal scandal. These results imply that, on average, firms tend to employ a positive sentiment when writing CEO announcements—and this strategy pays off in terms of abnormal returns. This provides support for the voluntary disclosure theory from Dye (2001) and the impression management literature (see Boudt and Thewissen 2019), which posit that firms intentionally emphasize positive information and downplay unfavorable information. Given the variation in the sentiment variable, it is likely that firms attempt to maximize the positivity of the announcement given the constraints of the objective facts surrounding the CEO turnover.

Furthermore, I find evidence that sentiment has a positive and significant influence on market reaction, with positive announcements receiving, on average, 354 basis points higher abnormal returns than neutral announcements. The interpretation of this market reaction, however, remains open. My reseach is not able to determine whether the reaction to sentiment is primarily behavioral bias, whether sentiment may be quasi-verifiable based on other substantive information in the CEO announcement, or whether unverifiable sentiment is a correlated with firm performance (akin to Spence's job market signaling (1973)). Therefore, although future research should clarify the precise informativeness of unverifiable sentiment, the initial implications of this research are important: if firms realize that narrative sentiment is an opportunity to strategically influence investors, this relationship could present an ongoing and worsening dynamic.

5.2 Theoretical framework and empirical evidence

5.2.1 Sentiment in corporate communications

Publicly traded companies provide information about firm performance to investors and the public through corporate communications (for an introduction to the field of corporate communications, see Argenti 1996). According to Yekini et al. (2016), within corporate communications, there are two primary types of information. The first type is quantitative information. Companies are required to publish quantitative information, such as audited financial statements, at regular intervals. By their nature as regulated financial figures, quantitative information represents an arguably objective measure of the financial position of the company, what Demers and Vega (2014) and Engelberg (2008) call "hard information." Analysts

use these figures to make investment decisions based on their assessments of future firm performance. In addition to quantitative information, companies publish qualitative information via corporate narratives. Examples of corporate narratives are letters to shareholders, earnings press releases, annual report narratives, and new CEO announcements (see, for example, Yekini et al. (2016) and Loughran and McDonald 2011). When quantitative and qualitative information are published simultaneously—which is often the case—the role of qualitative information is to narrate and contextualize quantitative information (Yekini et al. 2016).

Within corporate narratives, there are two sub-types of qualitative or "soft" information (Demers and Vega 2014, Engelberg 2008). First, substantive content includes discussion of past events, contextualization of audited figures, speculation about future performance, biographical information about executive, etc. Sentiment, which is distinct from substantive information, represents the overall tone in which the substantive content is conveyed (Demers and Vega 2014). In the context of this research, sentiment is the noticeable and measurable level of positivity or optimism in a corporate narrative. Adapting a common phrase: substantive content is what you say, and sentiment is how you say it.

Voluntary disclosures and narrative sentiment

In contrast to compulsory publication of financial figures, corporate narratives are voluntary disclosures. As such, narratives are generally not subject to strict regulatory oversight; narrative content and sentiment is at management's discretion. Healy and Palepu (2001) provide a valuable summary of the literature on voluntary disclosures, noting the existence of information asymmetries and conflicts of interest, which are generally to the advantage of firm insiders. One of the most widely cited theories from Dye (2001) indicates that managers have an information advantage over investors, and will only disclose information that is favorable to them and omit information that is unfavorable to them. In the context of this research, Dye's theory implies that firms will craft as positive a message as possible in their corporate narratives.

This relationship is echoed in the corporate communications literature, where evidence indicates that managers often use sentiment strategically, obfuscating negative events and highlighting positive ones in order to present a positive company image, a practice called impression management (Patelli and Pedrini 2014, Boudt and Thewissen 2019). Yuthas et al. (2002) succinctly summarize impression management as follows: "[Impression management] is designed to influence the decisions of a rational opponent." (p. 144). In impression management, the goal of corporate communication is not to sincerely mitigate information asymmetries, but to strategically manage the image of the firm and its performance to investors and the public.

Like other corporate narratives, CEO announcements are voluntary disclosures crafted at the discretion of management. However, as the formal announcement of new company leadership, CEO announcements represent a particularly important vehicle for a company to signal its future direction; for example, a passing of the torch to the next generation, or a necessary change in direction after poor firm performance. The announcement of a CEO is a moment where the public image is critical (see Park and Berger 2004), and, therefore, impression management is particularly likely. And, as voluntary disclosures, content and sentiment are at the discretion of the firm. As a regulatory measure, announcements are

simply required to end with a "forward looking statement disclaimer," indicating that managerial statements are valid only at of the time of publication. This unique setting indicates that CEO announcements are, even more so than other types of corporate narratives, a likely vehicle for impression management.

Based on the theory from Dye (2001) and the voluntary disclosure literature, as well as the particular setting of CEO announcements as an opportunity for companies to signal their future direction, I develop a prediction for the sentiment analysis of CEO announcements:

H1: CEO announcements will, on average, have a positive sentiment.

5.2.2 Narrative sentiment and stock prices

The ultimate goal of this research is to examine whether markets react to narrative sentiment in the context of CEO announcements. This is a particularly interesting research question, as Price et al. (2012) note that sophisticated investors should expect a positive skew to sentiment in voluntary disclosures and disregard sentiment. To explain the mechanism by which narrative sentiment could affect stock prices, I take the Efficient Markts Hypothesis (EMH), which underpins the modern understanding of financial markets, as a starting point. EMH posits that, when markets are efficient, a particular stock price at any given time is a perfect reflection of all information available to investors at that time (Fama 1970). Investors' expectations of future firm performance are constantly and almost instantaneously evolving as new information is released (Malkiel 2005). The release of novel, price-relevant information will result in a change in stock price as investors update their expectations about future firm performance; this mechanism underpins the method of an event-study (Fama et al. 1969). The implication of EMH for this context is as follows: if sentiment is perceived as price-relevant information to investors, it will be reflected in stock prices. If sentiment is not perceived as price-relevant information, it will not be reflected in stock prices.

Theoretical underpinnings

Whether sentiment is expected to be price-relevant depends on the institutional setting and the assumptions of investor rationality. In a straightforward interpretation of the standard model, Price et al. (2012) assert that rational investors understand the information asymmetries and incentives involved in voluntary disclosures. Expecting the positive skew, investors disregard sentiment information, as it cannot be a credible signal of future firm performance. However, other frameworks indicate that, under certain conditions, voluntary disclosures and "soft information" like sentiment can provide price-relevant information.

In a widely cited example, Demers and Vega (2012) note that financial figures, published in tandem with narratives, enable investors to verify sentiment. This is essentially an extension of the strategic communication theory from Benabou and Laroque (1992), where verifiability leads to truthful revelation as an equilibrium. When truthful sentiment is the expectation, sentiment becomes credible and therefore price-relevant to investors. In addition to verification via financial figures, Healy and Palepu note that third-party intermediaries, such as regulatory agencies and financial analysts, can also verify voluntary disclosures, rendering sentiment credible (2001). Without verification, sentiment is merely "cheap talk," because, ex post, investors cannot discern whether a mistaken prediction was due to honest uncertainty or dishonest manipulation (Demers and Vega 2014, p.4).

In the particular setting of CEO announcements, the mechanisms for verifying sentiment are notably absent. In general, there is no simultaneous publication of quantitative information, as is the case for other narratives, like annual reports or quarterly earnings. Furthermore, the narratives are not subject to rigorous regulation. With regard to the theory of voluntary disclosure and problems of verifiability, CEO announcements are a perfect storm: announcements are composed and published at management's discretion and subject to little third-party scrutiny; moreover, there are no accompanying financial figures against which to verify the sincerity of the narrative.

According to this framework, sophisticated investors understand that firms are incentivized and able take advantage of the public relations opportunity that the announcement presents. As a result, this framework implies that investors are, at the very least, highly skeptical of sentiment in such a setting. Therefore, announcement sentiment will not be reflected in stock prices. Based on this reasoning, I formulate a prediction as follows:

H2: Announcement sentiment is not significantly correlated with returns.

Empirical evidence

The theoretical framework implies that sentiment will affect stock prices when investors can verify sentiment, i.e. when it provides reliable information. In an apparent contrast to this, empirical evidence indicates that sentiment in other corporate narratives does influence stock prices. However, nearly all of the extant literature examines sentiment in the presence of at least one of the mechanisms for credibility: the narratives either accompany financial figures or are scrutinized by third-parties, or both. Therefore, most of the existing evidence on market reaction to sentiment is more accurately called evidence on market reaction to verifiable, and therefore, credible sentiment.

For example, Price et al. (2012) analyze the overall sentiment of the unscripted question-and-answer segment of earnings conference calls. Their findings indicate that there is a positive and significant relationship between sentiment and abnormal returns. Interestingly, the authors also find that sentiment is a stronger predictor of abnormal returns for firms who have more uncertainty in cash flows (non-dividend paying firms), indicating that investors relied more heavily on sentiment information in the face of higher uncertainty (p. 994). However, the Price et al. (2012) explicitly discuss sentiment credibility, noting that the active participation of third-parties in the Q&A session assures that sentiment is (at least somewhat) credible.

Yekini et al. (2016) examine market reactions to sentiment in annual report narratives for UK companies. The authors propose a cognitive processing-based model which formalizes how sentiment can affect stock prices. In this model, investors process sentiment as they read narratives; the repetition of positive (or negative) words incrementally changes cognitive structures, ultimately resulting in a changed attitude toward the initial position (i.e. updating of expectations about future firm performance). The authors do not address the role of credibility in their model; they simply indicate that investors must be "attentive" to the information (p. 418). However, in their explanation of the institutional background, the authors note that the UK has a Financial Reporting Council which monitors the transparency of corporate disclosures (although managers still exercise discretion over sentiment and content), which may lead investors to believe that sentiment is credible. The authors find that market returns are strongly and positively correlated with sentiment, concluding

"[A]lthough many claim that annual report narratives may have the tendency to suffer from subjective optimism, investors clearly believe that they also convey material information" (p. 425). However, given the regulatory oversight of disclosures, sentiment in Yekini et al. (2016) is likely percieved as at least somewhat credible to investors.

Henry (2008) examines earnings press releases using prospect theory from Kahneman and Tversky (1979) as a starting point for understanding the effect of sentiment on investors. According to Henry, the influence of positive sentiment in corporate narratives is analogous to the influence of positive framing in other decision-making contexts, where choices differ when the issue is framed in positive instead of negative terms (see Tversky and Kahneman 1986). In this regard, narrative sentiment influences investors by framing the decision to invest in positive terms. Moreover, Henry applies the concept of reference points, explicitly testing for the concavity of positive sentiment relative to a reference point (neutrality). Henry finds that sentiment is positively and significantly correlated with abnormal returns; she also finds a concave relationship between abnormal returns and sentiment, in line with prospect theory. However, although Henry's goal is to examine whether investor behavior can be explained by prospect theory, she uses the setting of earnings press releases—which by definition, narrate financial figures. Therefore, the significant result could be due to the verifiability of sentiment against the earnings figures.

In a setting closest to the one in this research, Huang et al. (2014) are interested in whether sentiment is used to inform or mislead investors. Using the setting of earnings press releases between 1997 and 2007, the authors utilize a novel measure called abnormal tone, where abnormal tone is a sentiment that is orthogonal to fundamentals. They explicitly test the informativeness of this measure, finding that positive abnormal tone is correlated with negative firm performance; i.e. abnormal sentiment is, in fact, generally misleading. The authors further indicate that abnormally positive narratives were significantly correlated with positive abnormal market returns. This implies that investors reacted positively to sentiment, even when sentiment was misleading, and even in a context particularly rife with management incentives to mislead. As Huang et al. (2014) conclude, "[M]anipulation succeeds in misleading investors" (p. 1111).

5.3 Contribution

The standard framework, applied to the particular setting of CEO announcements, where sentiment is unverifiable and prone to manipulation, implies that sophisticated investors would likely not consider sentiment price-relevant. Furthermore, most of the empirical evidence that attempts to measure the isolated effect of sentiment really measures the effect of credible sentiment, which, even under the strictest assumptions of the standard model, could affect prices. However, whether sentiment matters for its own sake remains largely open. The only evidence which may estimate the effect of sentiment itself is limited to Huang et al. (2014), whose sample is relatively outdated, especially in the context of access to information in financial markets. Therefore, my goal is to provide the much-needed evidence as to whether investors react to sentiment for its own sake.

I make two distinct and novel contributions in this research. First, this paper is the first to apply sentiment analysis to CEO announcements, which will have specific implications for the voluntary disclosure theory from Dye (2001) and the literature on impression

management (Boudt and Thewissen 2019), as one would expect firms to generally use positive sentiment in composing their CEO announcements (H1). Second, this paper provides the first empirical evidence of market reaction to truly unverifiable sentiment. Failure to reject H2 would imply that investors, in line with theory, disregard unverifiable sentiment information. On the other hand, a rejection of H2 entails more discussion. It is possible that, despite the lack of financial figures, investors gauge the sincerity of narrative sentiment against the substantive content in the announcement, such as information about CEO background, like insider status, gender, education, and age. In this case, substantive content gains salience when financial figures are absent. It is also possible that, given the expertise of institutional investors and analysts, even unverifiable sentiment can be incrementally informative against the backdrop of investors' expertise and prior knowledge. Similarly, along the lines of signaling theory from Spence (1973), it is possible that unverifiable sentiment could still be imperfectly correlated with firm performance and therefore informative.

Finally, it is possible that sentiment matters to investors even when it is unverifiable. To quote Farrell and Rabin: "talk is cheap ... but, given that people respond to it, talk definitely affects payoffs" (1996, p. 104). This would be in line with the behavioral school of thought, which argues that investors regularly exhibit behavioral biases (see Lo 2005 for a concise summary of the behavioral finance literature). As a whole, the results of this research are, empirically, an important contribution to the corporate communication and strategic management literatures; and, theoretically, may also provide important insights into the ongoing debate regarding efficient markets and investor behavior.

5.4 Data and empirical approach

5.4.1 Data

Executive data come from the 2021 version of Wharton Research Data Service (WRDS) EXECUCOMP and include all CEOs appointed at S&P 1500 firms between 2010 and 2020. After excluding utilities and financial services firms (SIC codes 4900-4999 and 6000-6999) due to their non-standard and highly regulated market structures, the resulting sample includes 1,174 CEO appointments with 79 female CEOs and 1,095 male CEOs.

CEO announcements were hand-gathered using Nexis Uni full-text search and downloaded in their entirety. In cases where multiple versions of an announcement were available, I selected the earliest published announcement for download. Some announcements were unavailable or otherwise too compromised for inclusion and were therefore excluded from the sample. The resulting sample at this stage includes 1,079 CEO appointments with 76 female CEOs and 1,003 male CEOs.

Company financials and returns data for the event study come from WRDS Compustat and WRDS Eventstudy software, respectively. Because some observations are missing financials data or did not have enough returns data to complete the event study, the sample at this stage comprises 785 announcements with 730 male CEOs and 55 female CEOs. Additionally, data on CEO turnover reason comes from Gentry et al. (2021). As some observations are not included in the CEO turnover dataset, the final sample comprises 738 CEO announcements with 54 female CEOs and 684 male CEOs.[1]

[1]Results are robust to the different samples, see Appendix C.1 for details.

Variable	N	Mean	SD	Min	Max
Count of positive words	738	14.47	9.77	0	60
Count of negative words	738	1.81	2.74	0	25
Announcement length	738	330.86	174.46	25	1366
Sentiment	738	2.01	1.13	-1.61	4.62

TABLE 5.1: Summary statistics - Sentiment analysis

Measuring sentiment

Within the context of this research, sentiment is a measure of the positivity or negativity (polarity) of overall tone in a corporate narrative. I use the Quanteda package in R (Benoit et al. 2018) to carry out sentiment analysis on the announcements, generating a single sentiment score for each. The process is as follows. First, I upload the CEO announcement texts into the package and generate a corpus object comprised of the texts. I then apply lemmatization to the corpus, which replaces each word with its base form. For example, I replace the terms "planned," "plans," and "planning," with the base word "plan." Then, to perform the sentiment analysis itself, I apply the Loughran and McDonald dictionary, which was specifically developed for sentiment analysis in finance contexts (2011). The dictionary is comprised of positive and negative word lists, which serve as reference for Quanteda to count the number of positive and negative words in each text. Some examples of positive words are "advantage," "boost," and "innovate." Some examples of negative words are "adverse," "shortfall," and "volatility." See Loughran and McDonald (2011) for the full word lists and Table 5.1 for summary statistics of word counts and sentiment scores in the sample.

