Essays on Asset Pricing with Big and Unconventional Data

by

Danilo Antonino Giannone

Thesis

Submitted to The University of Warwick

for the degree of

Doctor of Philosophy

in Finance & Econometrics

Warwick Business School

October 2021

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgements

Faber est suae quisque fortunae

Every man is the artisan of his own fortune

Appius Claudius Ceacus - 4^{th} century BC

I would like to express my deepest appreciation to my supervisor and co-author Professor Cesare Robotti for our discussions about research and his technical advice. He has been of invaluable help in every aspect of my journey as a PhD student. I am also deeply indebted to Dr Antoniou Constantinos for his suggestions and guidance throughout my PhD. He always found the right approach and words to encourage me, which helped me to grow personally and academically.

I would like to extend my sincere thanks to Professor Roman Kozhan for his feedback during my annual reviews, and also Professor Philippe Müller for his support in every organisational aspect of the PhD.

I would like to take this opportunity to thank my friends and colleagues at Warwick University: Jonathan Canioni, Andrea De Polis, Francesco-Saverio Gaudio, Elettra Latini and Karla Monterrubio, who provided me with both, emotional support as well as comments and feedback on my work. I am also grateful to my friends Biagio Esposito, Laura Massera, Nicholas Meli, Andrea Ostini and Giada Pizzoni for encouraging me throughout my academic career.

Finally, I express my gratitude to my family, who are always by my side and encourage me in all my endeavours. Therefore, I would like dedicate this thesis to my mother, who has been my rock over the years. She always showed me the bright side of a moment, urged me to pursue my objectives and never give up.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. The first and second chapter have been composed by myself and has not been submitted in any previous application for any degree. I declare that the third chapter is jointly authored Professor Cesare Robotti.

Abstract

This thesis comprises three papers on empirical asset pricing, where big and unconventional datasets are implemented through Artificial Intelligence, and recent advancements in econometrics. The first paper analyses intraday returns to investigate the time-varying characteristic of the systematic risk around unscheduled firm-level news. More precisely, I study firm-level news writing about secondary equity offering programs. I show that on the day that the press reports the news, the beta of the company drops by a statistically significant and economically important amount. I demonstrate that, through this variation, it is possible to explain more than a half of the abnormal return documented on the event day. The second paper leverages recent advancements in text analysis to derive a novel measure from the distribution of words in firm-level news. This proxy variable overcomes most of the limitations of classical text analysis and does not require to be trained on any quantitative variables. I show that this proxy captures the changes in the stock return volatility and in the systematic risk, and it explains the variations in market risk. Hence, I argue that this measure is a useful instrument to understand the time-varying uncertainty of the company, and by extension, of the market. Finally, the third paper investigates whether the New York Federal Reserve primary dealers are special marginal investors with very different characteristics relative to the non-primary broker-dealer sector. Using publicly available information, we identify the publicly traded ultimate parent company of each primary dealer and show that they are not special. Through recent advancement in econometrics, we also conclude that the factors derived from the primary dealer observations does not possess any pricing performance.

Introduction

Many research fields, such as engineering and medicine, are embracing unconventional and/or Big Data to create new knowledge. The concept of unconventional data refers to data under development or, more generally, data that need to be elaborated in order to be used (European, 2017). Conversely, Big Data refers to large amounts of data that requires advanced computational methodologies to infer information. The computational difficulties along with the unbounded insights that researchers can gain from these types of datasets motivate scholars and practitioners to develop new techniques in the area of Artificial Intelligence (AI) and econometrics. In the context of the finance academic literature, the use of AI and Big Data is still very much nascent. This thesis documents the importance of unconventional and Big Data, the adoption of AI and the recent advancement in econometrics in empirical asset pricing.

Due to the digitisation of current and past archival data, along our ability to accurately record market information, the financial literature expanded in areas that were hitherto unexplained. For example, these data allows us to study the relationship between quantitative and qualitative in information in low and high frequency setting. In addition, we currently are able to examine changes in intra-day prices for an extensive periods of time, thus permitting the analysis of short lived changes around firm-level news in stock prices that would have been impossible. Moreover, we can translate qualitative information in a readable computer format, to infer the meaning and the topics discussed in written documents. Finally, it is possible to reconstruct the relationship between companies (i.e., controlled and controlling companies), starting from more than half a century ago. This thesis leverages on these new capabilities, through three chapters.

The first two chapters of this thesis provide a novel lense to address the implications of firm-level news on the systematic risk. Despite considerable efforts in the financial literature to determine the relationship between these two variables, especially around earnings announcements (i.e., Ball and Kothari (1991), Patton and Verardo (2012) and Savor and Wilson (2016)), most studies provide incremental contribution to theory. For example, some reconceptualise established theories and/or support previous empirical findings by using different theoretical perspectives. Few studies attempt to challenge or develop new theories to conceptualise the links between firm-level news and systematic risk. The extent literature argue that firm-level news communicates information about the market, which increases the systematic risk through a spillover effect. Stated differently, when investors learn about the market from firm-level news, the systematic risk of the company increases.

The fact that firm-level news provide information about the market is an interesting idea. Accordingly, around the issuance of firm-level news that provides information about the market, we should observe an increase in the systematic risk of the company. My first chapter rejects this hypothesis. I study the systematic risk variation through four years of intra-day prices for all publicly listed companies in the US around firm-level news writing about secondary equity offering (SEO) programs. I show that on the day that news is reported, the systematic risk drops. However, when the systematic risk is calculated with daily data, this variation disappears. To investigate the economic relevance of this finding, I measure the abnormal return (AR) around SEO news. I demonstrate that the AR halves compared to a standard approach. Given these results, the paper concludes that when investors learn about the market, the systematic risk of the company does not necessarily increase. Building on this, in the second chapter I propose a proxy variable derived from the linguistic description of firm-level news. This proxy quantifies the relationship between the firm-level news and the market to explain the time-varying systematic risk of the company.

I argue that firm-level news is composed of a company-specific and market-wide component, and the latter component links the company to the market. This composition of firm-level news can increase the systematic risk in some instances, while in others can lead to a decrease in the systematic risk. Hence, the more the firm-level news is related to the market, the more the company is exposed to the market variation. By dissecting the two components, it is possible to classify companies according to their market-wide component, and examine whether it effects the variation of the systematic risk.

Using AI's Natural Language Processing (NLP) methods, in the second chapter, I provide a new methodology to quantify the above mentioned market-wide component in firm-level news. This methodology does not require the researcher to specify any dictionaries or to be trained on quantitative variables. This makes this methodologies applicable to an array of written documents. Through seventeen years of daily observations and approximately a half-million firm-level news items, I demonstrate that the proxy variable delivered by this methodology can reconcile the findings of the ongoing literature with those of the first chapter. I call this proxy, the Market Similarity (MS). Hence, companies with low (high) MS have a systematic risk that is lower (higher) than the average companies in the sample. This chapter also document that the MS explains the stock return volatility of the company. Furthermore, when the MS is aggregated across all companies in the sample, it explains and predicts the market volatility. I conclude that the MS measures the firm uncertainty when deployed for a single company, while when it is aggregated across all companies, it measures the market uncertainty.

The last chapter investigates a crucial aspect of empirical works: the replicability and reliability of pricing factors. The construction of trustworthy knowledge relies upon the strict adherence of researchers to these elements. In this paper we address this by focusing on the New York Federal Reserve (NY Fed) primary dealers factor of He et al. (2017). They propose two pricing factors: one derived from the stock returns of NY Fed PD, the other from their balance-sheet information. They show that both factors price a wide range of asset classes. In a following publication, Gospodinov and Robotti (2021) show that by applying the correct econometric framework to their factor, the pricing performance of these factors vanishes. It is certainly prestigious to become a NY Fed's primary dealer, and the fact that they are marginal investors in the major asset markets might suggest that they are indeed unique. However, after re-building the factors, we document a lack of pricing performance of the two factors. In addition, we employ an event-study analysis to show that becoming primary dealers of the NY Fed does not generate any relevant price effect for those companies. Finally, in order to develop the literature, we make an exhaustive list of the publicly-listed controlling companies.

A more inclusive introduction of the three chapters in a stand alone format, is reported below. Chapter 1 and 2 are my sole work, while Chapter 3 is co-authored with Professor Cesare Robotti.

Chapter 1: The Effect of Unscheduled News on Systematic Risk

Current literature shows that, around earnings announcements, the systematic risk of the announcing company increases, both at daily and intra-daily frequency. These studies argue that this increase is due to investors learning about the profitability of non-announcing firms from the earnings announcement of a company. Although there are many firm-level news that satisfy this condition, our current understanding is mostly limited to this announcement. In this paper, I show that unscheduled announcements - firm-level news writing about SEO - lead to a decrease in systematic risk of the company in the day that the news is reported.

There are a few worthwhile differences between earnings and SEO news. For instance, investors interpret earnings announcements as positive or negative based on the amount of earning that is expected compared to what will actually be released by the company. Conversely, investors usually interpret SEOs as negative signals for the company. Earnings are scheduled announcements and represent an information about the profitability of the company, whereas SEOs are unscheduled announcements and reflect the managers' belief about the valuation of the company. More generally, studying SEOs news contributes to our understanding concerning the link between firm-level news and systematic risk.

In more details, Patton and Verardo (2012) show that good (bad) news about the earnings announcements of a company, is interpreted as good (bad) news about non-announcing, which leads to an increase in systematic risk on the event date. Contrary to their work, I show that, in the presence of SEO news, the systematic risk of the company decreases, by a much larger amount than around earnings announcements. I argue that investors understand the implications of an SEO program in the company valuation and move their wealth to more fairly valued companies. Since part of the valuation of a company can be attributed to a market-wide component, they sell the announcing company in favour of a company outside the given sector. In support of this idea, I show that, when the systematic risk is calculated according to the respective sector mimicking portfolio of the company rather than the US market, the systematic risk variation is smaller. The finding of this paper are supported by the corporate finance literature, which asserts that SEO programs lead to a decrease in systematic risk.

This paper also provides two additional results. Firstly, the variation in systematic risk reported in this paper explains more than half of the negative AR around SEO news. This result is of interest to scholars in the finance and accounting fields who study AR around specific events. Thus, I suggest to scholars they check whether around the given event the systematic risk of the company varies before reporting the size of the abnormal return detected in their research. Secondly, after controlling for the change in news sentiment in and around the news date, I show that sentiment inferred from firm-level news does not explain the change in systematic risk. This insight is especially of interest to practitioners who use unstructured datasets to monitor their risk exposure. Indeed, they should closely monitor the taxonomy of the news, rather than its sentiment, to understand the risk they are bearing on a daily basis.