The average CEO announcement contains 331 words, 14.5 of which are positive and 1.8 of which are negative. The remaining words are neutral, or do not carry a sentiment according to the Loughran McDonald dictionary. There is large variation in the length of announcements, with the shortest announcement being only 25 words, and the longest announcement comprising 1366 words. Similarly, the count of positive and negative words varies substantially, with the minimum for both of zero words and the maximum of 60 and 25 words, respectively.

The Quanteda sentiment analysis package generates a sentiment score based on the log of the ratio of positive to negative words. This score is based on the work of Lowe et al. (2011), who, in their analysis of political manifestos, posit that a writer will manipulate the relative balance rather than the absolute quantity of polarized content when writing to communicate a particular position relative to two poles (2011, p. 131). Therefore, the overall sentiment of a CEO announcement is not a function of announcement length; it is a function of the ratio of positive to negative words. This scale is especially useful for this dataset, where the length of announcements varies greatly. The log scale also has the benefit of capturing realistic marginal effects as well, where the marginal effect of an additional positive (negative) word is decreasing in the number of positive (negative) words (Lowe et al. 2011). Therefore, $sentiment_i$ shown in Equation 5.1 gives a useful continuous measure of the overall sentiment of a CEO announcement:

$$sentiment_i = ln\ (\frac{count\ of\ positive\ words}{count\ of\ negative\ words}) \tag{5.1}$$

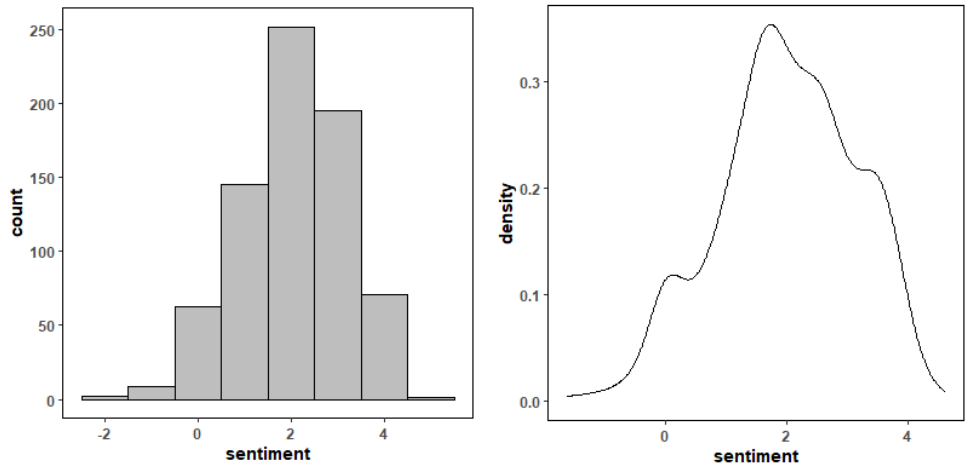


FIGURE 5.1: Distribution of sentiment

where negative scores indicate a negative overall sentiment, zero represents a tone that is neither negative nor positive, and positive scores indicate a positive sentiment. The Quanteda package uses a smoothing parameter of 0.5 to overcome situations in which there are no negative or positive words; that is, when either count is equal to zero, the smoothing parameter replaces zero with 0.5. This avoids the problem of taking the logarithm of zero. The distribution of sentiment can be seen in Figure 5.1. To give context, the announcement with the most positive sentiment (score of 4.62) has 50 positive words and 0 negative words. The most negative announcement (score of -1.61) has 2 negative words and 0 positive words. The distribution of sentiment scores for the CEO announcements follows a left-skewed distribution with a mean of 2.01. Examples of positive, neutral, and negative announcements can be found in Appendices C.2 through C.4.

In addition to the continuous measure of sentiment, I generate two additional specifications. First, I generate overall sentiment, which can take three values based on the continuous sentiment score: overall sentiment is equal to 1 when sentiment is positive, equal to 0 when sentiment is neutral, and equal to -1 when sentiment is negative. This specification allows me to capture the average effects of positivity and negativity relative to the baseline of sentiment neutrality. Second, I generate a discrete categorical variable (discrete sentiment), where the new variable is grouped by one-point buckets of the continuous sentiment score. For example, scores between 1.0 and 1.9 are grouped together and scores between 2.0 and 2.9 are grouped together, and so on. This specification allows me to analyze whether market reaction to positivity changes with degree of positivity. Table 5.2 shows summary statistics for discrete sentiment.

As shown in Table 5.2, a small portion of the sample (17 announcements or 2.3 percent) have a negative overall sentiment. In addition, there are 38 announcements with a score of zero. There are two ways for this to score to obtain. First, an announcement may contain zero sentiment words (neutral sentiment); there are 12 of these announcements. Second, an announcement may contain exactly as many positive words as negative words (balanced sentiment); there are 26 of these announcements. Due to the small sample size, I do not differentiate between balanced and neutral sentiment scores and consider all scores of 0 to

Variable	N	Percent
Discrete sentiment	738	
... < 0	17	2.33%
... 0 Neutral	38	5.15%
... 0 to 0.9	74	10.03%
... 1 to 1.9	239	32.38%
... 2 to 2.9	214	29.01%
... 3 or higher	156	21.14%

TABLE 5.2: Summary statistics - Discrete sentiment

be effectively sentiment-neutral.[2] Approximately 10 percent of the sample have a slightly positive sentiment between 0.0 and 0.9, and 32 percent have a positive score of between 1.0 and 1.9. Approximately 29 percent have a positive sentiment between 2.0 and 2.9, and 21 percent of the sample have a highly positive sentiment (3.0 or higher). Figure 5.1 and Table 5.2 make clear that, on average, the CEO announcements in this sample have a positive sentiment. This provides initial support for H1, where I reasoned, based on the voluntary disclosure theory from Dye (2001), that companies would intentionally a craft positive tone in their voluntary disclosures.

Controlling for substantive information

In order to empirically identify the market reaction to sentiment, it is necessary to control for the substantive information contained in the CEO announcements that investors may perceive as relevant for pricing. As noted in Price et al. (2012) and Leitch and Sherif (2017), the extant empirical literature indicates that certain CEO characteristics have a significant effect on stock pricing around CEO announcements. In this section, I briefly review the literature to derive and explain the covariates used in the empirical analysis.

First, the extant literature indicates that, in general, investors react more positively to outsider CEO announcements (Charitou et al. 2010, Jalal and Prezas 2012). However, the decision to appoint an insider is endogenous to firm decision making, and therefore highly situation-dependent. Insider appointments may reflect a Board that prefers fewer information asymmetries (Zajac 1990; Rose 2019), a policy of internal promotion to incentivize workers (Chan 1996), or a firm with greater managerial depth (Parrino 1997; Dalton and Kesner 1983). In contrast, outside succession is an effective signal of breaking with previous policies; both Farrell and Whidbee (2003) and Parrino (1997) find that outsider appointments are more likely following a forced resignation. In order to control for market reaction to insider status, I generate a binary variable which is equal to one when the announcement indicates that the incoming CEO is a company insider and equal to zero when the announcement indicates that the incoming CEO is an outside hire.

Second, the empirical literature indicates that CEO education (Cheng et al. 2010) and MBA status in particular (You et al. 2020; Bhagat et al. 2010; Betrand and Schoar 2003), are strongly correlated with firm performance, and, therefore, may influence market reaction to CEO announcement. To control for information on CEOs MBA status, I use Quanteda to

[2]Results are robust to specifications that differentiate between neutral and balanced sentiments, see Appendix C.5.

determine which announcements indicate that the incoming CEO has an MBA, specifying a binary variable which is equal to one when the announcement indicates that the incoming CEO has an MBA. It is important to note that the comparison group (binary variable equal to zero) for this variable is composed of both CEOs who have an MBA, but the announcement does not include this information, and all CEOS who do not have an MBA.

Third, the literature indicates that CEO age is also relevant for investors' perceptions of future firm performance. There is evidence that CEO age is negatively correlated with operating leverage (Serfling 2014), Tobin's Q (Cline and Yore 2016), R&D investment (Serfling 2014), acquisition activity (Yim 2013), and overall firm risk-taking (Peltomäkki et al. 2021). Indeed, approximately half of all S&P 1500 firms have mandatory retirement policies, typically at age 65 (Cline and Yore 2016). To control for the reporting of CEO age, I generate a binary variable which is equal to one when the age of the incoming CEO is reported.

Finally, investors may perceive the gender of the incoming CEO as relevant for future firm performance. Braegelmann and Ujah (2020), Lucy and Carron (2011), and Lee and James (2007) find that the market reacts negatively to female CEO announcements compared to male CEO announcements. Dixon-Fowler et al. (2013) find evidence of a negative contagion effect: not only does the market react negatively to new female CEOs, the market also discounts the stocks of companies already led by woman CEOs. To control for market reaction to CEO gender, I generate a binary variable which is equal to one when CEO gender is female. While only two of the female CEO announcements explicitly refer to the incoming CEO as the first woman or first female CEO, other announcements use gendered names, titles, or pronouns that effectively signal gender to investors. Therefore, investors are able to determine this information from the announcement, even when gender is not explicitly stated.

Controlling for CEO turnover reason

Finally, I use the CEO turnover dataset from Gentry et al. (2021) in order to control for the departure reason of the previous CEO. This is critical information, because it represents the firm-specific situation against which investors update their expectations of firm performance. For example, the literature on insider appointments indicates that outsider appointments may be more common following an involuntary turnover, as the Board of Directors may wish to signal a change from the status quo (Farrell and Whidbee 2003, Parrino 1997). Using the dataset from Gentry et al. (2021), I am able to distinguish between different types of voluntary and involuntary turnovers. I use voluntary turnovers due to retirement as the baseline for this factor variable. A detailed explanation behind the methodology and coding can be found in Gentry et al. (2021) p. 975. Basic summary statistics for the covariates of the final sample are shown in Table 5.3, and further descriptive analysis of covariates follows in the Empirical Approach section.

Summary statistics and descriptive analysis

In this dataset, female CEOs represent seven percent of the sample, and approximately 28 percent of CEO announcements indicate that the incoming CEO is a company insider. Companies in the sample are spread across the Standard and Poor's (SP) indices, with 38 percent of companies in the SmallCap 600 index, 27 percent of companies in the MidCap 400 index, and 36 percent of companies in the LargeCap 500 index.

Variable	N	Mean
Gender	738	
... Female	54	7%
... Male	684	93%
Insider	738	
... No	204	28%
... Yes	534	72%
SP Index	738	
... LargeCap	262	36%
... MidCap	199	27%
... SmallCap	277	38%
MBA reported	738	
... No	656	89%
... Yes	82	11%
Age reported	738	
... No	528	72%
... Yes	210	28%
Turnover reason	738	
... Voluntary - Retirement	485	66%
... Voluntary - New opportunity	19	3%
... Involuntary - CEO death	10	1%
... Involuntary - CEO illness	23	3%
... Involuntary - job performance	133	18%
... Involuntary - personal issues	10	1%
... Other	58	8%

TABLE 5.3: Summary statistics - Covariates

Variable	N	Mean	SD	Min	Max
Age when age reported	210	52.7	5.4	33	67
Age in Boardex subsample	454	61.1	6.5	42	83
Age in Peltomäki et al. 2021	14,839	56.3	6.9	28	89

TABLE 5.4: Summary statistics - Age

Furthermore, approximately 11 percent of announcements indicate that the incoming CEO has an MBA degree. As a voluntary disclosure, it is possible that firms strategically report MBA status of the incoming CEO; according to Dye's (2001) theory, we would expect firms to report this favorable information when they can. To briefly examine whether there is strategic reporting, I use the BoardEx (2023) dataset—a premiere source for personal data on executives—to compare how many CEOs have an MBA with how many announcements report this information. In the BoardEx sample, which covers about 60 percent of the same final sample used here, approximately 40 percent of incoming executives have an MBA. In comparison, only 11 percent of announcements in this sample report that the new CEO has an MBA. This relatively small number implies that firms actually underreport MBA status. This is surprising, as the empirical literature indicates that investors would likely respond positively to CEOs with MBAs.

Similar to the choice of reporting MBA status, approximately 28 percent of announcements in this sample voluntarily indicate the age of the incoming CEO in the announcement. Because reporting CEO age is also likely a strategic choice, I compare the reported ages from this sample with the overall S&P 1500 CEO age data reported in Peltomäki et al (2021), who analyze a similar period (2006 to 2018) and scope of data (S&P 1500 firms excluding financial institutions), and the executive age data from BoardEx. Summary statistics are shown in Table 5.4. In comparison to the data from Peltomäki et al. (2021) and BoardEx, the average voluntarily reported CEO age is lower, with a mean of 52.7 years compared to 56.3 years in Peltomäki and 61.1 years in Boardex. It also appears that the standard deviation for the voluntarily reported CEO age is smaller in this dataset than in the Peltomäki and BoardEx datasets, with a standard deviation of 5.4 compared to 6.9 and 6.5, respectively. Therefore, it appears that firms choose to report age for relatively young CEOs; moreover, it appears that firms generally refrain from reporting ages at either extreme of the age distribution. This is approximately in line with the expectations of Dye (2001) and indicates that there is some strategic reporting of CEO age. For more detail, a histogram of reported age is shown in Appendix C.6. Furthermore, Appendices C.7 and C.8 show the summary statistics by age and MBA reporting; the other variables of interest are generally not conditional on reporting.

Voluntary retirements comprise the majority of the data at 66 percent, whereas voluntary turnovers for a new opportunity represent 3 percent of the sample. Involuntary turnovers due to job performance (i.e. firings) comprise 18 percent of the data, and involuntary turnovers due to CEO illness, death, or personal issues (such as sexual harassment, improper behavior, or illegal activity) each represent 3 percent or less of the sample. The turnover category "other," which includes interim CEOs or turnovers due to mergers or acquisitions, comprise 8 percent of the sample.

Table 5.5 shows the summary statistics for sentiment by each turnover category. Average sentiment is positive for all categories. Moreover, for the two categories which represent

		Sentiment			
Variable	N	Mean	SD	Min	Max
Voluntary - Retirement	485	2.0941	1.0711	-1.0986	4.6151
Voluntary - New opportunity	19	1.9602	1.3606	-1.6094	4.2341
Involuntary - CEO death	10	0.7259	0.7279	-0.3677	1.8458
Involuntary - CEO illness	23	1.546	1.2133	0.0000	3.5553
Involuntary - job performance	133	1.9562	1.2209	-1.6094	4.4659
Involuntary - personal issues	10	1.9846	0.9151	0.7472	3.434
Other	58	1.8927	1.2071	-0.5108	3.9703

TABLE 5.5: Summary statistics - Sentiment by turnover reason

turnovers due to firings (involuntary - job performance and involuntary - personal issues), the average sentiment is positive, and maximum sentiments are over 3.4 in both categories. This provides further evidence in support of H1, as firms tend to craft positive sentiments even when the turnover is forced due to major failings in leadership. This follows the interpretation of firms who strategically manage their public image in the face of poor performance or scandal. The lowest average sentiments are in announcements following CEO death or illness, which may represent firms' respect for privacy and discretion following a CEO death or health issues. This provides some support for the interpretation that firms attempt to maximize the positive impact of their new CEO announcements given the constraints of the background reason for the turnover. For example, the use of highly positive sentiment following a CEO death would likely be considered to be in poor taste; in contrast, firms appear to use sentiment as a way to re-frame a break with the status quo after a period of poor performance or scandal.

5.4.2 Empirical approach

Measuring market reaction

In order to measure the stock market reaction to CEO announcements, I use an event study based on a Fama-French three-factor model. The event study includes an estimation period of 250 days with at least 200 days of trading activity, a gap period of 20 days to minimize noise, and a three-day event window $= \{-1, 0, 1\}$ where the announcement date is $t = 0$.[3] I use returns from the estimation period to estimate the counterfactual: how daily returns would have developed in the absence of the event. Equation 5.2 represents the expected daily return $\hat{R}_{RiskModel,i,t}$ for firm i:

$$\hat{R}_{RiskModel,i,t} = R_{f,t} + \hat{\alpha}_i + \hat{\beta}_{1i}\left(R_{m,t} - R_{f,t}\right) + \hat{\beta}_{2i}SMB_t + \hat{\beta}_{3i}HML_t \qquad (5.2)$$

where $R_{f,t}$ is the risk-free rate in the market, $\hat{\alpha}_i$ is an intercept, $\hat{\beta}_{1i}$ captures the market premium, $\hat{\beta}_{2i}$ captures the excess return of small stocks relative to large stocks, and $\hat{\beta}_{3i}$ captures the excess return of stocks with high market-to-book ratios relative to stocks with a low market-to-book ratio.

[3]Results are generally not conditional on window lengths or choice of estimation model, see Appendices C.9 and C.10.

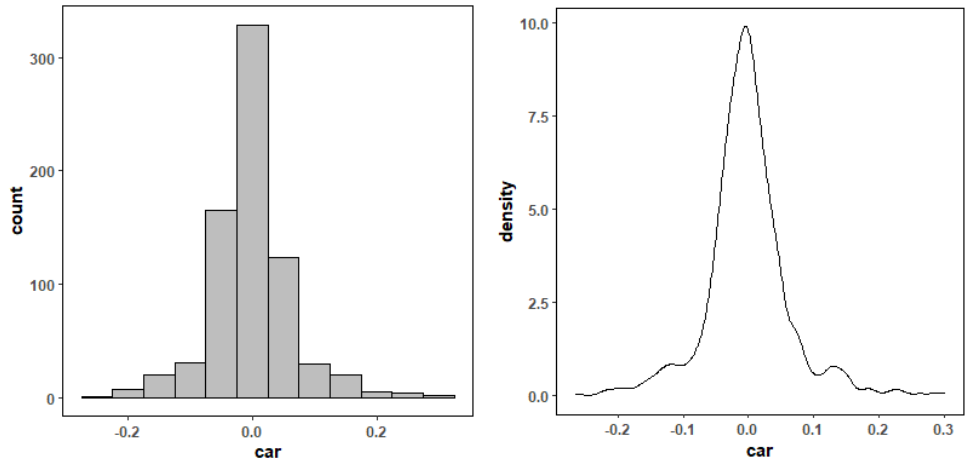


FIGURE 5.2: Distribution of CAR

The abnormal return $AR_{i,t}$ for firm i on day t is simply the difference between the actual stock performance and the expected stock performance as estimated by the Fama-French regression. The abnormal returns are then summed over the event window, resulting in a cumulative abnormal return (CAR) for each CEO announcement i, as shown in Equation 5.3:

$$CAR_i = \sum_{t=1}^{T} AR_{i,t} \tag{5.3}$$

Descriptive analysis of market reaction

The resulting CAR from the event study represents the market reaction to the new information gained from the CEO announcement. Here, I provide an initial descriptive analysis of the relationship between market reaction and sentiment, as well as the relationships between these variables of interest and the covariates. As shown in Figure 5.2 and Table 5.6, CARs are approximately normally distributed with a slightly negative mean market reaction of -17 basis points and a standard deviation of 634 basis points.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
CAR	738	-0.0017	0.0634	-0.2653	-0.0306	0.025	0.3026

TABLE 5.6: Summary statistics - CAR
Note: shows summary statistics for market reaction as measured by CAR

Figure 5.3 represents a scatterplot of CAR (y-axis) and sentiment (x-axis). In general, it appears that there is a slightly positive relationship between CAR and sentiment (correlation coefficient of 0.10 significant at the 1 percent level). A Breusch-Pagan test on a simple linear regression of sentiment on CAR finds no evidence of heteroskedasticity.

To examine the relationships between market reaction, sentiment, and covariates, Table 5.7 shows a correlation matrix of all variables in the analysis.[4] According to Table 5.7, CAR

[4]Density functions of CAR for covariates can be found in Appendix C.11.

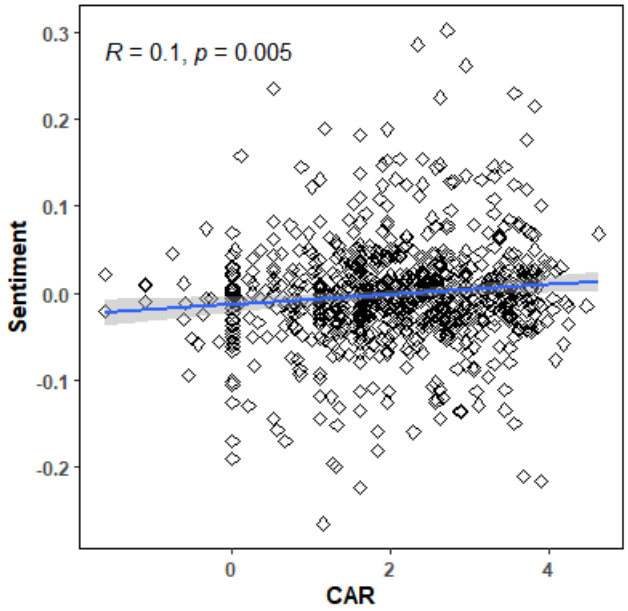


FIGURE 5.3: Scatterplot of announcement sentiment and CAR
Note: shows a scatterplot of announcement sentiment and CAR. The blue line indicates a simple linear regression of sentiment on CAR, with a Pearson correlation coefficient of 0.1 (p = 0.006)

and sentiment are indeed positively correlated, with a correlation coefficient of 0.10 (significant at the one percent level). Similarly, CAR and MidCap are weakly negatively correlated (significant at the ten percent level); furthermore, MBA reporting has a small but positive correlation with CAR (significant at the ten percent value). Sentiment, in addition to being significantly positively correlated with CAR, is also correlated with MBA reporting (at the one percent level) and with age reporting (at the one percent level). This implies that MBA and age—which are candidates for strategic reporting–are more likely to be reported in announcements with more positive sentiment. Furthermore, the correlation between MBA and age reporting (significant at the five percent level) indicates that they may be more likely to be reported together, and both of these strategic variables are correlated with announcement length (significant at the one percent level). Finally, sentiment is also positively correlated with both voluntary turnover and announcement length (at the one percent level).

	CAR	Sentiment	Female	Insider	SP	MD	SM	MBA reported	Age reported	Voluntary turnover	Ann. length
CAR	1.00										
Sentiment	0.10***	1.00									
Female	−0.01	−0.01	1.00								
Insider	−0.02	−0.04	0.02	1.00							
LargeCap	0.05	0.01	−0.02	0.13***	1.00						
MidCap	−0.10*	0.00	0.01	−0.01	—	1.00					
SmallCap	0.04	−0.01	0.02	−0.12***	—	—	1.00				
MBA reported	0.06*	0.16***	−0.03	−0.04	−0.04	0.02	0.02	1.00			
Age reported	−0.03	0.10***	−0.05	0.00	0.08**	−0.02	−0.07*	0.06*	1.00		
Voluntary turnover	0.01	0.10***	0.02	0.20***	0.00	0.02	−0.02	−0.02	0.01	1.00	
Ann. length	−0.01	0.25***	0.03	−0.02	0.03	−0.01	−0.02	0.21***	0.15***	0.02	1.00

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE 5.7: Correlation matrix

Note: shows a correlation matrix for all variables of interest and covariates.

The binary variable Female is not significantly correlated with any other variables. In contrast, insider appointments are positively correlated with LargeCap firms (one percent level) and negatively correlated with SmallCap firms (one percent level), which is in line with the managerial depth theory from Dalton and Kesner (1983). Furthermore, as expected, insider appointments are positively and significantly correlated with voluntary CEO turnovers (one percent level), consistent with the implications of Farrell and Whidbee (2003) and Parrino (1997). Finally, reporting of age is positively correlated with LargeCap firms (ten percent level) and negatively correlated with SmallCap firms (ten percent level), but it is not correlated with MidCap firms.

The significant correlations between sentiment and MBA reporting, age reporting, voluntary turnover, and announcement length lend themselves to an interpretation of impression management. This could be typified by a "passing of the torch" announcement, characterized by a voluntary retirement, announced with a relatively long press release, with strategic reporting of more background information about the incoming CEO, who is likely a hand-picked inside successor. An example of such a "passing of the torch" announcement can be found in Appendix C.12. However, the absolute magnitude of the correlations in question are relatively small.

Regression analysis of market reaction

The descriptive analysis indicates that market reaction is positively correlated with announcement sentiment. However, in order to more rigorously analyze the relationship and isolate the effect of sentiment on market reaction, I use a simple OLS regression with CAR_i as the dependent variable, estimating Equation 5.4:

$$CAR_i = \beta_1 Sentiment_i + \beta_2 X_i + \epsilon_i \tag{5.4}$$

where $Sentiment_i$ represents the continuous sentiment score and X_i represents the full battery of covariates based on the empirical literature, including: insider, where the information that an incoming CEO is a company insider is equal to one; age, which is a binary variable that is equal to one when the announcement indicates the age of the incoming CEO; MBA status, which is a binary variable equal to one when the announcement indicates that the incoming CEO has an MBA, and gender, where female is equal to one. I also control for firm size, which is proxied for by S&P index based on capital categorization (SmallCap 600, MidCap 400, and LargeCap 500 firms), with LargeCap firms as the reference level. To control for the situation-specific reason for the CEO turnover event, I use the dataset from Gentry et al. (2021) with voluntary retirement as the reference level, and I also control for the mention of a simultaneous event, such as a spin-off or merger. Finally, I include control variables based on the event study literature, which include annual return on equity (ROE) and leverage ratio. Because CAR_i is a generated dependent variable, I estimate my regression with White (1980) robust standard errors. This procedure accounts for heteroscedasticity present in the residuals due to sampling error in the estimation of the dependent variable (Lewis and Linzer 2005).

5.5 Results

I display the regression results in Table 5.8. All regressions contain control variables for ROE, debt ratios, the announcement of a simultaneous event, and announcement length. The three models represent the step-wise addition of covariates; Model 1 includes the basic control variables of firm size (S&P Index) and the reason for the CEO turnover event. Model 2 includes additional substantive information variables of insider and gender; Model 3, the full specification, includes MBA and age reporting, which may be strategically reported. Results are generally robust to different specifications of the sentiment measure; see Appendix C.13 for more detail.

Model 3 represents the final specification, which I explain in detail here. In H2, I reason that sentiment will not affect market reaction to CEO announcement, because sentiment represents unverifiable information that investors disregard when markets are efficient and investors understand firms' incentives to disclose favorable information. In fact, the coefficient for sentiment is positive and highly statistically significant; moreover, it is stable across all specifications. The coefficient of 0.0062 means that, on average, a one-point increase in sentiment score would be greeted by an average 62 basis point increase in stock price, all else equal.

The coefficient for MidCap firms is also negative and highly significant, indicating that, on average, announcements of CEOs to MidCap companies are reacted to negatively (-182 basis points) compared to LargeCap companies. There is no significant effect for SmallCap companies. According to Model 3, investors react negatively (-118 basis points) to involuntary turnovers due to job performance; although this effect is not surprising, it is only weakly significant. There is no significant effect of the information about the incoming CEO being a company insider, nor is there a significant effect for female CEO announcements. Model 3 indicates that, on average, markets react positively (135 basis points) to announcements that report that the incoming CEO has an MBA; however, this effect is only significant at the 10 percent level.

Based on the results of Model 3, I find evidence that sentiment is significantly and positively correlated with market reaction; this evidence results in a rejection of H2. In order to further examine the effect of sentiment on market reaction, I provide the regression results of the additional specifications of sentiment. Models 4 through 6 contain the full specification of control variables as shown in Model 3, but are not shown for the sake of clarity.

Model 4 in Table 5.9 shows the result of the regression with overall sentiment, which captures the average market reaction of positivity and negativity relative to neutrality. According to Model 4, there is no significant difference between market reaction to neutral and negative announcements. However, there is an economically meaningful and statistically significant difference between neutral and positive announcements. On average, the market reacts to positive announcements with a 354 basis point increase in stock price compared to neutral announcements. This is a large effect, representing an average 3.54 percent abnormal increase in stock price in the event window.

Furthermore, it is also possible that the investors react differently to different levels of positive sentiment. According to the standard model, investors are likely skeptical of unverifiable sentiment; it follows that the relationship between sentiment and returns may depend on the level of positivity. Therefore, Model 5 shows the regression for discrete sentiment.

	Model 1	Model 2	Model 3
Sentiment	0.0065***	0.0065***	0.0062***
	(0.0022)	(0.0022)	(0.0022)
MidCap	−0.0176***	−0.0176***	−0.0182***
	(0.0057)	(0.0057)	(0.0057)
SmallCap	−0.0047	−0.0048	−0.0055
	(0.0060)	(0.0061)	(0.0061)
Voluntary - New opportunity	−0.0130	−0.0130	−0.0140
	(0.0100)	(0.0102)	(0.0104)
Involuntary - CEO death	0.0103	0.0100	0.0104
	(0.0163)	(0.0164)	(0.0164)
Involuntary - CEO Illness	−0.0047	−0.0048	−0.0049
	(0.0111)	(0.0112)	(0.0112)
Involuntary - Job performance	−0.0112*	−0.0115*	−0.0118*
	(0.0066)	(0.0068)	(0.0068)
Involuntary - Personal issues	−0.0002	−0.0007	−0.0006
	(0.0233)	(0.0233)	(0.0227)
Departure - Other	0.0174*	0.0173	0.0169
	(0.0104)	(0.0109)	(0.0108)
Insider		−0.0008	−0.0007
		(0.0059)	(0.0059)
Female		−0.0046	−0.0043
		(0.0087)	(0.0087)
MBA reported			0.0135*
			(0.0076)
Age reported			−0.0039
			(0.0046)
Intercept	−0.0019	−0.0008	0.0009
	(0.0072)	(0.0090)	(0.0090)
R^2	0.0470	0.0474	0.0522
Adj. R^2	0.0299	0.0276	0.0298
Num. obs.	738	738	738

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE 5.8: Results of full specification

Note: displays the result of the full specifications

	Model 3	Model 4	Model 5
Sentiment	0.0062***		
	(0.0022)		
Overall sentiment negative		0.0223	
		(0.0140)	
Overall sentiment positive		0.0354***	
		(0.0100)	
Discrete sentiment < 0			0.0232
			(0.0141)
Discrete sentiment 0 to 0.9			0.0369***
			(0.0121)
Discrete sentiment 1 to 1.9			0.0296***
			(0.0103)
Discrete sentiment 2 to 2.9			0.0424***
			(0.0111)
Discrete sentiment 3 or higher			0.0404***
			(0.0113)
R^2	0.0522	0.0555	0.0623
Adj. R^2	0.0298	0.0319	0.0348
Num. obs.	738	738	738

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE 5.9: Detailed results
Note: displays the result of the full specifications (covariates not shown for clarity) for Sentiment, overall sentiment, and discrete sentiment.

The results of Model 5 are in line with Model 4, as there is no significant reaction to negative announcements (Lower than 0). For weakly positive scores between 0.0 and 0.9, the coefficient is 369 basis points compared to neutrality (significant at the one percent level). For positive scores between 1.0 and 1.9, the coefficient is 296 basis points (significant at the one percent level). For positive scores between 2.0 and 2.9, the coefficient is 424 basis points (significant at the one percent level) and for sentiment scores 3 or higher the coefficient is slightly lower at 404 basis points (significant at the one percent level). However, given the size of the standard errors, the coefficients for the different positive discrete groups are not significantly different from one another. Therefore, I am not able to conclude that investors react differently to different levels of sentiment.