Chapter 2: Dissecting Firm-Level News: a New Measure to Capture the Time-Varying Risk of the Company

In order to understand whether more information equates to less uncertainty, scholars analyse stock prices variation in the presence or absence of firm-level news. In this regard, the current literature proposes two distinct views. The first asserts that news increases the uncertainty of the company, while the second argues that news decreases the uncertainty of the company. I argue that an important element to understand the link between news and uncertainty, is given by the narrative of the news. Towards this end, I propose a new measure, which I call MS. I show that the MS explains the time-varying systematic risk of the company, the stock return volatility and the overall market risk.

The MS is based on the idea that each firm-level news comprises two components: a company-specific and a market-wide one. Given this idea, the paper investigates three predictions related to the market-wide component. First, when a considerable number of firms have news where the market-wide component dominates the company-specific one, the market risk increases. Second, the larger the market-wide component is, the less information investors obtain about the company, and the higher the stock return volatility. Third, the larger the market-wide component, the more the company is related to the market, and the higher the systematic risk.

In order to derive the MS, I rely on a multitude of advanced computational linguistics techniques. I propose several tests to document that, through this methodology, it is possible to measure both, the change in the market and company information environment. There are three notable differences between MS and other measures derived from news. The MS does not require: (i) the researcher to specify any specific word to be selected or to filter for particular news; (ii) extensive or discretionary coding rules; and, (iii) to be trained on any quantitative variable.

Boudoukh et al. (2019) show that by selecting fourteen firm-level news categories, it is possible to better explain price movements. They assert that these categories are the only one that are relevant for explaining the return volatility of the company. I extend their work by demonstrating that, through the words distribution in firm-level news, it is possible to reach a similar conclusion, but without the need of selecting any types of news items.

This paper documents several new insights about the implications of firm-level news. To the best of my knowledge, this is the first research linking the linguistic description of firm-level news to company uncertainty, through a proxy variable. Sorting companies on their MS confirms that this measure captures the company's uncertainty, as measured by the existing literature. That is, a higher MS leads to: high systematic risk, high stock return volatility and low R-squared. Furthermore, I show that the explanatory power and forecasting ability of the MS for the Market Volatility Index (VIX), is superior to a similar proxy derived from the verbal description of news. Finally, investors that hold the entire market can limit their expected loss by simply shorting companies with high MS. Hence, the MS is a real-time measure that quantifies the risk of the company, and by extension, of the overall market, which allows investor to better mange the risk exposure.

Chapter 3: Are the Primary Dealers of the New York Fed Really Special?

He et al. (2017) offer strong empirical support to the NY Fed's PDs as a priced factor across major asset classes. This leads them to conclude PD are special marginal investors who differ from the overall broker-dealer sector. Using information that is publicly available on the NY Fed's website, we reconstruct their factor and show that their results are likely driven by heavy and ad-hoc trimming of the list of PD. Based on our new sample, there is simply no evidence that the NY Fed's primary dealers are special marginal investors across any of the seven asset classes considered by the authors.

This paper investigates three main facts about the NY Fed's PD: (i) the intermediary factor of He et al. (2017) is priced in the cross-section of asset returns; (ii) there is an announcement effect in terms of risk and return from becoming active NY Fed's PDs; and (iii) the performance of the NY Fed's PDs deviates substantially from that of the non-primary broker-dealer sector. We show that none of these facts are confirmed in the data. The factor is not priced in the cross-section, there is no announcement effect. Furthermore, the performance of the NY Fed's PDs does not deviate from non-primary broker-dealers

The paper concludes that theNY Fed's PDs are certainly large finance intermediaries, and it is a prestigious position to cover, but they do not seem to differ substantially in performance from the non-primary sector, based on their risk/return profile and several other characteristics.

Chapter 1

The Effect of Unscheduled News on Systematic Risk

1.1 Introduction

From an asset-pricing point of view, firm-level news is commonly believed to have no or minimal implications to systematic risk variation (Engelberg et al., 2018a), merely because it is idiosyncratic by definition. However, recent evidence provided by Patton and Verardo (2012) and Savor and Wilson (2016) demonstrates that around earnings announcements the beta increases. By relying on this idea, I investigate the variation in the systematic risk in the presence of firm-level news writing about Secondary Equity Offering (SEO).[1] I show that around this information flow the systematic risk[2] of the company decreases by 33.4% and this variation explains more than 50% of the negative abnormal return on the SEO announcement date.

In this paper, I compare the prediction of the learning model of Patton and Verardo (2012) and that of corporate finance one (Eckbo et al., 2000; Carlson et al., 2010) to understand the behaviour of systematic risk around SEO announcements. Patton and Verardo (2012) hypothesise that earnings announcements are composed of a company-specific (i.e., idiosyncratic) and market-wide (i.e., common) component. When investors learn about the market from the latter component, they trade both the announcing firm and the market (i.e., non-announcing firms). Hence, investors interpret good (bad) news about the company as good (bad) news about the market. This information spillover effect (i.e., learning across stocks), emerging from the market-wide component of earnings announcements, spikes the systematic risk in the announcement date. Instead, Carlson

[1] I do not make any distinction between seasonal equity offering, secondary equity offering and rights issues. Furthermore, I use, news, firm-level news and announcement interchangeably.

[2] In this paper, I use the word beta, realised beta and systematic risk interchangeably.

et al. (2010) from a real options theory perspective, have formalised a model where the commitment to invest of the company plays a critical role in describing the behaviour of beta after stock issuance. Since real options theories build on rational expectations, there is an immediate price reaction. If the company has no commitment to invest, and time-to-build (e.g., the rate at which the investment evolves) then the systematic risk drops to the same level as the company's assets-beta, on the announcement date. Meanwhile, Eckbo et al. (2000) argue that SEO decreases the beta of the company since the issue of new shares deleverages the equity beta of the company.

If the learning across stocks prevails over the corporate finance prediction, then the beta should increase on the SEO announcement date. Instead, if the corporate finance prediction dominates, then the systematic risk will decrease on the announcement date. This consideration is formalised in the first hypothesis of the paper: the information learned by investors from SEO announcements is used to evaluate non-announcing firms, which increases the systematic risk of the announcing company. Since SEO programs are often perceived as bad news for the company, the systematic risk will only increase if investors also sell non-announcing firms. The alternative hypothesis relies on the prediction of the corporate finance literature, where SEO programs lead to a decrease in the company's systematic risk (Eckbo et al., 2000; Carlson et al., 2010). If, on the announcement date, investors immediately incorporate their expectation into the issuing company, and rebalance their portfolios by selling the announcing stock in favour of other stocks, then the beta will drop. That is, when investors rebalance their stocks' holding, the comovement of returns between the announcing stock and the market drops, leading to a decrease in the company's systematic risk. Based on the results reported in this paper, I reject the null in favour of the alternative hypothesis.

To study the effect of SEO announcements on systematic risk, I use similar methodology to that of Patton and Verardo (2012). I employ a panel regression to study the daily variation in realised beta. I calculate the daily realised beta for the twenty-one days surrounding the event date using the TAQ dataset from 9:45 a.m. to 4:00 p.m., for all stocks listed on the NYSE, AMEX and NASDAQ from January 1, 2010 to December 31, 2013.

There are two reasons why SEO announcements are a suitable choices to study systematic risk variation. First, this program is often linked to managers' willingness to signal the current valuation of the company. Corporate and behavioural finance theories suggest that the valuation of companies can be partially attributed to a market-wide component (Otchere and Ross, 2002). Because these events are rarely implemented by companies, it is a unique opportunity for investors to obtain information about the current and future valuation of the announcing as well as non-announcing companies. Second, corporate finance theory suggests that the issuance of new shares leads to a decrease in the systematic risk of the company. Hence, this announcement satisfies both the requirement of the learning model and the corporate finance one contemporaneously.

While companies increasingly roll out SEO programmes, our current understanding of investors' risk exposure around this announcement is limited. As such, the empirical motivation of this paper is to examine the effect of SEO announcements on the risk exposure of investors. For example, given a low level of interest rate, companies might undertake an SEO program to payback existing expensive debt and refinance it at a lower rate. Companies might also sell shares to take advantage of a current overvaluation (i.e., bull market) or, as during the recent health crisis, to finance existing operations or raise cash to stay afloat.[3]

From an academic point of view, this paper is motivated by the large adoption of event studies in the finance literature and to the increasing interest in the link between news and company risk and return. An earnings announcement might not be an isolated case, but other announcements might shift the systematic risk of the company on the announcement date. If this is the case, then some portion of the abnormal return (AR) commonly reported around information flow might be due to the change in the beta, which questions the real size of the AR. Furthermore, given the contrasting predictions between the learning and real options theory, this paper provides a formal test of the reason why systematic risk varies around SEO announcements.

I uncover statistically significant and economically important variations in the realised beta on the announcement date. The systematic risk spikes nine days before the announcement by 0.201 (t-stat of 2.88) and drops on the announcement day by -0.334 (t-stat of -2.61), reverting back to its pre-announcing level thereafter. Therefore, the effect documented around an earnings announcement is not an isolated case; rather, other announcements lead to a variation in the beta of the announcing company.[4] This result

[3]This trend is featured by Bloomberg in a recent article.

[4]In Appendix A.4, I show that the beta varies also around buyback announcements.

rejects the main hypothesis of the paper, in favour of the alternative one.

To understand the importance of the results presented in this paper, I compare the AR with and without accounting for the high-frequency variation in beta. I show that, by using a classical methodology, the AR in the announcement date is equal to -2.59% and statistically significant at the 10% level. Instead, when the beta is allowed to vary the AR decreases to -1.12% and is statistically significant at the 10% level.

Furthermore, I investigate whether the sentiment of the news about the announcing company over the twenty-one days around the announcement date explains the change in systematic risk. Toward this end, I calculate the daily average sentiment of all news items about the company, and use it as a controlling variable. This analysis reveals that news sentiment does not explain the documented change in systematic risk. In other words, the taxonomy related to the news has more explanatory power than sentiment regarding the change in systematic risk.

I then check the robustness of the results from two different points of view. The first is from the possibility of omitted variables and, the second, from the information environment (i.e., placebo tests). Denis and Kadlec (1994) assert that trading volume and bid-ask spread drives the change in beta around SEO announcements. Andersen et al. (2006) show that betas are autocorrelated. Forbes and Rigobon (2002) document that the realised volatility might affect the comovement of stocks. After controlling for all these variables, the results remain statistically significant, except for the variation nine days before the announcement date that disappears when controlling for the first lag of the beta. I then conduct two placebo tests. Both tests confirm the results presented in this paper: investors' trading proclivity drives the results.