5.6 Discussion

The goal of this paper is to analyze whether the sentiment of CEO announcements affects market reaction. First, based on the descriptive evidence from the sentiment analysis, I find that the CEO announcements in the sample have, on average, an overall positive tone. Specifically, the average announcement in the sample uses about seven positive words to every negative word. With respect to Dye's (2001) voluntary disclosure theory, I find evidence that companies do, on average, use a positive tone in their CEO announcements, a finding which holds even for CEO announcements following firings based on poor performance or personal scandal. This implies support for the theory that firms tend to emphasize positive

and downplay negative information in voluntary disclosures. This is in line with the evidence for impression management in the corporate communications literature (Patelli and Pedrini 2014, Boudt and Thewissen 2019). It appears likely that firms attempt to maximize the positive impact of the narrative subject to the constraints of the firm-specific situation surrounding the CEO turnover.

In the second step of this research, the regression results indicate that investors react significantly and positively to sentiment in CEO announcements. Compared to neutral announcements, investors reacted to positive announcements with an average 3.54 percent higher abnormal returns. This is an economically significant change in terms of stock prices. This significant reaction presents a contrast to the implications of the standard model, as sentiment in CEO announcements is neither verifiable against financial figures, nor do third parties validate the narratives. CEO announcements are forward-looking and highly subjective in nature, and represent an important public relations opportunity for companies. In this setting, investors cannot easily determine whether management is sincerely communicating or strategically obfuscating.

Therefore, the result that markets react significantly and positively to sentiment in CEO announcements is, in my view, at least partly attributable to the reaction of investors to unverifiable sentiment. In line with the initial findings of Huang et al. (2014), I find that investors appear to react to sentiment even when it likely does not provide a reliable signal of future firm performance. The unique contribution of this research is the setting: it appears that investors react to sentiment for its own sake, even when sentiment is unverifiable, the narrative is highly subjective in nature, and management have strong incentives to manage the company image.

However, the underlying reason why investors react to sentiment is not made clear by this research. It is possible that the reaction represents a behavioral bias, where investors are susceptible to positive sentiment even when it is clearly manipulable. Another possible explanation is that investors use the substantive information in CEO announcements as a quasi-hard facts for verification of sentiment. Similarly, it is also possible that sentiment is correlated with firm performance or sophistication, along the lines of a Spence (1973) signalling model. However, given the speculative nature of CEO announcements and the lack of any verification mechanism, the strong empirical result implies that, at least to some extent, investors seem susceptible to narrative sentiment.

The main finding, that markets appear to react significantly and positively to sentiment, has important implications for financial markets. If investors react positively to positive sentiment in corporate narratives, this represents an opportunity for exploitation by firms. If firms are aware that investors react to positive sentiment—even for a short time—they may have an incentive to use sentiment strategically when crafting corporate narratives. Taken to the extreme, this behavior would lead to noisy stock prices and additional costly information-gathering for investors.

As with any research, this paper has limitations that future research should address. Most importantly, this research is not able to clarify the underlying reason that investors react to sentiment in this setting; therfore, I cannot decisively rule out the possibility that unverifiable sentiment can be incrementally informative. Future research should seek to analyse whether, for example, sentiment in CEO announcements is correlated with future firm performance and may therefore act as a signal of firm performance or sophistication.

Additionally, this research was not able to conclude that market reaction to sentiment is conditional on level of positivity. Therefore, future research should more robustnly examine a possible non-linear relationship between CAR and sentiment along the lines of Henry (2008).

In addition, this research has methodological limitations. First, documents were hand-searched and hand-downloaded using Nexis Uni. Some variables, such as company insider, were hand-coded by researchers who read the articles. Assuming that a sufficiently sophisticated text analysis program could overcome human error in coding the variables, future research could potentially improve the analysis of the market reaction to company insiders. In addition, research that includes CEO gender is fundamentally constrained by the small number of female CEOs. Future research should attempt to replicate the findings around CEO gender using different and larger samples. Finally, market-level dynamics such a strategic noise (Graffin et al. 2010) and timing of corporate communications (Burchard et al. 2020) remain out of the scope of this paper, but could nevertheless be relevant for strategic communication behavior on behalf of firms. Future research should seek to address these dynamics.

5.7 Conclusion

Building on the empirical evidence that investors react to sentiment in other corporate narratives, I investigate whether investors similarly react to the sentiment expressed in CEO announcements, which are not subject to either of the two primary verification mechanisms. Analyzing a dataset of 738 CEO announcements in the S&P 1500 between 2010 and 2020, I find that that firms tend to craft positive sentiments in their CEO announcements. Moreover, I find an economically and statistically significant effect of sentiment on market reaction, with positive CEO announcements yielding substantially higher abnormal returns than neutral ones.

These results provide support for the voluntary disclosure theory from Dye (2001) and the literature on impression management (see Boudt and Thewissen 2019), indicating that firms tend to emphasize positive information while downplaying unfavorable details in CEO announcements. The results also suggest that investors react to unverifiable sentiment. Because of the specific setting of CEO announcements, this result implies that investors react to sentiment for its own sake. However, this interpretation is not clear-cut and requires futher empirical testing. While my findings highlight the possibility that firms could strategically employ sentiment as a means to influence investors, more research is needed to discern whether the influence is due to behavioral bias, or whether sentiment could be verified in other ways.

Chapter 6

General conclusions

In this dissertation, I contribute to the existing literature on market inefficiencies by providing empirical evidence regarding market power and possible behavioral biases. Because market inefficiencies can lead to a reduction in consumer welfare, inefficient resource allocation, and, in the worst case, total market failure, this research is important for recognizing and addressing such inefficiencies. Furthermore, the empirical nature of the dissertation leads to implications for the specific markets and societies studied in Chapters 2 through 5.

In Chapter 2, my co-author and I focus on the German private health insurance market, which is characterized by notable market power held by brokers. Because new business is limited and profitable, insurers are highly motivated to write new contracts. And, because contracts are complex and new business is almost exclusively sold through brokers, brokers had market power, leading to excessive commissions. These commissions were ultimately passed on to consumers, making premiums more expensive and reducing the value of insurance. In Chapter 2, we empirically examine the efficacy of a two-pronged reform in this market. We find evidence that the introduction of a minimum cancellation liability period was likely successful in reducing the inefficient reshuffling of customers between insurers. In contrast, the results imply that the introduction of a commission cap had no significant effect on the level of total acquisition costs. Rather, the commission cap appeared to contribute to a large and significant decline in new business. Therefore, it is likely that, faced with capped commissions, the insurers' resulting marketing mix is even less efficient. These findings underscore the difficulty of cost regulation in insurance markets.

In Chapter 3, my co-author and I provide an empirical test of social role theory by analyzing the relationship between gender equality and the perception of income fairness in European countries. The findings of Chapter 3 imply that higher gender equality correlates with an increased likelihood of individuals, regardless of gender, perceiving their income as fair. This highlights the potential long-term societal benefits of gender-equality policies. Moreover, our research presents initial evidence supporting the convergence of perception of income fairness with rising gender equality, aligning with social role theory's prediction that gender differences diminish as social roles lose their salience. The complex nature of the relationship between gender equality and gender differences warrants further exploration in future research, particularly within an international context.

In Chapter 4, my co-author and I analyze whether investors exhibit a gender bias. Such a behavioral bias would lead to the inefficient allocation of capital and potentially endanger

the pricing mechanism. Using event study methodology, we find that the market reacts negatively to female CEO announcements and positively to male CEO announcements; moreover, the difference is highly statistically significant. Therefore, we find some evidence that gender bias does exist in financial markets. More importantly, our findings imply that the bias is conditional on firm size, which supports the role of information in mitigating gender bais. Furthermore, the finding that gender bias changes over time supports the predictions of token theory, where gender bias should decrease as female CEOs become more common.

In Chapter 5, I examine market reaction to narrative sentiment in new CEO announcements. First, I use sentiment analysis to generate a measure of the overall sentiment of each CEO announcement. Then, I use event study methodology to identify market reaction to each CEO announcement; this allows me to analyze whether narrative sentiment is significantly correlated with market reaction. I find that the CEO announcements are, on average, written with a positive overall sentiment. This holds even for announcements following forced turnovers. Moreover, results indicate that investors react significantly and positively to the sentiment of CEO announcements. These results are important, because they are likely evidence of a market inefficiency with a potentially worsening dynamic. Given the particular setting, narrative sentiment in CEO announcements is unverifiable and firms have an incentive to manipulate sentiment; that investors react to sentiment for its own sake creates an opportunity for firms to exploit. If firms recognize that markets reward positive sentiment, firms may be more likely to engage in strategic impression management, creating noisy stock prices and generally decreasing the efficiency of capital allocation.

In summary, my dissertation provides key evidence regarding the efficacy of market reforms to overcome inefficiencies due to market power and the possible existence of behavioral biases in financial markets. This dissertation has four main implications. First, my research implies that commission caps may not be effective to regulate brokers' market power in insurance markets. Second, this research implies that gender equality may be a social good, where increasing gender equality is positively correlated with the likelihood of all individuals, regardless of gender, perceiving their income as fair. Third, my thesis provides evidence that gender bias exists in financial markets. More importantly, the results imply that information mitigates the bias and that gender bias appears to abate as female CEOs become more commonplace. Finally, my thesis provides initial evidence that investors react to sentiment for its own sake, even in a setting where it is unverifiable. Based on the research presented in this dissertation, future research should address whether commission caps are effective in other markets, whether firm performance is conditional on CEO gender, and whether and under what conditions unverifiable sentiment could provide price-relevant information to investors.

Appendix A

Appendix to Chapter 2

A.1 Development of inflation-adjusted JAEG and new SubstHI business

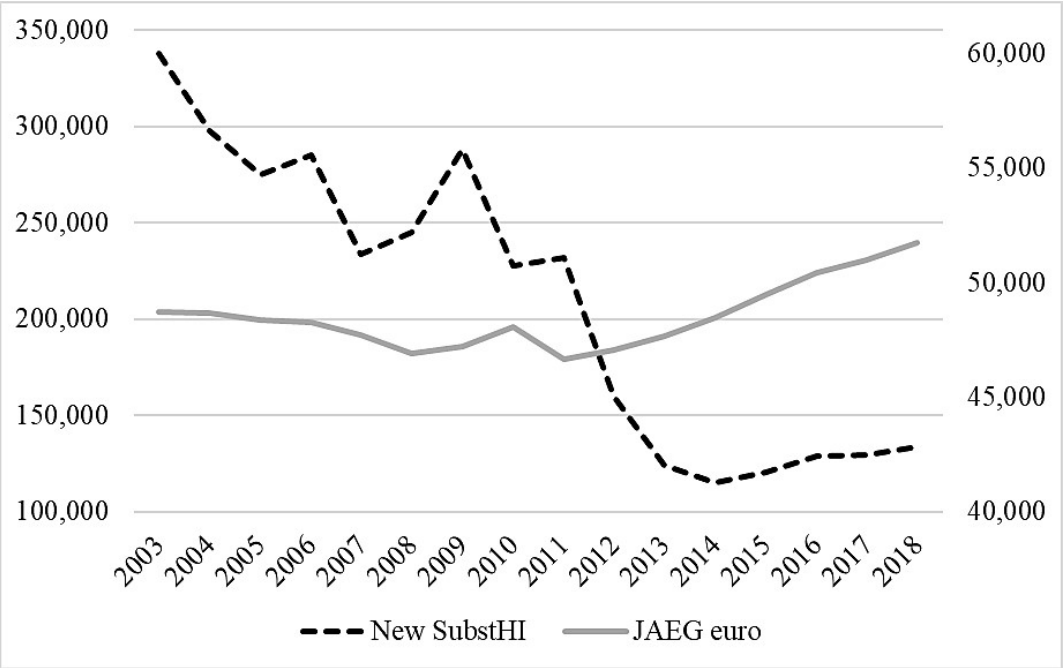


FIGURE A.1: Development of inflation adjusted JAEG and new SubstHI business

Note: Shows the development of the inflation adjusted JAEG (shown in solid grey) and the development of new SubstHI business (shown in dotted black). Data are from the website of the German Association of Private Insurers ("PKV Zahlenportal") Although the drop in New SubstHI appears to coincide with an increase in the slope of the JAEG, the two variables are not highly correlated (overall correlation coefficient of -0.43). More importantly, from an institutional perspective, changes in the JAEG are simply intended to keep up with wage growth and the calculation method was not affected by the 2012 reform. Furthermore, any remaining effect of the JAEG on New SubstHI should also controlled for by the two-way fixed effects model.

A.2 New SubstHI contracts in the market

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Total	534.6	517.2	566.1	498	488.3	413.2	316.4	267.8	265.4	280.2	280.6	284.6
Reported	302.5	353.7	410.9	331.2	327	283.2	219.1	198	198.1	220	222.6	221
Percent Reported	56.60%	68.40%	72.60%	66.50%	67.00%	68.50%	69.30%	74.00%	74.70%	78.00%	79.30%	77.60%

TABLE A.1: New SubstHI contracts in the market (in thousands)

Note: Reports the total new SubstHI contracts in the market as well as the number of new SubstHI contracts that are reported (i.e. that are available for the empirical analysis of this paper). We are able to observe between approximately 60 percent and up to nearly 80 percent of total new contracts in the market.

A.3 Reporting vs. non-reporting firms pre-reform

	Non-reporting					Reporting				
	Mean	Std. Dev.	Min	Max	*n*	Mean	Std. Dev.	Min	Max	*n*
Total Premiums[1]	867.6	1239.7	5.8	4580.2	14	1004	1022.6	34.7	4512.3	20
AC[1]	68.5	92.7	0.6	338.8	14	78.9	68	3.4	285.2	20
Relative AC (original)	0.1	0.05	0.02	0.23	14	0.09	0.04	0.02	0.21	20
Total SubstHI[2]	216.1	291	0.8	911.3	14	315.4	457.8	8.7	2149	20
Total SuppHI[2]	542.4	776.2	0.1	3449.2	14	631.2	502.8	71.2	1717.6	20
New SubstHI[2]	-	-	-	-	0	18.58	19.86	0.61	87.24	20
New SuppHI[2]	-	-	-	-	0	51.83	40.13	3.17	209.37	20
Stock	0.82	0.39	0	1	14	0.68	0.47	0	1	20
Treatment group	0.24	0.43	0	1	14	0.33	0.47	0	1	20
AC per new HI[3]	-	-	-	-	0	1.35	0.91	0.36	4.44	20
Loss Ratio	60.2	13.3	30.7	83.5	14	60.8	9.7	41.6	77.1	20

[1] in million Euros

[2] in thousands

[3] in thousand Euros

TABLE A.2: Reporting vs. non-reporting firms pre-reform

Note: displays an analysis of firms that report new business and firms who do not. Data are from the website of the German Association of Private Insurers ("PKV Zahlenportal") as well as Beenken (2011). Total premiums and total AC are inflation adjusted with the German CPI using 2007 as a baseline year. Table is based on the full sample. Some firms began reporting new business data after 2007 and therefore have firm-year observations in both non-reporting and reporting.