This article directly contributes to the work proposed by Patton and Verardo (2012), who study the variation of the systematic risk of the announcing company around earnings announcements.[5] They show that the beta of the company always increases on the announcement day. Conversely, the results documented in this paper show the exact opposite behaviour of the beta: the systematic risk of the announcing company decreases. This is a considerable difference, since the learning model proposed for earnings announcements does not hold for SEO announcements. More generally, I contribute to our understanding

[5] In Appendix A.1, I briefly show the main results of Patton and Verardo (2012), to emphasise the main difference between the change in beta around earnings announcements and that documented in this study.

of the importance of firm-level news for the time-varying beta. To the best of my knowledge, this is the first paper to illustrate the difference between scheduled and unscheduled announcements in the high-frequency change of company systematic risk.

This research also contributes to a broad range of event studies investigating the change in systematic risk around SEOs using daily returns. Denis and Kadlec (1994) test the prediction of the microstructure literature, which asserts that the variation in beta around SEOs is due to the change in trading volume and bid-ask spread. After controlling for these two variables, they find the the variation in systematic risk is indistinguishable from zero. Carlson et al. (2010), using real options theory with daily observations, demonstrate that systematic risk increases before and decreases after the announcement of SEOs. I show that the beta of the announcing company decreases on the announcement date and reverts to its pre-announcing level.

This article is also related to the literature studying the effect of information spillover across companies. Bradley and Yuan (2013) show that SEOs elicit a contagion effect. Hameed et al. (2015) argue that "bellwether" companies are used to price neglected stocks, especially those with high analyst coverage and companies with correlated fundamentals. Similarly, Piotroski and Roulstone (2004) investigate how the information revealed by analysts, institutional and insider investors is incorporated into stock prices. They find that firm-specific information is better reflected by the presence of insider and institutional investors, while analyst forecasting accelerates the inclusion of sectors and firm-specific information. I take a different perspective and use the information spillover effect to measure the change in systematic risk. I show that SEOs communicate relevant information about non-announcing firms operating in the same market or sector, and this is more relevant for the market than for the industry. This effect can be attributed to better pricing efficiency of companies operating in the same sector.

The remainder of the paper is organised as follows. In Section 2.2, I discuss the current literature about the systematic risk variation and the major characteristics of SEOs. In Section 1.3, I develop the two hypotheses. Section 2.3 describes the realised beta calculation and the cross-sectional model. Section 2.4 presents the derivation of the sample and the descriptive statistics. In Section 2.6, the empirical results are illustrated, while in Section 1.7, I present the robustness tests, and Section 2.7 concludes the paper.

1.2 Current Literature

According to Patton and Verardo (2012), earnings announcements are a suitable choice to study the variation of systematic risk, since they are reported regularly by companies and are well-documented information disclosures. If earnings information is cross-correlated, the returns of companies announcing together increase, which is reflected in an increase of the beta of the announcing firms (Ball and Kothari, 1991). Savor and Wilson (2016) report similar results. They assert that earnings announcements are noisy signals about market cash flow. Thus, any news about individual company profitability becomes a source of systematic risk, since it reveals salient information about the aggregate cash flow. Patton and Verardo (2012), studying the daily variation of the realised beta calculated with intraday returns around earnings announcements, show that, on the event date, the systematic risk of the announcing company spikes. They explain this increase as a consequence of investors learning about the profitability of non-announcing firms from the announcement of another company (i.e., learning across stocks). They theoretically demonstrate that when investors learn about non-announcing firms from the common component of earnings announcements, the systematic risk increases.

Another strand of the literature argues that SEOs convey information about the current valuation and future profitability of the company.[6] Managers issue shares when they need additional capital to finance future investments or when the market is temporarily optimistic about the firm's future performance (Ritter, 2003). Since managers are able to time the market, they issue shares when the company is overvalued (Myers and Majluf, 1984). The market perceives the issuance of new shares as an adverse signal, which is reflected in a negative price reaction around the announcement date. Bradley and Yuan (2013) show that the contagion effect dominates SEO announcements since rivals are more likely to initiate an SEO program.

Masulis (1983) demonstrates that investors perceive the choice of deleveraging the company through an equity issue as an increase in the riskiness of future earnings. Eckbo et al. (2000) report that companies issuing shares underperform when compared to similar companies that do not issue shares. Their explanation is that equity issuers undergo a reduction in systematic risk as a consequence of deleveraging their equity, which decrease

[6]For a review of the reasons and the effect of SEO announcements, I refer the reader to Ritter (2003).

the stock's expected returns. Carlson et al. (2010), formalised a model where the commitment a of company to invest plays a critical role in describing the behaviour of beta after stock issuance. If the company has no investment commitment, and time-to-build (e.g., the rate at which the investment evolves) then the systematic risk following an SEO announcement drops to the same level as the company's assets-beta. Unlike the above literature, Johnson et al. (2018) show that overvaluation is not the main reason for managers to issue shares. They demonstrate that SEO programs are used to signal that the issuer's major customers will suffer financial difficulty in the future. Hence, the stock price of the issuer and customer decline.

Denis and Kadlec (1994), by using daily observations, estimate the daily change in systematic risk following an SEO announcement. They show that an increase in trading volume leads to an automatic increase in systematic risk. Hence, they document that, around this announcement, the systematic risk of the company increases, but this increase is driven only by trading activity. In support of their results, the authors also found no significant change in the systematic risk for companies without strong trading activity around the announcement date. They conclude that the change in systematic risk following equity offering is only due to the change in trading activity.

In sum, the systematic risk around earnings announcements increases due to the ability of investors to learn about non-announcing firms from the announcement of the company. SEOs seem to allow investors to learn across stocks too. The learning across stocks around this event is mainly due to the commonality in fundamentals that link the announcing company to non-announcing firms. This information spillover effect suggests that the systematic risk of the announcing company might vary around SEO announcements, too. Furthermore, the corporate finance literature predicts a reduction in systematic risk due to the choice in deleveraging the company and the change in investment commitment.

1.3 Hypotheses Development

The first part of this section introduces the two hypotheses of the paper, which are describe in detail in the following two subsections. The first, and main hypothesis, is used to investigate whether the change in systematic risk follows the prediction of the learning model of Patton and Verardo (2012) or that of corporate finance (Eckbo et al., 2000;

Carlson et al., 2010). The second hypothesis is used as a support for the first one. Hence, it investigates whether investors move their wealth away from the given sector to avoid future losses.

These two hypotheses are defined as market and sector hypotheses, which are the names of the two samples. The comparison between the market and sector sample is done by looking at the size and statistical significance of the coefficients of the panel regression, as presented in Section 2.3. That is, the higher and more statistically significant the coefficients, the more the investors use the information from the announcement to evaluate non-announcing firms.

Current literature shows that the stock prices of the announcing company and its peers vary around an SEO announcement (Bradley and Yuan, 2013; Johnson et al., 2018) and SEOs allow an information spillover effect (Vermaelen, 1981; Myers and Majluf, 1984). Patton and Verardo (2012) assert that investors interpret good (bad) news about the announcing company as good (bad) news for non-announcing firms. If this investor learning is also observed around SEO announcements, then the systematic risk of the company is expected to increase on the announcement date. The alternative prediction follows the corporate finance literature which asserts that, by issuing the new shares, the company deleverages its operation, thereby reducing its systematic risk. Hence, if the corporate finance prediction is observed in the data, the systematic risk is expected to drop on the announcement date.

More formally, Patton and Verardo (2012) propose a rational model of learning around earnings announcements to guide their empirical analysis. The model assumes that the daily return of a stock is driven by the change in expectations of earnings:

$$R_{i,t} = (\mathbb{E}_t[\log X_{i,t}] - \mathbb{E}_{t-1}[\log X_{i,t-1}]) + \epsilon_{i,t}, \tag{1.1}$$

where $X_{i,t}$ is the level of earnings of stock i on day t. Thus, the daily return is driven by the change in investor expectation of the earnings from day t-1 to t, plus other effects reflected in the residuals $\epsilon_{i,t}$. Since the earnings of firm i are only observable on the announcement day, on all other days investors have to infer this information from past earnings information of firm i as well as information of other firms' earnings. This cross-learning is possible because the innovations to earnings ($w_{i,t}$) are assumed to have a common component and an idiosyncratic component.

$$\begin{aligned} \Delta \log X_{i,t} &= g_i + w_{i,t}, \\ w_{i,t} &= \gamma_i Z_t + u_{i,t}, \end{aligned} \tag{1.2}$$

where g_i is the average growth rate in earnings for firm i, $w_{i,t}$ are innovations in the earnings process for firm i, Z_t is the common component of the earnings innovations, and $u_{i,t}$ is the idiosyncratic component of the innovations. Parameter γ_i captures the importance of the common component for stock i. This implies that the earnings announcement of a given firm contains information about both announcing and non-announcing firms.[7]

In the specific case, if SEO announcements communicate relevant information about the valuation of fundamentals of non-announcing firms, then investors will revise their expectation about these firms in the same way they revise those about the announcing firm and the systematic risk of the company is expected to increase.

From a corporate finance point of view, Eckbo et al. (2000) and Carlson et al. (2010) provide a solid alternative view on what might occur in the SEO announcement date. The former authors argue that when companies issue shares their equity leverage decreases, driving down their systematic risk along with their future performance. If investors understand the future implication of the SEO program, they will sell the announcing companies in favour of non-announcing stocks. This trading proclivity, initiated by investors to avoid future losses, is then reflected in a temporary drop in systematic risk. The latter authors, instead, provide the same dynamic in systematic risk, but from a real options theory point of view. They show that the systematic risk increases before and decreases after the announcement of an SEO. However, when there is no investment commitment of the company, the systematic risk drops on the SEO announcement date, and becomes equal to the beta of the assets-in-place. It is worth highlighting that, due to the setting used in this paper, it is not possible to measure either the investment commitment or time-to-build. Therefore, the real options theory developed by the authors is used to guide the prediction of this paper, rather than test their results.

[7]For further information about the model, I refer the reader to the Appendix 2, page 43, of Patton and Verardo (2012).

1.3.1 First Hypothesis - Market Hypothesis

The first hypothesis tests whether SEOs convey information about non-announcing firms, and whether this information is reflected in a variation of individual firm beta:

H_1: The information learned by investors from SEO announcements is used to evaluate non-announcing firms, which increases the systematic risk of the announcing company.

If the information obtained from announcing companies is learned and used quasi-instantaneously for the valuation of non-announcing firms, then the systematic risk of the announcing company will increase. I expect this variation to be positive and statistically significant on the announcing date. Instead, the alternative hypothesis is that when investors understand the undervaluation of the company they invest in more fairly priced stocks, which leads to a decrease of the systematic risk of the company.

Through the results of this hypothesis, it is also possible to test the difference between scheduled and unscheduled announcements. Earnings releases are classified as scheduled announcements and receive great coverage by analysts and the press. Investors can form expectations on the amount of earnings that will be released and take the most profitable trading strategy. Conversely, SEOs are classified as unscheduled since they are not communicated to investors in advance. Investors become aware of the willingness of the company to undertake this program only on the release date, which does not allow them to form prior expectations. I expect that, if any, the change in the systematic risk on the announcing date will be larger for SEOs than for earnings announcements. This prediction is consistent with the results of Patton and Verardo (2012), who show that the spike in systematic risk is larger (smaller) for high (low) earnings surprise relative to the forecast consensus.

1.3.2 Second Hypothesis - Sector Hypothesis

The second hypothesis investigates whether there are any differences in the information learned about non-announcing firms operating in the whole market, as opposed to non-announcing firms operating in the same sector. Companies operating in the same sector are likely to be exposed to the same demand shock, having related investment growth opportunities, being on similar life cycles or, more simply, copying the actions of their peers (Massa et al., 2007; Bradley and Yuan, 2013).

H_2: Investors learn less information from SEO announcements about non-announcing firms in the same sector compared to non-announcing firms operating in the same market.

If the information conveyed by the announcement is less relevant for non-announcing firms operating in the same sector than for firms operating in the same market, then the variation in the beta will be smaller and less statistically significant for the sector-sample. This implies that the sector is more efficiently evaluated, and that investors are aware of the contagion effect among companies operating in the same sector. The latter consideration supports the finding of Bradley and Yuan (2013), who find a contagion effect around SEO announcements. In other words, if investors know that companies operating in the same sector are likely to be overvalued and will implement an SEO program in the future too, they will allocate their capital to companies operating outside the given sector. This mechanism will reduce the change in systematic risk compared to the market-sample. Instead, if the information conveyed by the announcement is more relevant for non-announcing firms operating in the same sector than for firms operating in the same market, then the variation in the beta will be larger and more statistically significant for the sector-sample. This implies that the sector was inefficiently evaluated, and the new information significantly rectifies investor expectations.

1.4 Methodology

In this section, I present the methodology used to calculate the variation in systematic risk of the announcing company around the news date. This involves the calculation of the daily realised beta through a sixteen intraday-return sampled every twenty-five-minute, followed by the cross-sectional variation in the twenty-one days around the event date (i.e., event window).

1.4.1 Cross-Sectional Variation of Beta

To estimate the variation in systematic risk around the event date, I use the definition provided by Barndorff-Nielsen and Shephard (2004):

$$R\beta_{i,t}^{(S)} \equiv \frac{R\text{Cov}_{i,m,t}^{(S)}}{RV_{m,t,k}^{(S)}} = \frac{\sum_{k=1}^{S} r_{i,t,k} r_{m,t,k}}{\sum_{k=1}^{S} r_{m,t,k}^2} \tag{1.3}$$

where $r_{i,t,k}$ and $r_{m,t,k}$ are the log return of company i and market m. The return is calculated as $r_{i,t,k} = \log P_{i,t,k} - \log P_{i,t,k-1}$. That is, the return of stock i at time t computed with the intraday k^{th} prices P (the same applies to the return of the market $r_{m,t,k}$). The realised covariance $R\text{Cov}_{i,m,t}^{(S)}$ and the realised volatility of the benchmark index $RV_{m,t,k}^{(S)}$ are calculated S times within the day, for each time t. The benchmark index is either the market (hypothesis 1) or the given sectors mimicking portfolios (hypothesis 2). For a high level of frequency S - but not too high to avoid microstructure issues - the realised beta can be seen as a noisy, but unbiased, estimate of the integrated beta. To estimate the behaviour of beta around SEO news, I regress the realised beta on a series of dummy variables, using the following panel regression:

$$\begin{aligned} R\beta_{i,t}^{(S)} = \; & \delta_{-10} I_{i,t+10} + ... + \delta_0 I_{i,t} + ... + \delta_{10} I_{i,t-10} \\ & + \bar{\beta}_{i,1} D_{i,1} + \gamma' \mathbf{X}_{i,t} + \epsilon_{i,t} \end{aligned} \tag{1.4}$$

where $R\beta_{i,t}^{(S)}$ is the daily beta estimated with Equation 1.3. $I_{i,t}$ is a dummy variable for announcement day ($I_{i,t} = 1$ announced day and 0 otherwise) defined over the twenty-one days estimation window. $\bar{\beta}_{i,1}$ is the coefficient of the firm-year fixed effect, which captures the average beta outside the estimation window. δ_j measures the deviation of the beta from its long-run average in each day (δ_{-10} for the first day, and δ_{10} for the last day). I also include a series of controlling variables through $\mathbf{X}_{i,t}$ (i.e., volume, bid-ask spread) used for robustness tests in Section 1.7.1. When the announcement takes place while the market is closed, the event is set to the next trading day (circa 55% of the times) where the most trading activity takes place. I estimate the standard errors through the methodology proposed by Petersen (2009). Hence, I cluster the residuals along two dimensions, by firm-year and days, obtaining standard errors that are robust to heteroskedasticity and arbitrary within clusters.

If the announcements communicate any valuable information about non-announcing firms, the coefficients δ_j should be statistically different from zero:

$$\begin{aligned} & H_0 \; : \; \delta_j = 0 \\ vs. \quad & H_a \; : \; \delta_j \neq 0 \quad for\; j = -10, -9, \ldots, 10 \end{aligned}$$

According to the above hypothesis, the interpretation of Equation 1.4 is as follows. If firm-level news writing about an SEO program, communicates valuable information about the announcing and non-announcing firms, the coefficient δ_t should be statistically significant different from zero. A positive and statistically significant coefficient means that investors are trading non-announcing the same as announcing firms, thereby confirming the learning model of Petersen (2009). Instead, if the coefficient is negative and statistically significant, investors trade non-announcing in the opposite way to the announcing firm, thereby confirming the corporate finance prediction. Hence, investors interpret the announcement as good (bad) for the company, but bad (good) for non-announcing firms. The day in which the δ_j becomes statistically significant is the day in which investors can predict (before the announcing date) or incorporate (after the announcing date) the new information into non-announcing firms.

1.5 Sample Construction and Descriptive Statistics

The realised beta is calculated using the quote of the TAQ database, while SEO news is from the published news in the Dow Jones newswire. The market capitalisations and volume are from CRSP and Compustat, respectively. I then apply several filters to the news and companies to obtain the study sample.

I focus on publicly traded stocks on the NYSE, AMEX and NASDAQ from January 1, 2010 to December 31, 2013.[8] I sort the TAQ dataset in ascending order according to the Amihud illiquidity ratio (Amihud, 2002) in the first week of February of each year and delete the bottom decile. I delete companies that have more than 10% of missing

[8]The sampling period covers all the available data after the recession. I do not take into account the recession period since investors respond less to firm-level news. In Appendix A.2, I document the effect of the recession in investors learning across stocks through an example.

quote over the event window, and penny stocks (i.e., stock price less than five dollars). These criteria avoid the inclusion of illiquid stocks, which cause a downward bias of the covariance toward zero (Epps, 1979). I retain only companies that are contemporaneously identifiable in the CRSP and Compustat datasets. I take from CRSP the companies' SIC codes and exclude firms that are classified in the utilities (SIC codes from 4900 to 4942) or financial sector (SIC codes from 6000 to 6300 and from 6700 to 6799), similar to Foucault and Fresard (2014).

I use the identified companies to calculate the daily betas with a sampling frequency of twenty-five-minutes from 9:45 a.m. to 4:00 p.m., plus an overnight return, for a total of sixteen intraday returns. The sampling time starts fifteen minutes after the market opens to ensure that every stock is traded. According to previous studies, twenty-five-minutes represents a conservative trade-off to avoid microstructure biases and, at the same time, maintain a reasonable number of observations to ensure an accurate estimation of the realised betas (Bollerslev et al., 2008; Patton and Verardo, 2012). The twenty-five-minute quotes for the stocks and indexes are based on the National Best Bid Offer (NBBO) mid-point retrieved from TAQ quotes (Holden and Jacobsen, 2014). This methodology avoids both price bounces and the effect of the bid-ask spreads arising from lack of liquidity.

To test the first hypothesis (market-sample), I use the SPDR as a proxy for the S&P 500, which is an ETF traded on AMEX under the ticker SPY. Meanwhile, for the second hypothesis (sector-sample), I use ETFs that mimic the various sectors.[9] I match the ETFs according to their two-digit Global Industry Classification Standard (GICS) sector code taken from Compustat and the description provided by Vanguard and SPDR. Using the ETFs to calculate the betas has two main benefits. First, the quotes of ETFs are observed approximately every second, which ease the synchronisation between the stocks' mid-points with the index's mid-points, avoiding nonsynchronous observations when estimating the covariance (Bannouh et al., 2012). Secondly, the high liquidity of the ETFs can be assumed to be free from microstructure noise.

Table 1.1 reports the sample statistics of the betas for the two samples in each year.[10] The means, medians and standard deviation for the market-sample are consistently higher than for the sector sample. A higher (lower) standard deviation and means might lead to a

[9]I report the full list of the mimicking sectors and their two-digit GICS sector code in Appendix A.3.

[10]The top and bottom 1% of the distributions are winsorised. The number of announcements *No. Ann.* and average size are not reported for the sector-sample since they are the same as the market-sample.

higher (lower) variation of the realised beta over the event window. Based on Table 1.1, the information conveyed by the SEO news seems to be more relevant for the market-sample rather than for the sector-sample, which might indicate that the variation in beta will be larger for the former sample.

Table 1.2 shows the correlations between the variables used in this study, where above (below) the diagonal the pairwise correlations about market (sector) along with the p-values between brackets are reported. The table shows that the realised beta is positive and statistically significant correlated, at least at the 10% level with: stock, market and ETFs returns, first lag of the realised beta, volume and turnover. However, the size is not statistically significant across the two samples, and the bid-ask spread is statistically significant only for the sector-sample. On average, the correlation between the beta, stock and market returns is higher for the market-sample than for the sector sample.