A.4 Summary statistics – pre-reform by business strategy

	I-Type					M-Type					T-Type				
	Mean	Std. Dev.	Min	Max	*n*	Mean	Std. Dev.	Min	Max	*n*	Mean	Std. Dev.	Min	Max	*n*
Tot. Premiums	589.29	514.75	34.69	2077.94	10	1529.32	1169.79	187.99	4580.15	10	739.82	1255.14	5.83	4512.29	10
Loss Ratio	0.56	0.06	0.45	0.67	10	0.67	0.08	0.5	0.77	10	58.81	14.83	30.7	83.5	10
Tot. AC*	57.69	67.42	3.43	285.17	10	113.05	72.82	9.83	302.52	10	40.86	55.91	0.6	195.88	10
RelAC	0.09	0.03	0.02	0.15	10	0.1	0.04	0.06	0.21	10	0.09	0.06	0.02	0.23	10
Tot. SubstHI+	147.34	132.29	8.71	508.99	10	373.83	271.59	34.68	911.3	10	313.4	617.06	0.79	2148.96	10
Tot. SuppHI+	401.56	370.71	71.17	1332.44	10	1018.64	747.68	126.47	3449.22	10	373.56	442.59	0.08	1579.59	10
New SubstHI+	13.06	17.08	0.61	69.95	8	19.13	10.84	5.02	45.82	7	26.55	29.51	2.58	87.24	6
New SuppHI+	35.46	31.12	3.17	106.23	7	66.3	43.23	5.16	209.37	7	54.44	40.33	10.64	174.63	6
Stock	0.8	0.41	0	1	10	0.6	0.5	0	1	10	0.8	0.41	0	1	10

* in million Euros

+ in thousands

TABLE A.3: Summary statistics – pre-reform by business strategy

Note: displays an analysis of pre-reform data by business strategy type. Data are from the website of the German Association of Private Insurers ("PKV Zahlenportal") as well as Beenken (2011). Total premiums and total AC are inflation adjusted with the German CPI using 2007 as a baseline year. Table is based on the full sample.

A.5 Matched sample DD regression

	(1) Annual premiums (millions)	(2) Total AC (millions)	(3) RelAC	(4) No. new SubstHI contracts (thousands)	(5) No. new SuppHI contracts (thousands)	(6) AC per new HI (thousands)
DD 2007	15.57	-4.075	0.00312	1.619	10.92	-0.364
	(132.6)	(18.33)	(0.0101)	(7.486)	(21.85)	(1.897)
DD 2008	149.9	5.106	-0.00226	2.956	1.011	-0.361
	(132.6)	(18.33)	(0.0101)	(7.337)	(21.42)	(1.860)
DD 2009	159.4	11.29	0.00267	1.000	0.209	-0.148
	(132.6)	(18.33)	(0.0101)	(7.228)	(21.10)	(1.832)
DD 2011	-0.539	12.82	0.00443	-2.203	-3.768	0.0875
	(132.6)	(18.33)	(0.0101)	(7.591)	(22.16)	(1.924)
DD 2012	0.820	7.618	-0.00504	-3.821	12.76	-0.199
	(132.6)	(18.33)	(0.0101)	(7.591)	(22.16)	(1.924)
DD 2013	15.89	0.432	-0.0154	-8.413	8.148	-0.391
	(132.6)	(18.33)	(0.0101)	(7.694)	(22.46)	(1.950)
DD 2014	1.085	-6.893	-0.0168*	-8.171	6.241	-0.199
	(132.6)	(18.33)	(0.0101)	(7.591)	(22.16)	(1.924)
DD 2015	9.671	-6.960	-0.0174*	-9.553	4.603	-0.147
	(132.6)	(18.33)	(0.0101)	(7.591)	(22.16)	(1.924)
DD 2016	24.06	-7.140	-0.0193*	-9.918	5.341	2.602
	(132.6)	(18.33)	(0.0101)	(7.366)	(21.81)	(1.867)
DD 2017	-141.9	-5.984	-0.0142	-11.13	20.77	-0.589
	(132.6)	(18.33)	(0.0101)	(7.366)	(21.50)	(1.867)
DD 2018	-131.4	-4.613	-0.0125	-9.080	16.11	-0.625
	(132.6)	(18.33)	(0.0101)	(7.366)	(21.50)	(1.867)
Obs.	180	180	180	125	124	125
R-squared	0.535	0.119	0.460	0.435	0.416	0.180
No. Comp	15	15	15	13	13	13
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

(Standard errors shown in parentheses) $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE A.4: Matched sample DD regression
Note: displays the results of the matched sample differences-in-differences model including the variable AC per new HI

A.6 DD regression without fixed effects

	(1) Annual premiums (millions)	(2) Total AC (millions)	(3) RelAC	(4) No. new SubstHI contracts (thousands)	(5) No. new SuppHI contracts (thousands)	(6) AC per new HI (thousands)
DD 2007	-147.5*	-14.64	-0.008	-1.272	38.88***	-0.696
	-76.11	-9.664	-0.008	-3.85	-14.47	-1.389
DD 2008	-21.41	-3.282	-0.007	1.324	16.64	-0.605
	-76.11	-9.664	-0.008	-3.713	-13.97	-1.341
DD 2009	17.28	5.258	0.000	2.697	12.1	-0.375
	-76.11	-9.664	-0.008	-3.611	-13.6	-1.305
DD 2011	12.38	5.946	-0.003	-1.431	-6.101	0.0521
	-76.11	-9.664	-0.008	-3.648	-13.78	-1.323
DD 2012	18.01	-3.766	-0.011	-4.654	2.077	-0.521
	-76.11	-9.664	-0.008	-3.648	-13.95	-1.323
DD 2013	14.51	-12.85	-0.012	-10.17***	-2.021	-1.792
	-76.11	-9.664	-0.008	-3.724	-14.35	-1.349
DD 2014	-0.818	-20.19**	-0.0144*	-11.07***	-6.342	-0.2
	-76.11	-9.664	-0.008	-3.586	-13.5	-1.298
DD 2015	7.231	-17.73*	-0.013	-11.68***	-9.234	-0.0899
	-76.11	-9.664	-0.008	-3.571	-13.44	-1.291
DD 2016	26.25	-15.1	-0.011	-10.88***	-4.954	2.597**
	-76.11	-9.664	-0.008	-3.56	-13.76	-1.286
DD 2017	2.111	-13.32	-0.00789	-11.19***	6.516	-0.597
	-76.11	-9.664	-0.008	-3.56	-13.39	-1.286
DD 2018	5.971	-12.19	-0.00484	-10.14***	8.869	-2.196*
	-76.11	-9.664	-0.008	-3.56	-13.39	-1.286
Obs.	360	360	360	241	225	241
R-squared	0.351	0.1	0.42	0.423	0.266	0.168
No. Comp	30	30	30	25	24	25
Controls	No	No	No	No	No	No
Fixed Effects	No	No	No	No	No	No

(Standard errors shown in parentheses) $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE A.5: DD regression without fixed effects
Note: displays the results of the differences-in-differences model without fixed effects including the variable AC per new HI.

A.7 Summary statistics – 2010

	Untreated					Treated				
	Mean	Std. Dev.	Min	Max	*n*	Mean	Std. Dev.	Min	Max	*n*
Total Premiums*	870.43	1317.86	8.66	4580.15	21	1399.92	898.04	274.90	3074.96	9
Loss Ratio	59.39	11.46	30.70	75.70	21	64.02	8.32	55.60	77.10	9
Total AC*	57.61	82.57	0.64	338.82	21	128.29	71.64	34.12	258.20	9
RelAC	0.08	0.04	0.02	0.15	21	0.10	0.05	0.06	0.21	9
Total SubstHI+	261.57	483.62	1.04	2148.96	21	358.87	243.67	45.28	737.57	9
Total SuppHI+	567.41	797.03	0.14	3449.22	21	860.45	487.71	229.41	1717.60	9
New SubstHI+	14.50	23.34	0.61	80.92	11	24.37	15.95	6.55	45.99	7
New SuppHI+	33.95	24.01	3.17	80.96	10	51.04	28.43	7.17	86.93	7
Stock	0.71	0.46	0.00	1.00	21	0.78	0.44	0.00	1.00	9
AC per new HI3	1.30	0.90	0.46	3.33	11	2.27	1.10	1.28	4.61	7

* in million Euros

+ in thousands

TABLE A.6: Summary statistics – 2010

Note: displays summary statistics for the 2010, which is the baseline year before the reform. Data are from the website of the German Association of Private Insurers ("PKV Zahlenportal") as well as Beenken (2011). Total premiums and total AC are inflation adjusted with the German CPI using 2007 as a baseline year. Table is based on the full sample.

Appendix B

Appendix to Chapter 4

B.1 Summary of firm performance literature

Author	Year	*n*	Data	Methodology	Findings
Powell & Ansic	1997	Unspecified	Experiment	Psychology experiment	Females more risk averse
Shrader et al.	1997	Unspecified	EEOC	OLS	Mixed results
Atkinson et al.	2003	72 female	Principia	One & Three index models	Male & female similar behavior
Erhardt et al.	2003	112 firms	Custom (Fortune)	Correlation & Regression	Diversity improves performance
Wolfers	2006	64 female	S&P	Portfolio analysis	No difference
Smith et al.	2006	2500 firms	Statistics Denmark	Panel regression	Positive effect of female leadership
Dezső & Ross	2008	Firm-years, unspecified	S&P	Regression	Males overconfident relative to females
Krishnan & Parsons	2008	353	Fortune 500	Regressions, measures of earnings, conservatism	Earnings quality positively related to diversity
Peni & Vähämaa	2010	215	S&P 500	Dechow & Dichev (2002) model, regression	Female executives related to conservative financial reporting
Kolev	2012	64 female	S&P	Portfolio method, risk-adjusted returns	Females underperform relative to risk
Huang & Kisgen	2013	116 female	S&P	Measures of risk aversion/confidence	Males overconfident relative to females
Khan & Vieito	2013	141 female	S&P	OLS	Females risk averse, better firm performance
Lam et al.	2013	196	Shanghai Shenzen exchanges	Panel regression	Mixed results

Continued on next page

Table B.1 – continued from previous page

Author	Year	n	Data	Methodology	Findings
Faccio et al.	2016	Firm-years, unspecified	Amadeus 250,000	Use a variety of matching types	Females more risk-averse
Amore & Garofalo	2016	Firm-years, unspecified	S&P	OLS	Females more risk averse

TABLE B.1: Summary of firm performance literature

B.2 Summary of existing event study results

Author	Year	Female *n*	Data	Methodology	Findings
Lee & James	2007	17 CEOs	S&P	Event study: standard, convenience sample	Negative reaction to female CEO
Gondhalekar & Dalmia	2007	50 CEOs	Russell 3.5000	Event study: Fama-French 3.5-factor model, random sample	No difference
Martin et al.	2009	70 CEOs	S&P	Event study: single index market model, matched sample	No difference
Coxbill et al.	2009	33 CEOs	S&P	Event study: market, market adjusted, & Fama-French 3.5-factor model	No difference
Cook & Glass	2011	200 C-suite	S&P	Event study: random matched sample, OLS	Mixed results depending on industry
Lucey & Carron	2011	77 exec. directors	FTSE	Event study: CAPM, random sample	Negative reaction to females

TABLE B.2: Summary of existing event studies

B.3 Variable definitions

Term	Source	Definition
age	EXECUCOMP	Age as reported, year of appointment
all other compensation	EXECUCOMP	All other unspecified compensation, including signing bonuses, life insurance premiums, debt forgiveness, 1501k contributions. Valued in thousands
bonus	EXECUCOMP	Bonus earned during the fiscal year, in thousands, year of appointment
cash-to-total assets	COMPUSTAT	Ratio of cash to total assets
debt-to-asset ratio	COMPUSTAT	Ratio of total debt to total assets
executive director	EXECUCOMP	Dummy variable - serves as an executive director (1) otherwise (0)
insider	EXECUCOMP	Dummy variable – if an insider (1) otherwise (0)
invested capital to assets	COMPUSTAT	ICAPT to total assets
log market value	COMPUSTAT	Natural logarithm of the market value of the firm's equity
market-to-book	COMPUSTAT	Ratio of common equity to the market value
other compensation	EXECUCOMP	Other compensation benefits like tax reimbursements, discounted stock purchases, gross ups, etc. valued in thousands
research and development investment	COMPUSTAT	Research and development expense to total assets
salary	EXECUCOMP	Base salary, in thousands, at the year of appointment
total compensation	EXECUCOMP	Total Compensation (Salary + Bonus + Other Annual + Restricted Stock Grants + LTIP Payouts + All Other + Value of Options) in thousands

TABLE B.3: Variable definitions

B.4 Return behavior around event (day 0), study 1

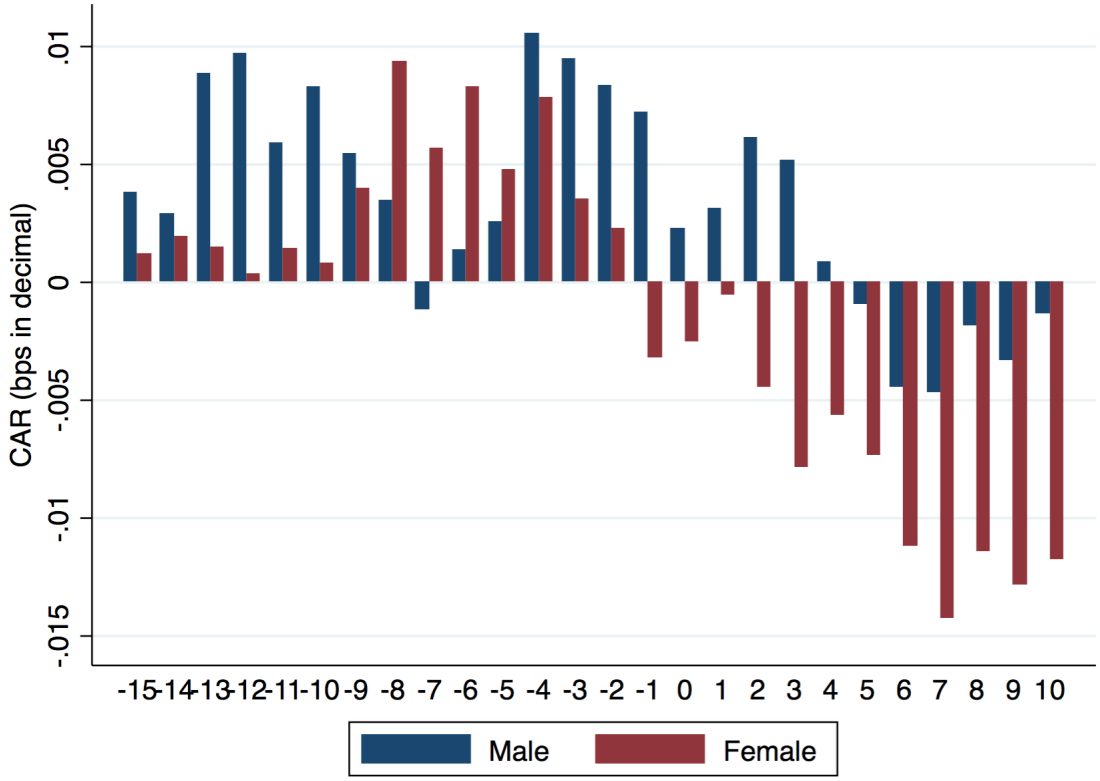


FIGURE B.1: Return behavior around event (day 0), study 1

Appendix C

Appendix to Chapter 5

C.1 Results with different sub-samples

	sentiment $>= 0$	sentiment > 0	voluntary retirements
Sentiment	0.0067^{***}	0.0045^{*}	0.0085^{***}
	(0.0023)	(0.0026)	(0.0027)
MidCap	-0.0186^{***}	-0.0199^{***}	-0.0145^{**}
	(0.0061)	(0.0063)	(0.0071)
SmallCap	-0.0057	-0.0080	-0.0039
	(0.0057)	(0.0059)	(0.0068)
Insider	-0.0008	-0.0018	-0.0020
	(0.0055)	(0.0056)	(0.0070)
Female	-0.0038	0.0007	-0.0121
	(0.0090)	(0.0094)	(0.0106)
MBA reported	0.0135^{*}	0.0142^{*}	0.0083
	(0.0076)	(0.0076)	(0.0093)
Age reported	-0.0037	-0.0036	-0.0057
	(0.0053)	(0.0054)	(0.0063)
Involuntary - CEO death	0.0143	0.0252	
	(0.0216)	(0.0243)	
Involuntary - CEO Illness	-0.0044	-0.0039	-
	(0.0135)	(0.0152)	
Involuntary - Job performance	-0.0120^{*}	-0.0128^{*}	-
	(0.0064)	(0.0066)	
Involuntary - Personal issues	-0.0003	-0.0007	-
	(0.0202)	(0.0203)	
Departure - Other	0.0180^{*}	0.0181^{*}	-
	(0.0094)	(0.0098)	
Voluntary - New opportunity	-0.0167	-0.0183	-
	(0.0152)	(0.0157)	
R^2	0.0541	0.0571	0.0431
Adj. R^2	0.0312	0.0330	0.0209
Num. obs.	721	683	485
Controls	Yes	Yes	Yes

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE C.1: Results with different sub-samples

Note: Shows robustness of results to other subsamples of the data.