For the identified companies, I select the SEO news from the Dow Jones newswire archive, with a company relevance score of at least 80%, as in Fang and Peress (2009). According to Chan (2003), this dataset is by far the most comprehensive data source regarding stocks, and it can be seen as an approximation of public news for traders. Furthermore, it is commonly used among scholars in event studies since it contains news from various sources such as newspapers, periodicals and newswires. Hence, I interpret the days and time of the published news as days and time of the company announcement.

Tetlock (2011) points out the difference between news and coverage. According to the author, news is defined as non-public information that becomes public for the first time, while coverage refers to repackaging or reprinting of old news. To avoid the introduction of stale news, as well as having an overlapping announcement in the event window for the same company about the same event, I discard the effect of coverage by applying the following criterion.[11] I define an announcement date t_0 as the first time that an article is published about company x regarding event z. I then check if other journals have published the same news on the same date; if so, I take the former. From day t_0 to date t_{22} I do not take into account any other news about event z concerning firm x, since I consider this second news as stale. From day t_{22} I reset the counter, so the next article about company x on event z is accounted as a new event.

This sample selection identifies a total of 530 news items writing about SEOs, for a

[11] A similar sampling criterion is applied by Peyer and Vermaelen (2008).

total of approximately 98% of unique companies. The distribution of the announcement is reported in the last column of Table 1.1. The number of announcements reported in Table 1.1 are plotted in Figure 1.1 at a monthly frequencies. Figure 1.2 shows the distribution of the announcements at quarterly frequency for each sector individually. Interestingly, except for one year, in the remaining three years there is a cluster in both time and industry. Companies mostly announce SEO programs in the first quarter. This is the preferred period of the year, since it is the moment that companies file their annual report; hence, when the information asymmetry is at its lowest level. Furthermore, Figure 1.3 shows a negative return for the announcing company on the event date, while the market shows a slight positive return. The opposite direction between the stock and market returns might lead to a negative variation in beta.

One of the major differences between this paper and most other research investigating SEO programs is the sample selection. Here, I do not distinguish whether the news is writing about the implementation or willingness of the company to undertake the program. This feature allows the effect of firm-level news to br more broadly studied.

1.6 Results

In this section I present the results of the paper. The tables report the number of observations used to estimate the model, the adjusted R-squared and between brackets the t-statistics calculated as in Petersen (2009). The dashed lines in the figures represent the 95% confidence bounds.

I start by presenting the results for the market-sample, which are used to test the first hypothesis. I then present the results for the sectors-sample, used to test the second hypothesis.[12] I also demonstrate the implication of the change in beta documented in the market-sample when calculating the Abnormal Return (AR) over the event window. Finally, I illustrate the effect of firm-level news sentiment in the systematic risk variation.

[12] In Appendix A.4, I present the results for the market and sector sample for buyback announcements.

1.6.1 Market Sample

Table 1.3 shows the results for the market-sample. The beta of the announcing company increases nine days before by 0.201 (t-stat of 2.88) then it reverts back to its long-run average except on the announcement day where it decreases by -0.334 (t-stat of -2.61). Hence, investors predict well before the upcoming event, and reflect the same information about the announcing firm on non-announcing companies before the event date. However, on the announcement date, investors trade non-announcing companies in the opposite way to the announcing company. This result is supported by Figure 1.3, where the stock and market return moves in the opposite direction on the announcement date. To ease the understanding of the table, Figure 1.4 plots the coefficients of Table 1.3.[13]

The beta variation around SEO is very different compared to earnings, where the beta spikes in the event date by 0.150 (t-stat of 3.02).[14] This implies that, when investors observe SEO news, they sell the announcing stock and buy non-announcing stocks, leading to a decrease in systematic risk that is higher (in absolute terms) compared to earnings. The fact that around SEO announcements the systematic risk varies more than around earnings, confirms that when investors cannot form expectations the variation in beta is higher.

From these results, I conclude that there is possible information leakage before an SEO announcement. Moreover, investors positively interpret the SEO announcement for non-announcing companies, which casts a drop in beta on the announcing date. It is also possible to notice that the stock returns reported in Figure 1.3 move exactly as the beta. Hence, the drops in returns might be due to a drop in systematic risk. Furthermore, these results are supported by Carlson et al. (2010) and Eckbo et al. (2000). Interestingly, the drops reported by the former authors is circa 30%, which is almost the same as the one documented in this paper. It is, however, true that in their paper the systematic risk shifts level while here, after the announcement, the beta revert backs to its pre-announcement level. I believe that - from an asset-pricing point of view - the beta reverts back to its pre-announcing level since once the new information is incorporated into announcing and

[13]In Appendix A.5, I show the results of the cross-sectional variation in realised beta when the companies are sorted according to characteristics, such as turnover, realised volatility and size.

[14]The comparison is based on the results for earnings announcement in 2006 reported in Appendix A.2. The results from 2006 to 2008 reported in Appendix A.1 show an increase of 0.092 (t-stat of 3.21). While, Patton and Verardo (2012) in Table 2 report an increase of 0.162 (t-stat of 8.08) for the period from 1996 to 2006.

non-announcing firms the comovement reverts back to its previous long-run average.

1.6.2 Sectors Sample

Table 1.4 shows the results for the sector sample. The beta of the announcing company never deviates from its long-run average, expect on the news date where it decreases by -0.230 (t-stat of -2.10). Hence, also for the sector sample investors positively interpret the news about the SEO for non-announcing firms.[15]

When the two samples are compared, some differences emerge. For the market sample, the beta increases nine days before and then decreases on the announcement date. In the sectors sample, the first variation disappears and the drop in the announcement date is smaller and less statistically significant. The fact that, in the sector-sample, the first variation in beta disappears, might indicate that this variation in the market-sample is spurious.

A smaller decrease in beta on the announcement date implies that the information conveyed by firm-level news reveals less relevant information about companies operating in the same sector. Hence, investors trade companies outside the sector casting a smaller variation in beta for the sector sample. This implies that investors pay attention to sector-wide news, which reduces the surprise, emerging from SEO announcements. Such a conclusion is supported by Engelberg et al. (2018b) who assert that companies in the same sector are priced more accurately, given the higher attention dedicated by analysts and investors to study categories of stocks rather than companies (Piotroski and Roulstone, 2004). This result confirms the prediction that companies in the same sector are priced more fairly, and that investors move their wealth away from an overvalued sector.

1.6.3 Abnormal Return with Time-Varying Beta

Having a reliable and precise measure of systematic risk is of critical importance when calculating the AR. A standard approach to calculate the AR is to estimate the market beta first:

$$r_t = \alpha_t + \beta_t * r_{m,t} + \epsilon_t \tag{1.5}$$

[15] To better understand the difference between sectors, I perform a more granular analysis, which accounts for the variation of the realised beta in each individual sector. The results are reported in Appendix A.6.

where, r_t is the excess return of the company at time t, $r_{m,t}$ is the excess return of the market, β_t is the systematic risk of the company and α_t is the intercept of the model. The subscripts in the coefficients of the model imply that the model is estimated through a daily rolling-window and the coefficients are stored. More precisely, I store the twenty-one coefficients estimated over the event window, using the previous 252 daily returns and one day rolling window. Once the parameters are estimated, the AR is calculated as follows:

$$AR_t = r_t - (\hat{\alpha}_t + \hat{\beta}_t * r_{m,t}) \tag{1.6}$$

where $\hat{\alpha}_t$ and $\hat{\beta}_t$ are the parameters estimated with Equation 1.5.

It is easy to see that if the AR of the company is calculated according to the two equations above, the variation in beta documented in this paper is neglected. Specifically, if the change in beta reported in Table 1.3 is economically meaningful, then the negative - and probably abnormal - return reported in Figure 1.3 should shrink.

To test whether, after accounting for the time-varying beta, there is a reduction in AR, I first estimate Equation 1.5 and 1.6 for each company separately. I use the same companies as in the market-sample, except that for this analysis I require that a company has 252 daily returns before the event window. In total I identify 352 companies.

Under column IR in Table 1.5, I report the cross-sectional mean of the AR in each of the twenty-one days along with the p-values between brackets. The AR is statistically significant at the 5% level five days before, and seven and eight days after the event date, with a major drop of -2.6% on the announcement date, which is statistically significant at the 10% level.

To account for the high-frequency variation in beta, I rely on a similar methodology to Phin et al. (2018). I calculate the intercept of the model for each company as follows:

$$\hat{\alpha}_T = T^{-1} \sum_{t=1}^{T} (r_t - \hat{\beta}_t * r_{m,t}) \tag{1.7}$$

where $\hat{\alpha}_T$ is the 30 days average intercept of the company before the event window. I then calculate the AR over the event window for each company:

$$AR_t = r_t - (\hat{\alpha}_T - R\hat{\beta}_t * r_{m,t}) \tag{1.8}$$

where $R\hat{\beta}_t$ is calculated as the long-run average beta of the company plus the variation from Equation 1.4. I report the results of Equation 1.8 in Table 1.6 under the column RB for each of the twenty-one days in the event window. Notably, the AR in the event day is now -1.12% and statistically significant at the 10% level, which represents a reduction of more than 50% compared to the standard methodology. Stated differently, the AR in the event date documented through a rolling window approach is partially due to the change in beta. Furthermore, the other three statistically significant ARs disappear.

I also document the importance of the time-varying beta when calculating the Cumulative Abnormal Return (CAR), and report the results in Figure 1.11. The figure shows the CAR in % over the event window when the AR is calculated with Equation 1.6 (dashed line) and when calculated with Equation 1.8 (solid line). Figure 1.11 depicts a completely different pattern between the two series. When the standard approach is used, the AR starts to decrease four days before the event and up to ten days after. Contrarily, when the CAR is estimated by accounting the time-varying beta, the CAR drops around the announcement date but, overall, is steady over the entire event window. In conclusion, this analysis shows that the variation document in this paper is statistically significant and economically important when evaluating the AR around information flow.

1.6.4 Effect of Sentiment on Beta

In this section, I investigate whether the sentiment of firm-level news explains the variation in systematic risk. To the best of my knowledge, no previous research links firm-level sentiment to the high-frequency change in systematic risk. Using the Baker and Wurgler (2006) sentiment index, Carlson et al. (2010) argue that market sentiment is negatively associated with beta after the issuance of security.

I hypothesis that investors judge the valuation of the company from the news sentiment. Consequently, the more pessimistic the news, the larger the investor reaction, which leads to a higher drop in systematic risk. Because I track news sentiment not only on the event date, but on the entire event widow, I also account for the sentiment surrounding the event, similar to Carlson et al. (2010)

I select all news about a given company with a company relevance score of at least 80% in the event window. In case of days without news, I carry forward the last available sentiment indicator. Such an approach leads to a time-series of sentiment indicators for each company over the event window. Following the same criterion as for the event selection, if the news is reported while the market is closed, I input the sentiment of that news to the next trading day.