C.2 Anonymized excerpt of a positive sentiment announcement

[Company] announced today that [new CEO] has been named the company's next chief executive officer, effective Jan. 1, 2015. He was also elected to the company's board of directors, effective immediately. [New CEO] currently serves as president and chief operating officer responsible for leading and running the company's global operations.

...

"I am honored to have the opportunity to lead our great company and team into the future," [New CEO] said. "Under previous leadership, we have made exceptional progress creating a high-performance, results-focused company. My commitment is to build on that success by accelerating our pace of progress. This is an exciting time, with tremendous opportunities, and all of us at [Company] are committed to leveraging our core model to achieve new levels of value-creation. I look forward to working with my colleagues to extend our more than 100-year record of success."

C.3 Anonymized excerpt of a neutral sentiment announcement

[Company] said early Wednesday that president and CEO [previous CEO] will serve as executive chairman and that chief operating officer [new CEO] will assume the role of president and CEO. ...

[New CEO] joined [company] in 1994 and served as chief administrative officer, senior vice president, general counsel and secretary before being promoted to chief operating officer.

C.4 Anonymized excerpt of a negative sentiment announcement

[Company] co-founder and Chairman [New CEO] is returning to the helm as the company faces declining subscribers and a changing pay-TV industry.

...

[New CEO] co-founded [Company] in 1980 and has held executive positions including president and CEO during his tenure. [Company] is facing a changing pay-TV landscape as more viewers stream services like [Competitor 1] and [Competitor 2] and cable channels like [Competitor 3] begin to offer standalone streaming services for its own programming.

C.5 Results when balanced and neutral are differentiated

	Model 4	Model 4B	Model 4N	Model 5	Model 5B	Model 5N
Negative	0.0223	0.0213	0.0246			
	(0.0184)	(0.0196)	(0.0240)			
Positive	0.0354***	0.0344***	0.0377**			
	(0.0110)	(0.0129)	(0.0189)			
Neutral		−0.0033				
		(0.0220)				
		(0.0129)	(0.0189)			
Balanced			0.0033			
			(0.0220)			
Sentiment < 0				0.0232	0.0256	0.0221
				(0.0184)	(0.0239)	(0.0196)
- 0 to 0.9				0.0369***	0.0393*	0.0358**
				(0.0130)	(0.0201)	(0.0147)
- 1 to 1.9				0.0296***	0.0320*	0.0285**
				(0.0113)	(0.0190)	(0.0132)
- 2 to 2.9				0.0424***	0.0449**	0.0413***
				(0.0118)	(0.0193)	(0.0135)
- 3 or higher				0.0404***	0.0428**	0.0393***
				(0.0119)	(0.0194)	(0.0137)
- balanced					0.0035	
					(0.0220)	
- neutral						−0.0035
						(0.0220)
R^2	0.0555	0.0555	0.0555	0.0623	0.0624	0.0624
Adj. R^2	0.0319	0.0305	0.0305	0.0348	0.0335	0.0335
Num. obs.	738	738	738	738	738	738

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE C.2: Results with neutral and balanced differentiated

Note: Shows robustness of results to when balanced and neutral sentiments are treated as distinct from one another. Model 4 and 5 are the same as Model 4 and 5 in Table 5.7. The suffix B indicates that "balanced" represents the baseline group, and the suffix N indicates that "neutral" represents the baseline group.

C.6 Histogram of age if age reported

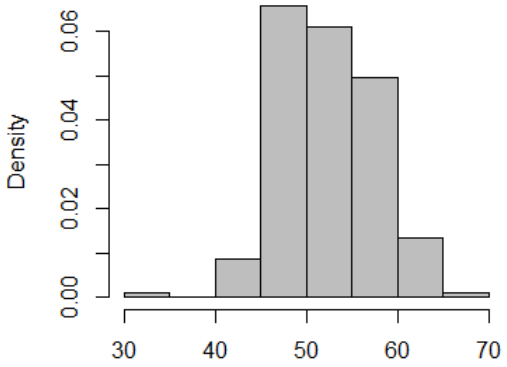


FIGURE C.1: Distribution of the variable age when age is reported

C.7 Summary statistics by MBA reporting

Variable	N	Mean	SD	Min	Max
	MBA not reported				
CAR	656	-0.00313	0.06342	-0.26533	0.30261
Sentiment	656	1.95059	1.13278	-1.60944	4.61512
Gender	656				
... Male	606	92.378%			
... Female	50	7.622%			
ROE	656	0.11165	0.39749	-3.61598	4.19044
Ann. length	656	318.09451	169.79452	25	1234
Turnover reason	656				
... Voluntary - Retirement	434	66.159%			
... Involuntary - CEO death	10	1.524%			
... Involuntary - CEO illness	21	3.201%			
... Involuntary - job performance	114	17.378%			
... Involuntary - personal issues	9	1.372%			
... Other	52	7.927%			
... Voluntary - New opportunity	16	2.439%			
	MBA reported				
CAR	82	0.00971	0.06237	-0.15218	0.22521
Sentiment	82	2.51134	0.96714	0.25131	4.46591
Gender	82				
... Male	78	95.122%			
... Female	4	4.878%			
ROE	82	0.12372	0.17254	-0.45343	0.79539
Ann. length	82	432.93902	178.70573	191	1366
Turnover reason	82				
... Voluntary - Retirement	51	62.195%			
... Involuntary - CEO death	0	0%			
... Involuntary - CEO illness	2	2.439%			
... Involuntary - job performance	19	23.171%			
... Involuntary - personal issues	1	1.22%			
... Other	6	7.317%			
... Voluntary - New opportunity	3	3.659%			

TABLE C.3: Summary statistics by MBA reporting

C.8 Summary statistics by age reporting

Variable	N	Mean	SD	Min	Max
	Age not reported				
CAR	528	-0.0005	0.0676	-0.2653	0.3026
Sentiment	528	1.9437	1.1901	-1.6094	4.6151
Gender	528				
... Male	485	91.86%			
... Female	43	8.14%			
ROE	528	0.0912	0.3857	-3.616	4.1904
Ann. length	528	313.9886	175.3479	25	1234
departure_code	528				
... Voluntary - Retirement	345	65.34%			
... Involuntary - CEO death	8	1.52%			
... Involuntary - CEO illness	16	3.03%			
... Involuntary - job performance	88	16.67%			
... Involuntary - personal issues	8	1.52%			
... Other	49	9.28%			
... Voluntary - New opportunity	14	2.65%			
	Age reported				
CAR	210	-0.0048	0.0513	-0.1696	0.1766
Sentiment	210	2.1868	0.9381	-0.619	4.4659
Gender	210				
... Male	199	94.76%			
... Female	11	5.24%			
ROE	210	0.1679	0.3568	-0.8127	3.2289
Ann. length	210	373.2619	165.1512	40	1366
Turnover reason	210				
... Voluntary - Retirement	140	66.67%			
... Involuntary - CEO death	2	0.95%			
... Involuntary - CEO illness	7	3.33%			
... Involuntary - job performance	45	21.43%			
... Involuntary - personal issues	2	0.95%			
... Other	9	4.29%			
... Voluntary - New opportunity	5	2.38%			

TABLE C.4: Summary statistics by age reporting

C.9 Results with different event windows

	Model 3 -2/+2	Model 3 -3/+3	Model 3 -1/+2	Model 3 -1/+3
Sentiment	0.0054***	0.0060***	0.0049**	0.0048**
	(0.0020)	(0.0022)	(0.0019)	(0.0020)
MidCap	−0.0127**	−0.0176***	−0.0107**	−0.0124**
	(0.0057)	(0.0057)	(0.0052)	(0.0052)
SmallCap	−0.0043	−0.0057	−0.0063	−0.0090
	(0.0062)	(0.0061)	(0.0057)	(0.0057)
Involuntary - CEO death	0.0090	0.0122	0.0069	0.0077
	(0.0180)	(0.0160)	(0.0182)	(0.0182)
Involuntary - CEO Illness	−0.0069	−0.0050	−0.0033	−0.0055
	(0.0096)	(0.0112)	(0.0087)	(0.0112)
Involuntary - Job performance	−0.0113*	−0.0124*	−0.0122**	−0.0109*
	(0.0063)	(0.0068)	(0.0058)	(0.0062)
Involuntary - Personal issues	0.0120	−0.0006	0.0134	0.0086
	(0.0202)	(0.0228)	(0.0177)	(0.0186)
Departure - Other	0.0182*	0.0172	0.0101	0.0127
	(0.0110)	(0.0110)	(0.0096)	(0.0101)
Voluntary - New opportunity	−0.0136	−0.0139	−0.0067	−0.0038
	(0.0101)	(0.0105)	(0.0080)	(0.0070)
Insider	0.0001	−0.0007	−0.0021	−0.0022
	(0.0054)	(0.0059)	(0.0050)	(0.0053)
Female	0.0008	−0.0042	−0.0023	−0.0074
	(0.0090)	(0.0087)	(0.0078)	(0.0081)
MBA reported	0.0112*	0.0136*	0.0120**	0.0145**
	(0.0067)	(0.0076)	(0.0060)	(0.0067)
Age reported	0.0007	−0.0031	0.0008	−0.0021
	(0.0040)	(0.0046)	(0.0037)	(0.0041)
R^2	0.0486	0.0504	0.0457	0.0482
Adj. R^2	0.0262	0.0279	0.0231	0.0257
Num. obs.	738	738	738	738

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE C.5: Results with different event windows

Note: Shows robustness of results to different specifications of the event window. Model 3 (full specification) is used in each case.

C.10 Results with CAPM model of returns

	CAPM
Sentiment	0.0034*
	(0.0020)
MidCap	−0.0117**
	(0.0050)
SmallCap	−0.0077
	(0.0056)
Involuntary - CEO death	0.0028
	(0.0185)
Involuntary - CEO Illness	−0.0013
	(0.0064)
Involuntary - Job performance	−0.0151**
	(0.0060)
Involuntary - Personal issues	0.0017
	(0.0141)
Departure - Other	0.0079
	(0.0102)
Voluntary - New opportunity	−0.0137
	(0.0102)
Insider	−0.0035
	(0.0050)
Female	0.0020
	(0.0078)
MBA reported	0.0126**
	(0.0058)
Age reported	0.0032
	(0.0037)
R^2	0.0493
Adj. R^2	0.0269
Num. obs.	738

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE C.6: Results with CAPM model of returns

Note: Shows robustness of results to use of CAPM model for estimating abnormal returns.

C.11 Density functions of covariates and CAR

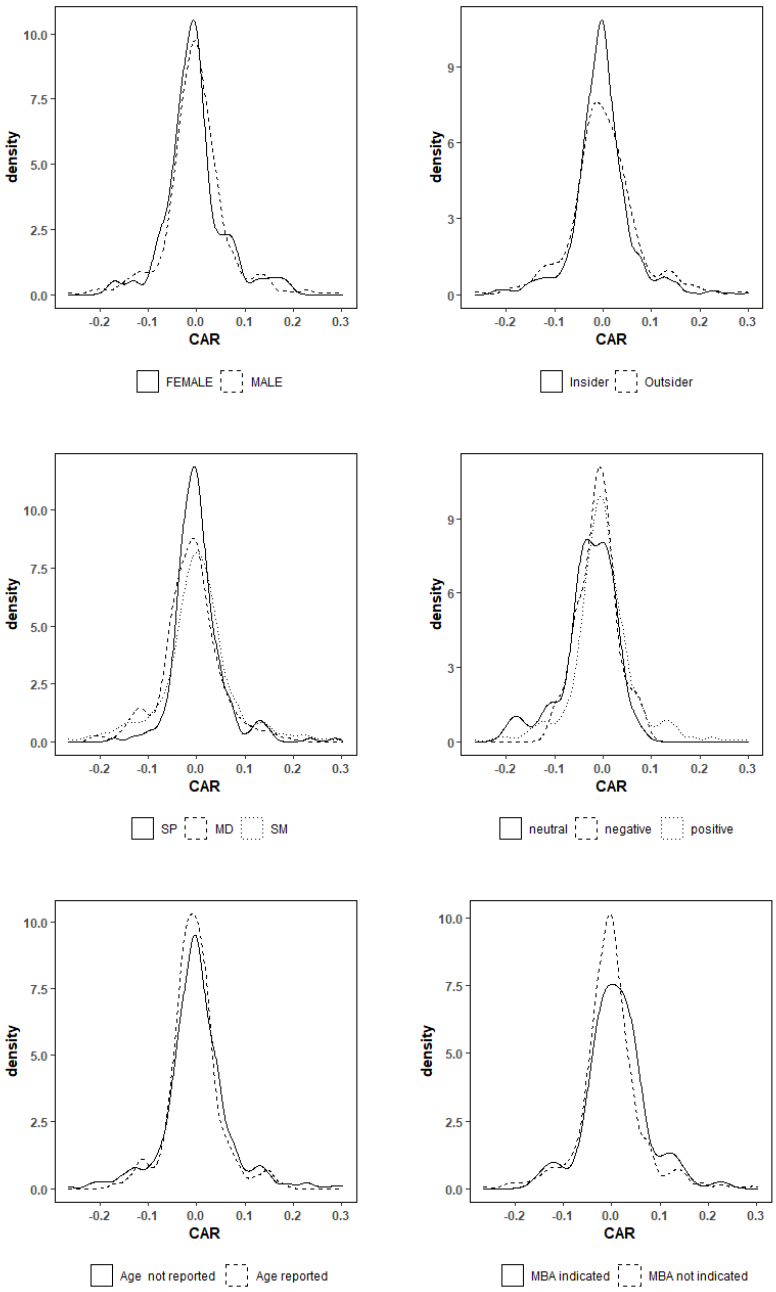


FIGURE C.2: CAR density functions by covariates

C.12 Anonymized version of "passing the torch"

“[Company] announced today that its Board of Directors has elected [new CEO], 51, to succeed [previous CEO] as president and chief executive officer (CEO).

[New CEO], who currently serves as a [Company] group president, joined [Company] in 1998 as vice president of manufacturing for [subsidiary]. In 2007, Mr. Allman became president of [subsidiary]. In 2011, he was promoted to group president and has responsibility for [Company’s] Plumbing and North American Cabinet businesses. He holds a Mechanical Engineering degree from General Motors Institute (Kettering University) and an MBA from the University of Michigan.

[New CEO] will replace [previous CEO], 65, who is retiring after a 37-year career with [Company]. [Previous CEO], only the third CEO in Masco’s 85-year history, was elected CEO in July 2007. [Previous CEO], will complete his current term on [Company’s] Board of Directors.

"I am pleased and excited that [new CEO] will succeed me as [Company’s] president and CEO," said [previous CEO]. "[New CEO] has a demonstrated track record of results reflecting his strong leadership and operational skills. [New CEO] has not only played an integral role in developing our strategies to strengthen our brands and improve our execution, he has also led large, complex businesses in executing against these strategies for growth and value creation. I have worked closely with [New CEO] and our leadership team over the past several years and we have established the foundation on which [Company] can build a successful future."

Chairman of the Board of Directors, stated, "The board is delighted to elect [New CEO] as [Company’s] president and CEO. With his knowledge of [Company] and its people, his operational expertise, his customer relationships and his skill in leading large organizations, we are confident that this will be a seamless transition. We strongly believe that [new CEO], leading the outstanding management team that we have across the enterprise, will continue to drive value creation for our shareholders."