The mean (median) sentiment over the event window is -0.077 (-0.001) with a standard deviation of 0.533, while the average sentiment in the event day is -0.122. Figure 1.6 shows the cross-sectional average sentiments across all companies over the event window. From the figure, it emerges that over the event window the sentiment is always negative, and it drops on the announcement day. The fact that the sentiment is remarkably negative on the event date matches the return pattern reported in Figure 1.3 or, more generally, the well-known drop in price around SEO announcements.

When the sentiment indicator is added as a control variable in Equation 1.4, it reveals that the sentiment does not explain the variation in systematic risk since it is indistinguishable from zero.[16] This implies that sentiment is a useful indicator for returns, but it does not explain the change in systematic risk. Thus, investors should closely monitor the event taxonomy to understand the systematic risk fluctuation of the company.

1.7 Robustness Test

In the first subsection, I check whether the change in realised beta is due to any omitted but potentially relevant variables. In the second subsection, I test investors' trading proclivity move the beta, through two placebo tests.

1.7.1 Possible Explanatory Variables

Some company-omitted variable might drive the results presented in this paper. Denis and Kadlec (1994) demonstrate that, after controlling for the trading volume and bid-ask spreads the variation in systematic risk disappears. Forbes and Rigobon (2002) show that the volatility of the stock might affect the covariance estimates. For a similar reason, I also control for the realised volatility of the market. To check that the results are not driven by autocorrelation in realised betas, I control for the first lag of the beta (Andersen et al.,

[16]Results available on request.

2006). Finally, I include firm size to control for omitted firm-specific factors. The company size is positively associated with the firm's information environment, such as media and analyst coverage (Piotroski and Roulstone, 2004).

Table 1.6 reports the results for the market-sample. The coefficients are of similar magnitude and statistical significance after controlling for all the above variables, except for the lagged value of beta. Once the first lag of the beta is used as a controlling variable the spike nine days before the event date disappears.[17]

Table 1.7 shows the results for the sector-sample. When controlling for the realised volatility of the stock, the variation in the event date becomes equal to -0.198 (t-stat of -1.81). Furthermore, when controlling for all variables simultaneously, the beta decreases (in absolute terms) to -0.188 (t-stat of -1.72). Since the coefficient on the event date is only statistically significant at the 10% level, it could be argued that investors only marginally trade companies in the same sector. This confirms the assertion of the second hypothesis, that investors are aware of the contagion effect around an SEO, and mostly trade companies outside the sector.

1.7.2 Placebo Tests

The first placebo test uses the the death of a board member announcement. The second placebo test investigates whether the variation in beta is due to an artefact of the methodology. To do so, I randomise the companies and the event selection.

1.7.2.1 Death of Board Member

To make this test as idiosyncratic as possible, I take the death of a board member that was in charge of only one mandate at the time of death. Unless this announcement is interpreted as an erosion of the competitive advantage of the company, resulting in a possible gain for competitors, it should not communicate any relevant information about non-announcing firms. Another exception is given by the fact that the death of a board member will reveal important information about the market cash flow (Savor and Wilson, 2016). If either of these two assumptions is violated, then the systematic risk will vary in

[17]The standard errors are checked with the Arellano and Bond (1991) methodology, to account for the feedback of the lagged variable in the dependent variable.

the event window.

I define a board member as a person that has a sensible position in the company, or that creates value for shareholders.[18] I obtain the death of board members from BoardEx, by matching the day of the death with the day that they resigned from the position.[19]

I use the same sampling methodology described in Section 2.4. However, since this announcement is observed sporadically, I take into account the period from January 1, 2006 to December 31, 2014. This sampling methodology results in 305 death events, or else circa two events per month.[20] Figure 1.7 shows the distribution of the event at a monthly frequency. The highest number of deaths is observed over 2007 and 2008, and over the entire period there is at least one announcement per month. Moreover, every company appears in the sample only once. Panel A of Table 1.8 shows the sample statistics of the realised beta.

I estimate Equation 1.4 with 7,038 daily observations and report the results in Table 1.9. The Table shows that, over the estimation window, the systematic risk varies ten and six days before the event day, where the coefficients are statistically significant at the 10 and 5% level.[21] The highest increase is observed nine days before the announcement by 0.166 (t-stat of 2.01). The coefficients of Table 1.9 are plotted in Figure 1.8. The figure shows that the coefficients are fluctuating around zero. To summarise, the result of this placebo test indicates that there is no clear link between the death of the CEO and the change in systematic risk, which implies that the main results of this paper are driven by investors' trading proclivity.

1.7.2.2 Random Dates

The second placebo test randomises the selection of firms and dates. I select 500 companies from those that appear at least once in this study, and I assign them a random date (i.e., non-event date) between January 1, 2010 and December 31, 2013. The idea is to assign to each company a date, and use it as event date. Furthermore, I check that around the

[18] More precisely I take into account the following rules: Senior Manager, Executive Director, Supervisory Director, Senior Manager as defined by BoardEx.

[19] I do not distinguish between sudden death (i.e., unexpected) and a death that is expected (i.e., after hospitalisation).

[20] The number of deaths is comparable to Nguyen and Nielsen (2010) for the same period.

[21] Given the results from the previous robustness test, I control for the first lag of the realised beta and check the standard errors with the Arellano and Bond (1991) methodology.

twenty-one days of the assigned non-event dates there are no identifiable announcements made by the company (i.e., earnings announcements, buyback). The aim of this placebo test is to investigate whether the beta varies also around the non-event date. After applying the same filtering steps discussed in Section 2.4 and replicating the same characteristics of the SEOs samples (98% of the companies appear only once and 55% announce after 4:00 p.m.), I identify 378 non-event dates for a total of 8,744 observations.

The distribution of the non-event dates is plotted in Figure 1.9, while the sample statistics are reported in Panel B of Table 1.8. The mean of the betas used in this analysis is similar to the SEO market-sample, but their standard deviation is almost 50% lower. This indicates that around non-event dates the beta is more stable than for the SEO market-sample.

I estimate Equation 1.4 and report the results in Table 1.10. None of the coefficients are significantly different from zero at the 10% level. Figure 1.10 plots the coefficients. From the figure, it is clear that when the variation in beta is estimated around non-event dates, the beta does not vary from its long-run average. Therefore, I conclude that what moves the beta around SEO news is investors' trading proclivity.

1.8 Concluding Remarks

This paper investigates the cross-sectional variation in the daily realised beta around firm-level news writing about an SEO program for companies listed on the NYSE, AMEX and NASDAQ from January 1, 2010 to December 31, 2013. I calculate the realised beta through intraday returns, with the S&P 500 (i.e., market-sample) and the sectors mimicking portfolios (i.e., sector-sample). I find that the realised beta varies by a statistically and economically significant amount around this event for both samples.

The variation in systematic risk documented is short-lived and, therefore, difficult to capture at a lower frequency. I show that, for the market-sample, the systematic risk decreases by 33.4% on the event day. Hence, the variation around unscheduled news is larger (in absolute terms) compared to scheduled news and it is in the opposite direction. For the sector-sample, the systematic risk also decreases on the event date. However, the variation in beta is smaller and less statistically significant compared to the market-sample, which implies that investors understand that companies in the same sector are likely to be overvalued too. Therefore, they trade companies outside the given sector.

The results reported in this paper are completely different from those documented by Patton and Verardo (2012), who show that the beta in the earnings announcement date increases. This implies that the learning model provided by the authors does not support the beta variation of this paper but, instead, is supported by the findings of Eckbo et al. (2000) and Carlson et al. (2010).

I demonstrate that after accounting for the change in high-frequency beta, the abnormal negative return observed around SEO announcements is reduced by more than 50%. Furthermore, I show that the sentiment of firm-level news does not explain the variation in systematic risk. This second finding implies that investors should evaluate their risk exposure from a news taxonomy point of view, rather than from a sentiment perspective.

In Appendix A.2, I show that investors pay less attention to firm-level news over a recession period, since the variation in beta around earnings announcements is indistinguishable from zero. This result is consistent with the idea that investors closely monitor news that has a larger effect on stock prices (i.e., macro-level news). In Appendix A.4, I report empirical evidence that the beta also varies around buyback announcements. Future research could examine these two effects in greater detail, in order to better understand the links between periods, firm-level news and systematic risk.

My results have important implications for portfolio managers for hedging purposes, especially for those implementing market-neutral portfolios. For example, they should consider the risk of unrelated companies announcing SEO programs. This announcement might shift the comovement between stocks in their portfolio, exposing them to short-lived undesired risks. Finally, I suggest scholars check whether, around the given event, the systematic risk of the company varies before reporting the size of the abnormal return detected in their research.

Table 1.1: Summary Statistics

The table shows the sample statistics of the companies considered in the analysis, from January 1, 1990 to December 31, 2017 where: *Size* is the average size (in thousands) calculated as number of outstanding shares multiplied by closing price; *Mean* is the mean of the realised beta; *Std* is the standard deviation of the realised beta; *Skew* is the Skewness of the realised beta; *Q1* is the value of the first quartile of the realised beta; *Median* is the median of the realised beta; *Q3* is the value of the third quartile of the realised beta, and *No. Ann* is the total number of announcement per year. **Panel A** shows the sample statistics when the realised beta is calculated with the SPY (i.e., market-sample), while **Panel B** shows the same statistics when the beta is calculated against the various sectors mimicking portfolios (i.e., sector-sample). The returns used to calculate the realised beta are from the TAQ dataset for the NYSE, AMEX and NASDAQ, by excluding the bottom decile of illiquid stocks according to the Amihud illiquidity ratio (Amihud, 2002), stock with a price less than five dollars and companies classified in the Utilities or Banking sector. The realised beta is calculated from the National Best Bid Offer (NBBO) mid-point as in Holden and Jacobsen (2014) with a twenty-five minute sample frequency from 9:45 a.m. to 4:00 p.m. plus an overnight return, for a total of sixteen intraday observations. The bottom and top percentile of the distribution are winsorised.

Panel A: Market-Sample

Year	***Size***	***Mean***	***Std***	***Skew***	***Q1***	***Median***	***Q3***	***No. Ann***
2010	8,667.77	1.110	1.423	-0.026	0.434	1.032	1.709	164
2011	7,057.35	0.988	1.354	0.492	0.351	0.916	1.527	121
2012	8,866.14	0.987	1.468	-1.136	0.374	0.929	1.599	119
2013	6,090.55	1.040	1.973	-0.002	0.151	0.937	1.869	126
							Tot.	530

Panel B: Sectors-Sample

Year	***Size***	***Mean***	***Std***	***Skew***	***Q1***	***Median***	***Q3***	***No. Ann***
2010	-	0.879	1.142	-0.386	0.371	0.833	1.374	-
2011	-	0.799	1.124	-0.076	0.274	0.763	1.252	-
2012	-	0.859	1.300	-1.523	0.345	0.819	1.368	-
2013	-	0.916	1.747	-0.509	0.143	0.815	1.637	-

Table 1.2: Correlation

The table shows the pairwise correlation between from January 1, 2010 to December 31, 2013 of the variables used in the analysis, where above (bellow) the diagonal the correlations of the market-sample (sector-sample) are reported. *Beta* is the daily realised beta calculated from the National Best Bid Offer (NBBO) mid-point as in Holden and Jacobsen (2014) with a twenty-five-minute sampling frequency from 9:45 a.m. to 4:00 p.m. plus an overnight return, for a total of sixteen intraday observations; *Ret* is the daily log return of the stocks; *Mkt* is the daily log return of the market or the sector ETFs; $Beta_{t\text{-}1}$ is the one-day lagged realised beta; *Vol* is the daily stock trading volume; *Size* is the daily closing price multiplied by the number of share outstanding; *Bid-Ask* is the average daily bid-ask spread; *Turn* is the stock turnover calculated as the daily trading volume divided by the number of shares outstanding. Between brackets the p-value is reported. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively.

	Beta	***Ret***	***Mkt***	$\boldsymbol{Beta_{t\text{-}1}}$	***Vol***	***Size***	***Bid-Ask***	***Turn***
Beta		0.051***	-0.028**	0.137***	0.028***	-0.012	0.014	0.054***
		(0.001)	*(0.002)*	*(0.001)*	*(0.001)*	*(0.187)*	*(0.105)*	*(0.001)*
Ret	0.020**		0.355***	0.028***	-0.015	0.008	0.025**	-0.063***
	(0.020)		*(0.001)*	*(0.001)*	*(0.102)*	*(0.343)*	*(0.004)*	*(0.001)*
Mkt	-0.025**	0.366***		0.004	-0.003	0.011	-0.002	-0.038***
	(0.004)	*(0.001)*		*(0.601)*	*(0.780)*	*(0.202)*	*(0.888)*	*(0.001)*
$\boldsymbol{Beta_{t\text{-}1}}$	0.120***	0.023*	0.001		0.029***	-0.010	-0.010	0.042***
	(0.001)	*(0.008)*	*(0.900)*		*(0.001)*	*(0.268)*	*(0.286)*	*(0.001)*
Vol	0.037***	-0.015	-0.009	0.035***		0.409***	-0.040***	0.215***
	(0.001)	*(0.102)*	*(0.350)*	*(0.001)*		*(0.001)*	*(0.001)*	*(0.001)*
Size	0.010	0.008	0.012	0.010	0.409***		-0.048***	-0.067***
	(0.231)	*(0.343)*	*(0.152)*	*(0.251)*	*(0.001)*		*(0.001)*	*(0.001)*
Bid-Ask	0.017**	0.025***	-0.001	-0.010	-0.040***	-0.048***		-0.030***
	(0.042)	*(0.004)*	*(0.967)*	*(0.291)*	*(0.001)*	*(0.001)*		*(0.001)*
Turn	0.042***	-0.063***	-0.042***	0.037***	0.215***	-0.067***	-0.030***	
	(0.001)	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*	

Table 1.3: Changes in Beta - Market Sample

The table shows the results of Equation 1.4 from January 1, 2010 to December 31, 2013 for the twenty-one-day around firm-level news writing about SEO, for a total of 530 announcements. The panel regression accounts for the realised beta of the company as a dependent variable calculated with sixteen intraday returns benchmarked against the S&P 500, which is regressed against the firm-year fix effect, as well as a dummy for each day in the estimation window (equals one if it is the announcement day and zero otherwise). *Beta* represents the deviation of the systematic risk from long-run average in a given day (i.e., *Event Day*) over a twenty-one day estimation window, thus Event Day 0 is the announcement day. *No. Obs* represents the number of observations used to estimate the panel regression of daily realised beta, and $Adj.R^2$ denotes the Adjusted R-squared. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm-year and days.

Event Day	***Beta***	***Event Day***	***Beta***	***Event Day***	***Beta***
-10	-0.012	-3	-0.046	4	-0.034
	(-0.16)		*(-0.57)*		*(-0.53)*
-9	0.201***	-2	0.049	5	-0.004
	(2.88)		*(0.63)*		*(-0.05)*
-8	0.077	-1	0.011	6	0.060
	(1.14)		*(0.14)*		*(0.90)*
-7	0.090	0	-0.334***	7	0.095
	(1.29)		*(-2.61)*		*(1.43)*
-6	0.082	1	-0.092	8	-0.035
	(1.23)		*(-1.44)*		*(-0.55)*
-5	0.050	2	-0.029	9	0.094
	(0.70)		*(-0.40)*		*(1.39)*
-4	0.079	3	0.013	10	0.043
	(0.92)		*(0.20)*		*(0.62)*
No.Obs	12,702				
$\boldsymbol{Adj.R^2}$	0.4045				

Table 1.4: Changes in Beta - Sectors Sample

The table shows the results of Equation 1.4 from January 1, 2010 to December 31, 2013 for the twenty-one-day around a firm-level news writing about SEO, for a total of 530 announcements. The panel regression accounts for the realised beta of the company as a dependent variable calculated with sixteen intraday returns benchmarked against the sector ETF (see Appendix A.3) of each company according to its two-digit GICS sector code, which is regressed against the firm-year fix effect, as well as a dummy for each day in the estimation window (equals one if it is the announcement day and zero otherwise). *Beta* represents the deviation of the systematic risk from long-run average in a given day (i.e., *Event Day*) over a twenty-one day estimation window, thus Event Day 0 is the announcement day. *No. Obs* represents the number of observations used to estimate the panel regression of daily realised beta, and $Adj.R^2$ denotes the adjusted R-squared. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm-year and days.

Event Day	***Beta***	***Event Day***	***Beta***	***Event Day***	***Beta***
-10	-0.012	-3	-0.073	4	-0.041
	(-0.48)		*(-1.08)*		*(-0.73)*
-9	0.022	-2	-0.003	5	0.012
	(0.37)		*(-0.03)*		*(0.24)*
-8	-0.004	-1	-0.039	6	-0.007
	(-0.06)		*(-0.56)*		*(-0.12)*
-7	0.057	0	-0.230^{**}	7	0.060
	(0.91)		*(-2.10)*		*(1.12)*
-6	0.035	1	-0.066	8	-0.061
	(0.55)		*(-1.18)*		*(-1.13)*
-5	0.044	2	0.054	9	0.023
	(0.67)		*(0.98)*		*(0.41)*
-4	0.044	3	-0.037	10	0.010
	(0.61)		*(-0.62)*		*(0.20)*
No.Obs	12,702				
$\boldsymbol{Adj.R^2}$	0.3869				

Table 1.5: Abnormal Return - Realised Beta vs. Integrated Beta

The table shows the Abnormal Return (AR) on a given day (*Event Day*) over twenty-one-day around the day that a firm-level news about SEO is reported, thus Event Day 0 is the announcement day. Under columns *IB* (integrated beta) and *RB* (realised beta) the results of Equation 1.6 and 1.8 are reported, along with the p-value between brackets. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively.

Event Day	***IB***	***RB***	***Event Day***	***IB***	***RB***	***Event Day***	***IB***	***RB***
-10	-0.272	-0.071	-3	-0.929	0.301	4	-0.173	-0.516
	(0.60)	*(0.85)*		*(0.35)*	*(0.57)*		*(0.60)*	*(0.19)*
-9	0.545	0.599	-2	-0.326	-1.105	5	-0.878	0.419
	(0.54)	*(0.52)*		*(0.41)*	*(0.14)*		*(0.38)*	*(0.43)*
-8	-0.087	-0.050	-1	-1.105	0.137	6	-0.299	0.749
	(0.83)	*(0.90)*		*(0.23)*	*(0.77)*		*(0.73)*	*(0.09)*
-7	-0.338	-0.152	0	-2.592*	-1.117*	7	0.745**	-0.350
	(0.48)	*(0.66)*		*(0.04)*	*(0.02)*		*(0.01)*	*(0.66)*
-6	-0.479	0.465	1	-0.322	1.626	8	0.779**	-0.123
	(0.59)	*(0.34)*		*(0.83)*	*(0.11)*		*(0.01)*	*(0.85)*
-5	0.588**	0.039	2	-0.180	-0.307	9	-0.858	-0.083
	(0.01)	*(0.94)*		*(0.64)*	*(0.41)*		*(0.25)*	*(0.82)*
-4	0.067	0.136	3	-0.461	0.250	10	-1.691	1.491
	(0.87)	*(0.65)*		*(0.50)*	*(0.45)*		*(0.38)*	*(0.32)*

Table 1.6: Robustness - Changes in Beta (Market Sample)

The table shows the results of Equation 1.4 from January 1, 2010 to December 31, 2013 for the twenty-one-day around firm-level news writing about SEO, for a total of 530 announcements. The panel regression accounts for the realised beta of the company as a dependent variable, calculated with sixteen intraday returns benchmarked against the S&P 500, regressed against the firm-year fix effect, as well as a dummy for each day in the estimation window (equals one if it is the announcement day and zero otherwise). Each column represents the variable that is added in Equation 3.5 through the matrix $\mathbf{X}_{i,t}$. *Trd* is the daily trading volume of the stock; $Beta_{t\text{-}1}$ is the one-day lagged realised beta; *Size* is the average size (in thousands) calculated as number of outstanding shares multiplied by closing price; RV_m is the realised volatility of the SPY; RV_i is the realised volatility of the stock; *Bid-Ask* is the bid-ask spread; *All* are all the variables together. Under each column the deviation of the company systematic risk from its long-run average is reported. *Day* represents the twenty-one days estimation, thus Day 0 is the announcement day. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm-year and days.