The Chairman continued, "On behalf of the board and [Company’s] employees, we thank [previous CEO] for his tireless commitment and numerous contributions to [Company’s] success during his 37-year career. During his tenure with the company, [previous CEO] has proven himself to be a strong and stable leader during periods of growth as well as during our more challenging times. His leadership, including his courage and positive energy, has guided us through difficult times and has positioned [Company] to take advantage of the recent improvement in housing dynamics as evidenced by our 2012 and 2013 performance. We also appreciate his willingness to continue to lead [Company] past age 65, as our board worked on his succession. It has been a great relationship and we wish [previous CEO] an enjoyable and rewarding retirement."

C.13 Results with different specifications of sentiment

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Sentiment score	0.0062***	0.0097*					
	(0.0022)	(0.0052)					
Sentiment squared		−0.0009					
		(0.0013)					
Pos. words / (Pos. + Neg. Words)			0.0003**				
			(0.0001)				
Pos. words / All words				0.0029**			
				(0.0013)			
Pos. - Neg. Words					0.0004		
					(0.0003)		
(Pos. - Neg. Words) / All words						0.3299***	
						(0.1147)	
(Pos. - Neg. Words) / (Pos. + Neg. Words)							0.0151**
							(0.0070)
R^2	0.0522	0.0527	0.0501	0.0477	0.0431	0.0517	0.0501
Adj. R^2	0.0298	0.0290	0.0273	0.0252	0.0206	0.0293	0.0273
Num. obs.	738	738	726	738	738	738	726

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

TABLE C.7: Results with different specifications of sentiment

Note: Shows robustness of results to different specifications of sentiment.

References

Adams, Susan M, Atul Gupta, and John D Leeth (2009). "Are female executives overrepresented in precarious leadership positions?" In: *British Journal of Management* 20.1, pp. 1–12.

Amore, Mario Daniele and Orsola Garofalo (2016). "Executive gender, competitive pressures, and corporate performance". In: *Journal of Economic Behavior & Organization* 131, pp. 308–327.

Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Dr. Gerhard Schick, Dr. Danyal Bayaz, Lisa Paus, weiterer Abgeordneter und der Fraktion BÜNDNIS 90/DIE GRÜNEN (2018). Drucksache 19/3802.

Argenti, Paul A (1996). "Corporate communication as a discipline: toward a definition". In: *Management Communication Quarterly* 10.1, pp. 73–97.

Asher, A (2006). "Unfinished Accounting Issues: Incorporating Fair Value and Prudence in Accounting Theory". In: *Annals of Actuarial Science* 1.2, pp. 271–290.

Atal, Juan Pablo et al. (2019). "Exit, Voice, or Loyalty? An Investigation Into Mandated Portability of Front-Loaded Private Health Plans". In: *Journal of Risk and Insurance* 86.3, pp. 697–727.

Atiase, Rowland Kwame (1985). "Predisclosure information, firm capitalization, and security price behavior around earnings announcements". In: *Journal of Accounting Research*, pp. 21–36.

Atkinson, Stanley M, Samantha Boyce Baird, and Melissa B Frye (2003). "Do female mutual fund managers manage differently?" In: *Journal of Financial Research* 26.1, pp. 1–18.

Averch, Harvey and Leland L Johnson (1962). "Behavior of the firm under regulatory constraint". In: *The American Economic Review* 52.5, pp. 1052–1069.

Baker, George P. and Brian J. Hall (2004). "CEO incentives and firm size". In: *Journal of Labor Economics* 22.4, pp. 767–798.

Beatty, Randolph P. and Edward J. Zajac (1987). "CEO change and firm performance in large corporations: Succession effects and manager effects". In: *Strategic Management Journal* 8.4, pp. 305–317.

Beenken, Matthias (2011). *Provisionen und Courtagen - was die Versicherer ihren Vermittlern zahlen*. Ahrensburg: VersicherungsVerlag Journal. ISBN: 978-3-938226-23-0.

Beenken, Matthias and Michael Radtke (2015). *Provisionen und Courtagen: was die Versicherer ihren Vermittlern zahlen; Umfrageergebnisse 2015*. Versicherungsjournal-Verlag.

Beenken, Matthias and Michael Radtke (2017). *Provisionen und Courtagen - was die Versicherer ihren Vermittlern zahlen*. Ahrengsburg: VersicherungsJournal Verlag. ISBN: 978-3-938226-59-9.

Benabou, Roland and Guy Laroque (1992). "Using privileged information to manipulate markets: Insiders, gurus, and credibility". In: *The Quarterly Journal of Economics* 107.3, pp. 921–958.

Benoit, Kenneth et al. (2018). "quanteda: An R package for the quantitative analysis of textual data". In: *Journal of Open Source Software* 3.30, p. 774. DOI: https://doi.org/10.21105/joss.00774.

Berger, Allen N, J David Cummins, and Mary A Weiss (1997). "The coexistence of multiple distribution systems for financial services: the case of property-liability insurance". In: *The Journal of Business* 70.4, pp. 515–546.

Bertrand, Marianne and Antoinette Schoar (2003). "Managing with style: The effect of managers on firm policies". In: *The Quarterly journal of Economics* 118.4, pp. 1169–1208.

Bhagat, Sanjai, Brian J. Bolton, and Ajay Subramanian (2010). "CEO education, CEO turnover, and firm performance". In: *Available at SSRN 1670219*.

Boardex (2023). *BoardEx from Wharton Research Data Services*.

Boudt, Kris and James Thewissen (2019). "Jockeying for position in CEO letters: Impression management and sentiment analytics". In: *Financial Management* 48.1, pp. 77–115.

Braegelmann, Kylie A. and Nacasius U. Ujah (2020). "Gender matters: market perception of future performance". In: *Managerial Finance* 46.10, pp. 1247–1262.

Brockett, Patrick L et al. (2005). "Financial intermediary versus production approach to efficiency of marketing distribution systems and organizational structure of insurance companies". In: *Journal of Risk and Insurance* 72.3, pp. 393–412.

Burchard, Cord et al. (2021). "Bad news for announcers, good news for rivals: Are rivals fully seizing transition-period opportunities following announcers' top management turnovers?" In: *Strategic Management Journal* 42.3, pp. 579–607.

Chan, William (1996). "External recruitment versus internal promotion". In: *Journal of Labor Economics* 14.4, pp. 555–570.

Charitou, Melita, Andreas Patis, and Adamos Vlittis (2010). "The market reaction to the appointment of an outside CEO: An empirical investigation". In: *Journal of Economics and International Finance* 2.11, p. 272.

Christensen, Theodore E, Toni Q Smith, and Pamela S Stuerke (2004). "Public predisclosure information, firm size, analyst following, and market reactions to earnings announcements". In: *Journal of Business Finance & Accounting* 31.7-8, pp. 951–984.

Cialdini, Robert B., Carl A. Kallgren, and Raymond R. Reno (1991). "A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior". In: *Advances in Experimental Social Psychology*. Vol. 24. Elsevier, pp. 201–234.

Cialdini, Robert B., Raymond R. Reno, and Carl A. Kallgren (1990). "A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places." In: *Journal of Personality and Social Psychology* 58.6, p. 1015.

Cialdini, Robert B. and Melanie R. Trost (1998). *Social influence: Social norms, conformity and compliance.* McGraw-Hill.

Cicala, Steve, Ethan MJ Lieber, and Victoria Marone (2019). "Regulating markups in US health insurance". In: *American Economic Journal: Applied Economics* 11.4, pp. 71–104.

Clark, Andrew E. (1996). "Job satisfaction in Britain". In: *British Journal of Industrial Relations* 34.2, pp. 189–217.

– (1997). "Job satisfaction and gender: why are women so happy at work?" In: *Labour economics* 4.4, pp. 341–372.

Clark, Andrew E. and Andrew J. Oswald (1996). "Satisfaction and comparison income". In: *Journal of Public Economics* 61.3, pp. 359–381.

Cline, Brandon N and Adam S Yore (2016). "Silverback CEOs: Age, experience, and firm value". In: *Journal of Empirical Finance* 35, pp. 169–188.

Compustat (2020). *Compustat Industrial [Annual Data]*. Wharton Research Data Service.

Cook, Alison and Christy Glass (2011). "Leadership change and shareholder value: How markets react to the appointments of women". In: *Human Resource Management* 50.4, pp. 501–519.

– (2014). "Women and top leadership positions: Towards an institutional analysis". In: *Gender, Work & Organization* 21.1, pp. 91–103.

Coxbill, Amanda L., Lee W. Sanning, Sherrill Shaffer, et al. (2009). *Market reaction to the announcement of a male-to-female CEO turnover*. Tech. rep. Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy ...

CRSP (2020). *CRSP Stocks*. Center for Research in Sercurity Prices, Wharton Research Data Services.

Dalton, Dan R. and Idalene F. Kesner (1983). "Inside/outside succession and organizational size: The pragmatics of executive replacement". In: *Academy of Management Journal* 26.4, pp. 736–742.

d'Ambrosio, Conchita, Andrew E. Clark, and Marta Barazzetta (2018). "Unfairness at work: Well-being and quits". In: *Labour Economics* 51, pp. 307–316.

Danzon, Patricia M. and Scott E. Harrington (2001). "Workers' compensation rate regulation: How price controls increase costs". In: *The Journal of Law and Economics* 44.1, pp. 1–36.

Davison, Heather K. and Michael J. Burke (2000). "Sex discrimination in simulated employment contexts: A meta-analytic investigation". In: *Journal of Vocational Behavior* 56.2, pp. 225–248.

Demers, Elizabeth and Clara Vega (2014). "The impact of credibility on the pricing of managerial textual content". In: *Available at SSRN 1153450*.

Dezső, Cristian L. and David Gaddis Ross (2008). "Girl Power: Female participation in top management and firm performance". In: *University of Maryland Robert H Smith School of Business*.

Dixon-Fowler, Heather R., Alan E. Ellstrand, and Jonathan L. Johnson (2013). "Strength in numbers or guilt by association? Intragroup effects of female chief executive announcements". In: *Strategic Management Journal* 34.12, pp. 1488–1501.

Dye, Ronald A. (2001). "An Evaluation of "Essays on Disclosure" and the Disclosure Literature in Accounting". In: *Journal of Accounting and Economics* 32.1-3, pp. 181–235.

Eagly, Alice H. (1987). *Sex Differences in Social Behavior: A Social-role Interpretation*. Hillsdale, NJ.: LawrenceErlbaum.

Eagly, Alice H and Steven J Karau (2002). "Role congruity theory of prejudice toward female leaders." In: *Psychological Review* 109.3, p. 573.

Eagly, Alice H and Wendy Wood (1991). "Explaining sex differences in social behavior: A meta-analytic perspective". In: *Personality and Social Psychology Bulletin* 17.3, pp. 306–315.

Eagly, Alice H. and Wendy Wood (1999). "The origins of sex differences in human behavior: Evolved dispositions versus social roles." In: *American Psychologist* 54.6, p. 408.

Eagly, Alice .H, Wendy Wood, and Amanda B. Diekman (2000). "Social role theory of sex differences and similarities: A current appraisal". In: *The Developmental Social Psychology of Gender* 12.174, pp. 9781410605245–12.

Eckardt, Martina (2007). *Insurance intermediation: An economic analysis of the information services market*. Springer Science & Business Media.

Elsaid, Eahab and Nancy D Ursel (2018). "Re-examining the glass cliff hypothesis using survival analysis: The case of female CEO tenure". In: *British Journal of Management* 29.1, pp. 156–170.

Engelberg, Joseph (2008). "Costly information processing: Evidence from earnings announcements". In: *AFA 2009 San Francisco meetings paper*.

Erhardt, Niclas L., James D. Werbel, and Charles B. Shrader (2003). "Board of director diversity and firm financial performance". In: *Corporate Governance: an international review* 11.2, pp. 102–111.

EXECUCOMP (2017). Wharton Research Data Service.

Faccio, Mara, Maria-Teresa Marchica, and Roberto Mura (2016). "CEO gender, corporate risk-taking, and the efficiency of capital allocation". In: *Journal of Corporate Finance* 39, pp. 193–209.

Falk, Armin, Ernst Fehr, and Christian Zehnder (2006). "Fairness perceptions and reservation wages—the behavioral effects of minimum wage laws". In: *The Quarterly Journal of Economics* 121.4, pp. 1347–1381.

Fama, Eugene F. (1970). "Efficient capital markets: A review of theory and empirical work". In: *The Journal of Finance* 25.2, pp. 383–417.

Fama, Eugene F. et al. (1969). "The adjustment of stock prices to new information". In: *International Economic Review* 10.1, pp. 1–21.

Farrell, Joseph and Matthew Rabin (1996). "Cheap talk". In: *Journal of Economic Perspectives* 10.3, pp. 103–118.

Farrell, Kathleen A. and David A. Whidbee (2003). "Impact of firm performance expectations on CEO turnover and replacement decisions". In: *Journal of Accounting and Economics* 36.1-3, pp. 165–196.

Federal Financial Supervisory Authority, Germany (2020). *Statistik der BaFin – Krankenversicherungsunternehmen 2020*. Tabelle 460.

Federal Statistical Office, Germany (2023). *Statistisches Bundesamt (Destatis), GenesisOnline*. Online. URL: https://www-genesis.destatis.de/genesis/online?operation=statistic&levelindex=0&levelid=1682586224162&code=61111#abreadcrumb.

Fehr, Ernst and Simon Gächter (2000). "Fairness and retaliation: The economics of reciprocity". In: *Journal of Economic Perspectives* 14.3, pp. 159–182.

Focht, Uwe, Andreas Richter, and Jörg Schiller (2013). "Intermediation and (Mis)-Matching in Insurance Markets—Who Should Pay the Insurance Broker?" In: *Journal of Risk and Insurance* 80.2, pp. 329–350.

Fors Connolly, Filip, Mikael Goossen, and Mikael Hjerm (2020). "Does gender equality cause gender differences in values? Reassessing the gender-equality-personality paradox". In: *Sex Roles* 83, pp. 101–113.

Gaucher, Danielle, Justin Friesen, and Aaron C Kay (2011). "Evidence that gendered wording in job advertisements exists and sustains gender inequality." In: *Journal of Personality and Social Psychology* 101.1, p. 109.

Gender Equality, European Institute for (2017). *Gender Equality Index*. Institute for Gender Equality, European. URL: https://eige.europa.eu/gender-equality-index/about.

Gentry, Richard J. et al. (2021). "A database of CEO turnover and dismissal in S&P 1500 firms, 2000–2018". In: *Strategic Management Journal* 42.5, pp. 968–991.

Gneezy, Uri, Kenneth L Leonard, and John A List (2009). "Gender differences in competition: Evidence from a matrilineal and a patriarchal society". In: *Econometrica* 77.5, pp. 1637–1664.

Gondhalekar, Vijay and Sonia Dalmia (2007). "Examining the stock market response: a comparison of male and female CEOs". In: *International Advances in Economic Research* 13, pp. 395–396.

Graffin, Scott, Mason Carpenter, and Steven Boivie (2011). "What's all that (strategic) noise? Anticipatory impression management in CEO succession". In: *Strategic Management Journal* 32.7, pp. 748–770.

Harrington, Scott E (2004). "Effects of Prior Approval Rate Regulation of Auto". In: *Deregulating Property-Liability Insurance: Restoring Competition and Increasing Market Efficiency* 285.

Haslam, S Alexander and Michelle K Ryan (2008). "The road to the glass cliff: Differences in the perceived suitability of men and women for leadership positions in succeeding and failing organizations". In: *The Leadership Quarterly* 19.5, pp. 530–546.

Hauret, Laetitia and Donald R. Williams (2017). "Cross-national analysis of gender differences in job satisfaction". In: *Industrial Relations: A Journal of Economy and Society* 56.2, pp. 203–235.

Healy, Paul M. and Krishna G. Palepu (2001). "Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature". In: *Journal of Accounting and Economics* 31.1-3, pp. 405–440.

Heilman, Madeline E. (2012). "Gender stereotypes and workplace bias". In: *Research in Organizational Behavior* 32, pp. 113–135.

Heilman, Madeline E., Caryn J. Block, and Peter Stathatos (1997). "The affirmative action stigma of incompetence: Effects of performance information ambiguity". In: *Academy of Management Journal* 40.3, pp. 603–625.

Heilman, Madeline E. and Brian Welle (2006). "Disadvantaged by diversity? The effects of diversity goals on competence perceptions 1". In: *Journal of Applied Social Psychology* 36.5, pp. 1291–1319.

Heilman, Madeline E. et al. (1989). "Has anything changed? Current characterizations of men, women, and managers." In: *Journal of Applied Psychology* 74.6, p. 935.

Hekman, David R. et al. (2017). "Does diversity-valuing behavior result in diminished performance ratings for non-white and female leaders?" In: *Academy of Management Journal* 60.2, pp. 771–797.

Henry, Elaine (2008). "Are investors influenced by how earnings press releases are written?" In: *The Journal of Business Communication* 45.4, pp. 363–407.

Hofmann, Annette and Mark Browne (2013). "One-sided commitment in dynamic insurance contracts: Evidence from private health insurance in Germany". In: *Journal of Risk and Uncertainty* 46, pp. 81–112.

Huang, Jiekun and Darren Kisgen (2013). "Gender and corporate finance: Are male executives overconfident relative to female executives?" In: *Journal of Financial Economics* 108.3, pp. 822–839.

Huang, Xuan, Siew Hong Teoh, and Yinglei Zhang (2014). "Tone management". In: *The Accounting Review* 89.3, pp. 1083–1113.

Inderst, Roman and Marco Ottaviani (2012a). "Competition through commissions and kickbacks". In: *American Economic Review* 102.2, pp. 780–809.

– (2012b). "How (not) to pay for advice: A framework for consumer financial protection". In: *Journal of Financial Economics* 105.2, pp. 393–411.

Jalal, Abu M. and Alexandros P. Prezas (2012). "Outsider CEO Succession and Firm Performance". In: *Journal of Economics and Business* 64.6, pp. 399–426.

Jamieson, Kathleen Hall (1995). *Beyond the double bind: Women and leadership*. Oxford University Press, USA.

Jones, Edward E and Victor A. Harris (1967). "The attribution of attitudes". In: *Journal of Experimental Social Psychology* 3.1, pp. 1–24.

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision Under Risk". In: *Econometrica* 47.2, pp. 263–291.

Kan, Man Yee, Oriel Sullivan, and Jonathan Gershuny (2011). "Gender convergence in domestic work: Discerning the effects of interactional and institutional barriers from large-scale data". In: *Sociology* 45.2, pp. 234–251.

Kanter, Rosabeth Moss (1977). "Some effects of proportions on group life: Skewed sex ratios and responses to token women". In: *American Journal of Sociology* 82.5, pp. 965–990.

Karsten, Margaret Foegen (1994). *Management and Gender: Issues and Attitudes*. Bloomsbury Academic.

Khan, Walayet A and João Paulo Vieito (2013). "CEO gender and firm performance". In: *Journal of Economics and Business* 67, pp. 55–66.

Klumpes, Paul JM (2004). "Performance benchmarking in financial services: evidence from the UK life insurance industry". In: *The Journal of Business* 77.2, pp. 257–273.

Knittel, Christopher R and Victor Stango (2003). "Price ceilings as focal points for tacit collusion: Evidence from credit cards". In: *American Economic Review* 93.5, pp. 1703–1729.

Kolev, Gueorgui I (2012). "Underperformance by female CEOs: A more powerful test". In: *Economics Letters* 117.2, pp. 436–440.

Krishnan, Gopal V. and Linda M. Parsons (2008). "Getting to the bottom line: An exploration of gender and earnings quality". In: *Journal of Business Ethics* 78, pp. 65–76.

Lam, Kevin C.K., Paul B. McGuinness, and João Paulo Vieito (2013). "CEO gender, executive compensation and firm performance in Chinese-listed enterprises". In: *Pacific-Basin Finance Journal* 21.1, pp. 1136–1159.

Lazear, Edward P. (1979). "Why is there mandatory retirement?" In: *Journal of Political Economy* 87.6, pp. 1261–1284.

Lee, Jim (2009). "Does size matter in firm performance? Evidence from US public firms". In: *International Journal of the Economics of Business* 16.2, pp. 189–203.

Lee, P. M. and E.H. James (2007). "She'-e-os: Gender Effects and Investor Reactions to the Announcements of Top Executive Appointments". In: *Strategic Management Journal* 28.3, pp. 227–241.

Leitch, Darren and Mohamed Sherif (2017). "Twitter mood, CEO succession announcements and stock returns". In: *Journal of computational science* 21, pp. 1–10.

Lier, M (2010). "Hohe Provisionen stoßen BaFin übel auf". In: *Versicherungsvertrieb* 6.2.

Lo, Andrew W. (2005). "Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis". In: *Journal of Investment Consulting* 7.2, pp. 21–44.

Loughran, Tim and Bill McDonald (2011). "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks". In: *The Journal of Finance* 66.1, pp. 35–65.

Lowe, Will et al. (2011). "Scaling policy preferences from coded political texts". In: *Legislative Studies Quarterly* 36.1, pp. 123–155.

Lucey, Brian M. and Brenda Carron (2011). "The effect of gender on stock price reaction to the appointment of directors: The case of the FTSE 100". In: *Applied Economics Letters* 18.13, pp. 1225–1229.

Malkiel, Burton G. (2005). "Reflections on the efficient market hypothesis: 30 years later". In: *Financial Review* 40.1, pp. 1–9.

Martin, Anna D., Takeshi Nishikawa, and Melissa A. Williams (2009). "CEO gender: Effects on valuation and risk". In: *Quarterly Journal of Finance and Accounting*, pp. 23–40.

Mayraz, Guy, Gert G. Wagner, and Jürgen Schupp (2009). "Life satisfaction and relative income-perceptions and evidence". In: *IZA Discussion Papers*. URL: `http://ftp.iza.org/dp4390.pdf`.

Milbrodt, Hartmut and Volker Röhrs (2016). *Aktuarielle Methoden der deutschen Privaten Krankenversicherung*. Vol. 34. VVW GmbH.

Moeller, Sara B., Frederik P. Schlingemann, and René M. Stulz (2004). "Firm size and the gains from acquisitions". In: *Journal of Financial Economics* 73.2, pp. 201–228.

Mohan, Nancy (2014). "A review of the gender effect on pay, corporate performance and entry into top management". In: *International Review of Economics & Finance* 34, pp. 41–51.

Nelson, Julie A. (2014). "The power of stereotyping and confirmation bias to overwhelm accurate assessment: The case of economics, gender, and risk aversion". In: *Journal of Economic Methodology* 21.3, pp. 211–231.

Oakley, Judith G. (2000). "Gender-based barriers to senior management positions: Understanding the scarcity of female CEOs". In: *Journal of Business Ethics* 27, pp. 321–334.

Park, Dong-Jin and Bruce K. Berger (2004). "The presentation of CEOs in the press, 1990-2000: Increasing salience, positive valence, and a focus on competency and personal dimensions of image". In: *Journal of Public Relations Research* 16.1, pp. 93–125.

Parrino, Robert (1997). "CEO turnover and outside succession a cross-sectional analysis". In: *Journal of Financial Economics* 46.2, pp. 165–197.

Patelli, Lorenzo and Matteo Pedrini (2014). "Is the optimism in CEO's letters to shareholders sincere? Impression management versus communicative action during the economic crisis". In: *Journal of Business Ethics* 124, pp. 19–34.

Paul, Maureen (2006). "A cross-section analysis of the fairness-of-pay perception of UK employees". In: *The Journal of Socio-Economics* 35.2, pp. 243–267.

Pauly, Mark V., Howard Kunreuther, and Richard Hirth (1995). "Guaranteed renewability in insurance". In: *Journal of Risk and Uncertainty* 10, pp. 143–156.

Peltomäki, Jarkko et al. (2021). "Age, gender, and risk-taking: Evidence from the S&P 1500 executives and market-based measures of firm risk". In: *Journal of Business Finance & Accounting* 48.9-10, pp. 1988–2014.

Peni, Emilia and Sami Vähämaa (2010). "Female executives and earnings management". In: *Managerial Finance* 36.7, pp. 629–645.

Pfeifer, Christian and Gesine Stephan (2019). "Why women do not ask: gender differences in fairness perceptions of own wages and subsequent wage growth". In: *Cambridge Journal of Economics* 43.2, pp. 295–310.

Posey, Lisa L. and Sharon Tennyson (1998). "The coexistence of distribution systems under price search: Theory and some evidence from insurance". In: *Journal of Economic Behavior & Organization* 35.1, pp. 95–115.

Posey, Lisa Lipowski and Abdullah Yavaş (1995). "A search model of marketing systems in property-liability insurance". In: *Journal of Risk and Insurance*, pp. 666–689.

Powell, Gary N. and D. Anthony Butterfield (1994). "Investigating the "glass ceiling" phenomenon: An empirical study of actual promotions to top management". In: *Academy of Management Journal* 37.1, pp. 68–86.

Price, S. McKay et al. (2012). "Earnings conference calls and stock returns: The incremental informativeness of textual tone". In: *Journal of Banking & Finance* 36.4, pp. 992–1011.

Private Health Insurers, German Association of (2020). *PKV-Zahlenportal*. Online. URL: `https://www.pkv-zahlenportal.de`.

Regan, Laureen (1997). "Vertical integration in the property-liability insurance industry: A transaction cost approach". In: *Journal of Risk and Insurance*, pp. 41–62.

Regan, Laureen and Sharon Tennyson (1996). "Agent discretion and the choice of insurance marketing system". In: *The Journal of Law and Economics* 39.2, pp. 637–666.

Rendall, Michelle (2017). "Brain versus brawn: the realization of women's comparative advantage". In: *University of Zurich, Institute for Empirical Research in Economics, Working Paper* 491.

Research Data, Norwegian Centre for (2018). *European Social Survey Round 9*. electronic. DOI: `10.21338/NSD-ESS9-2018`.

Rose, Caspar (2019). "Stock market reactions to CEO succession announcements: inside versus outside recruitment?" In: *Journal of Management and Governance* 23.1, pp. 33–65.

Ryan, Michelle K. and S. Alexander Haslam (2005). "The glass cliff: Evidence that women are over-represented in precarious leadership positions". In: *British Journal of Management* 16.2, pp. 81–90.

Schmitt (Dec. 2010). "Die Versicherer gegen die Verkäufer, Handelsblatt". In: *Handelsblatt*, p. 32.

Schneck, Stefan (2014). "My wage is unfair! Just a feeling or comparison with peers?" In: *Review of Behavioral Economics* 1.3, pp. 245–273.

Sent, Esther-Mirjam and Irene van Staveren (2019). "A feminist review of behavioral economic research on gender differences". In: *Feminist Economics* 25.2, pp. 1–35.

Seog, S Hun (1999). "The coexistence of distribution systems when consumers are not informed". In: *The Geneva Papers on Risk and Insurance Theory* 24, pp. 173–192.

Serfling, Matthew A. (2014). "CEO age and the riskiness of corporate policies". In: *Journal of Corporate Finance* 25, pp. 251–273.

Shores, David (1990). "The association between interim information and security returns surrounding earnings announcements". In: *Journal of Accounting Research* 28.1, pp. 164–181.

Shrader, Charles B., Virginia B. Blackburn, and Paul Iles (1997). "Women in management and firm financial performance: An exploratory study". In: *Journal of Managerial Issues*, pp. 355–372.

Smith, Nina, Valdemar Smith, and Mette Verner (2006). "Do women in top management affect firm performance? A panel study of 2,500 Danish firms". In: *International Journal of Productivity and Performance Management* 55.7, pp. 569–593.

Sousa-Poza, Alfonso and Andrés A. Sousa-Poza (2000). "Taking another look at the gender/job-satisfaction paradox". In: *Kyklos* 53.2, pp. 135–152.

South, Scott J. and Glenna Spitze (1994). "Housework in marital and nonmarital households". In: *American Sociological Review*, pp. 327–347.

Spence, Michael (1973). "Job Market Signaling". In: *The Quarterly Journal of Economics* 87.3, pp. 355–374.

Staveren, Irene van (2013). "An exploratory cross-country analysis of gendered institutions". In: *Journal of International Development* 25.1, pp. 108–121.

Surminski, Marc (2019). "Die PKV im Jahr 2018". In: *Zeitschrift für Versicherungswesen* 19, p. 568.

Swim, Janet et al. (1989). "Joan McKay versus John McKay: Do gender stereotypes bias evaluations?" In: *Psychological Bulletin* 105.3, p. 409.

Tica, Amra and Barbara E. Weißenberger (2022). "How regulatory changes are driven by a need for control in reputational scandals: a case study in the German insurance industry". In: *Journal of Accounting & Organizational Change* 18.1, pp. 57–76.

Tosi, Henry L. and Steven W. Einbender (1985). "The effects of the type and amount of information in sex discrimination research: A meta-analysis". In: *Academy of Management Journal* 28.3, pp. 712–723.

Trigo-Gamarra, Lucinda and Christian Growitsch (2010). "Comparing single-and multichannel distribution strategies in the German life insurance market: an analysis of cost and profit efficiency". In: *Schmalenbach Business Review* 62, pp. 401–417.

Tversky, Amos and Daniel Kahneman (1986). "The framing of decisions and the evaluation of prospects". In: *Studies in Logic and the Foundations of Mathematics*. Vol. 114. Elsevier, pp. 503–520.

Tyrowicz, Joanna, Lucas van der Velde, and Irene van Staveren (2018). "Does age exacerbate the gender-wage gap? New method and evidence from Germany, 1984–2014". In: *Feminist Economics* 24.4, pp. 108–130.

Uhlmann, Eric Luis and Geoffrey L Cohen (2005). "Constructed criteria: Redefining merit to justify discrimination". In: *Psychological Science* 16.6, pp. 474–480.

VersicherungsWiki, e.K. (2022). Accessed October 2022. URL: https://www.pkv.wiki/cms/pkv/info/anbieter/uebersicht-aller-privaten-krankenversicherer.

Watson, Towers (2012). *Vertriebswege-Survey zur Krankenversicherung: Unabhängige Vermittler sichern sich weitere Marktanteile im PKV-Vertrieb*. Online. URL: https://www.assekuranz-info-portal.de/presse/versicherungen/2012/12/vertriebswege-survey-zur-krankenversicherung-unabhaengige-vermittler-sichern-sich-weitere-marktante/.

Weiss, Mary A., Sharon Tennyson, and Laureen Regan (2010). "The effects of regulated premium subsidies on insurance costs: An empirical analysis of automobile insurance". In: *Journal of Risk and Insurance* 77.3, pp. 597–624.

Wolfers, Justin (2006). "Diagnosing discrimination: Stock returns and CEO gender". In: *Journal of the European Economic Association* 4.2-3, pp. 531–541.

Wood, Wendy and Alice H. Eagly (2012). "Biosocial construction of sex differences and similarities in behavior". In: *Advances in experimental social psychology*. Vol. 46. Elsevier, pp. 55–123.

Yekini, Liafisu Sina, Tomasz Piotr Wisniewski, and Yuval Millo (2016). "Market reaction to the positiveness of annual report narratives". In: *The British Accounting Review* 48.4, pp. 415–430.

Yim, Soojin (2013). "The acquisitiveness of youth: CEO age and acquisition behavior". In: *Journal of Financial Economics* 108.1, pp. 250–273.

You, Ya et al. (2020). "How CEO/CMO characteristics affect innovation and stock returns: findings and future directions". In: *Journal of the Academy of Marketing Science* 48, pp. 1229–1253.

Yuthas, Kristi, Rodney Rogers, and Jesse F. Dillard (2002). "Communicative action and corporate annual reports". In: *Journal of Business Ethics* 41, pp. 141–157.

Zajac, Edward J. (1990). "CEO selection, succession, compensation and firm performance: A theoretical integration and empirical analysis". In: *Strategic Management Journal* 11.3, pp. 217–230.

Zeitung, Ärzte (Nov. 2010). "Gerangel um Kunden: PKV sucht Hilfe bei der BaFin". In: *Ärzte Zeitung*.