Day	*Trd*	$Beta_{t\text{-}1}$	*Size*	RV_m	RV_i	*Bid-Ask*	*All*
-10	-0.011	-0.016	-0.012	-0.012	-0.014	-0.014	-0.019
	(-0.14)	*(-0.20)*	*(-0.16)*	*(-0.15)*	*(-0.17)*	*(-0.18)*	*(-0.24)*
-9	0.201***	0.025	0.201***	0.201***	0.201***	0.203***	0.021
	(2.88)	*(1.18)*	*(2.82)*	*(2.88)*	*(2.89)*	*(2.92)*	*(1.26)*
-8	0.077	0.072	0.077	0.077	0.077	0.074	0.066
	(1.14)	*(1.05)*	*(1.14)*	*(1.14)*	*(1.14)*	*(1.09)*	*(0.96)*
-7	0.091	0.086	0.090	0.090	0.088	0.098	0.091
	(1.29)	*(1.22)*	*(1.29)*	*(1.29)*	*(1.26)*	*(1.45)*	*(1.35)*
-6	0.082	0.079	0.083	0.083	0.080	0.080	0.074
	(1.23)	*(1.16)*	*(1.23)*	*(1.23)*	*(1.20)*	*(1.19)*	*(1.08)*
-5	0.050	0.046	0.050	0.050	0.047	0.042	0.034
	(0.70)	*(0.64)*	*(0.70)*	*(0.70)*	*(0.67)*	*(0.60)*	*(0.49)*
-4	0.080	0.076	0.079	0.080	0.072	0.089	0.076
	(0.92)	*(0.87)*	*(0.92)*	*(0.92)*	*(0.84)*	*(1.03)*	*(0.88)*

Day	*Trd*	$Beta_{t\text{-}1}$	*Size*	RV_m	RV_i	*Bid-Ask*	*All*
-3	-0.044	-0.050	-0.046	-0.046	-0.048	-0.054	-0.060
	(-0.55)	*(-0.62)*	*(-0.57)*	*(-0.57)*	*(-0.60)*	*(-0.67)*	*(-0.74)*
-2	0.051	0.046	0.049	0.049	0.046	0.046	0.041
	(0.66)	*(0.59)*	*(0.63)*	*(0.63)*	*(0.60)*	*(0.59)*	*(0.53)*
-1	0.013	0.007	0.011	0.011	0.007	0.021	0.014
	(0.17)	*(0.09)*	*(0.14)*	*(0.14)*	*(0.09)*	*(0.26)*	*(0.17)*
0	-0.321**	-0.337***	-0.334***	-0.333***	-0.345***	-0.328***	-0.332***
	(-2.52)	*(-2.63)*	*(-2.61)*	*(-2.61)*	*(-2.72)*	*(-2.59)*	*(-2.61)*
1	-0.088	-0.092	-0.092	-0.092	-0.091	-0.082	-0.075
	(-1.37)	*(-1.44)*	*(-1.43)*	*(-1.44)*	*(-1.43)*	*(-1.27)*	*(-1.17)*
2	-0.027	-0.032	-0.029	-0.029	-0.029	-0.023	-0.022
	(-0.38)	*(-0.44)*	*(-0.40)*	*(-0.40)*	*(-0.39)*	*(-0.31)*	*(-0.30)*
3	0.016	0.010	0.013	0.013	0.015	0.019	0.021
	(0.24)	*(0.16)*	*(0.20)*	*(0.20)*	*(0.23)*	*(0.28)*	*(0.31)*

Day	*Trd*	$Beta_{t\text{-}1}$	*Size*	RV_m	RV_i	*Bid-Ask*	*All*
4	-0.034	-0.038	-0.034	-0.034	-0.034	-0.021	-0.024
	(-0.52)	*(-0.57)*	*(-0.52)*	*(-0.53)*	*(-0.52)*	*(-0.33)*	*(-0.37)*
5	-0.005	-0.006	-0.003	-0.003	-0.002	0.003	0.002
	(-0.06)	*(-0.09)*	*(-0.04)*	*(-0.04)*	*(-0.02)*	*(0.05)*	*(0.03)*
6	0.059	0.057	0.060	0.060	0.061	0.069	0.066
	(0.88)	*(0.85)*	*(0.90)*	*(0.90)*	*(0.92)*	*(1.02)*	*(0.97)*
7	0.095	0.091	0.095	0.095	0.096	0.100	0.097
	(1.43)	*(1.36)*	*(1.43)*	*(1.43)*	*(1.44)*	*(1.50)*	*(1.45)*
8	-0.035	-0.039	-0.034	-0.034	-0.034	-0.028	-0.033
	(-0.56)	*(-0.61)*	*(-0.55)*	*(-0.55)*	*(-0.55)*	*(-0.44)*	*(-0.51)*
9	0.093	0.091	0.094	0.094	0.095	0.102	0.100
	(1.38)	*(1.34)*	*(1.39)*	*(1.39)*	*(1.40)*	*(1.49)*	*(1.46)*
10	0.042	0.039	0.043	0.043	0.044	0.049	0.046
	(0.61)	*(0.55)*	*(0.62)*	*(0.62)*	*(0.63)*	*(0.71)*	*(0.66)*

Table 1.7: Robustness - Changes in Beta (Sector-Sample)

The table shows the results of Equation 3.5 from January 1, 2010 to December 31, 2013 for the twenty-one-day around firm-level news writing about SEO, for a total of 530 announcements. The panel regression accounts for the realised beta of the company as dependent variable, calculated with sixteen intraday returns benchmarked against the various ETFs mimicking sectors (see Appendix A.3) of each company according to its two-digit GICS sector code, regressed against the firm-year fix effect, as well as a dummy for each day in the estimation window (equal one if is the announcement day and zero otherwise). Each column represents the variable that is added in Equation 3.5 through the matrix $\mathbf{X}_{i,t}$. *Trd* is the daily trading volume of the stock; $Beta_{t\text{-}1}$ is the one-day lagged realised beta; *Size* is the average size (in thousands) calculated as number of outstanding shares multiplied by closing price; RV_m is the realised volatility of the SPY; RV_i is the realised volatility of the stock; *Bid-Ask* is the bid-ask spread; *All* are all the variables together. Under each column the deviation of the company systematic risk from its long-run average is reported. *Day* represents the twenty-one days estimation, thus Day 0 is announcement day. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm-year and days.

Day	***Trd***	$Beta_{t\text{-}1}$	***Size***	RV_m	RV_i	***Bid-Ask***	***All***
-10	-0.029	-0.032	-0.031	-0.031	-0.027	-0.034	-0.029
	(-0.46)	*(-0.49)*	*(-0.48)*	*(-0.48)*	*(-0.43)*	*(-0.52)*	*(-0.47)*
-9	0.021	0.021	0.022	0.022	0.022	0.021	0.020
	(0.37)	*(0.36)*	*(0.38)*	*(0.38)*	*(0.37)*	*(0.37)*	*(0.35)*
-8	-0.004	-0.005	-0.004	-0.004	-0.004	-0.001	-0.001
	(-0.06)	*(-0.08)*	*(-0.06)*	*(-0.06)*	*(-0.05)*	*(-0.01)*	*(-0.01)*
-7	0.057	0.056	0.057	0.057	0.062	0.064	0.066
	(0.91)	*(0.89)*	*(0.91)*	*(0.91)*	*(1.01)*	*(1.06)*	*(1.09)*
-6	0.034	0.034	0.035	0.035	0.040	0.036	0.039
	(0.54)	*(0.52)*	*(0.55)*	*(0.55)*	*(0.64)*	*(0.56)*	*(0.61)*
-5	0.044	0.043	0.044	0.044	0.051	0.036	0.04
	(0.67)	*(0.65)*	*(0.68)*	*(0.67)*	*(0.79)*	*(0.55)*	*(0.63)*
-4	0.044	0.042	0.044	0.044	0.063	0.048	0.065
	(0.61)	*(0.59)*	*(0.61)*	*(0.61)*	*(0.92)*	*(0.67)*	*(0.95)*

Day	***Trd***	$Beta_{t\text{-}1}$	***Size***	RV_m	RV_i	***Bid-Ask***	***All***
-3	-0.071	-0.074	-0.072	-0.072	-0.067	-0.075	-0.069
	(-1.06)	*(-1.09)*	*(-1.08)*	*(-1.08)*	*(-1.02)*	*(-1.10)*	*(-1.05)*
-2	0.00	-0.004	-0.003	-0.003	0.005	-0.005	0.004
	(0.00)	*(-0.04)*	*(-0.03)*	*(-0.03)*	*(0.08)*	*(-0.07)*	*(0.06)*
-1	-0.037	-0.04	-0.039	-0.039	-0.027	-0.03	-0.018
	(-0.53)	*(-0.58)*	*(-0.56)*	*(-0.56)*	*(-0.40)*	*(-0.42)*	*(-0.26)*
0	-0.217**	-0.231**	-0.229**	-0.230**	-0.198*	-0.223**	-0.188*
	(-1.99)	*(-2.10)*	*(-2.10)*	*(-2.10)*	*(-1.81)*	*(-2.04)*	*(-1.72)*
1	-0.061	-0.066	-0.065	-0.066	-0.067	-0.053	-0.054
	(-1.10)	*(-1.18)*	*(-1.17)*	*(-1.18)*	*(-1.21)*	*(-0.96)*	*(-0.99)*
2	0.056	0.054	0.055	0.054	0.052	0.061	0.057
	(1.01)	*(0.96)*	*(0.98)*	*(0.98)*	*(0.93)*	*(1.08)*	*(1.01)*
3	-0.034	-0.038	-0.036	-0.037	-0.041	-0.031	-0.037
	(-0.58)	*(-0.64)*	*(-0.62)*	*(-0.62)*	*(-0.70)*	*(-0.52)*	*(-0.62)*

Day	***Trd***	$Beta_{t\text{-}1}$	***Size***	RV_m	RV_i	***Bid-Ask***	***All***
4	-0.041	-0.042	-0.041	-0.041	-0.042	-0.031	-0.035
	(-0.73)	*(-0.74)*	*(-0.72)*	*(-0.73)*	*(-0.77)*	*(-0.55)*	*(-0.62)*
5	0.010	0.011	0.012	0.012	0.007	0.018	0.01
	(0.22)	*(0.23)*	*(0.25)*	*(0.24)*	*(0.14)*	*(0.37)*	*(0.21)*
6	-0.009	-0.008	-0.007	-0.007	-0.011	-0.001	-0.008
	(-0.14)	*(-0.14)*	*(-0.12)*	*(-0.12)*	*(-0.19)*	*(0.00)*	*(-0.13)*
7	0.060	0.059	0.060	0.06	0.058	0.067	0.061
	(1.12)	*(1.10)*	*(1.13)*	*(1.13)*	*(1.07)*	*(1.24)*	*(1.12)*
8	-0.061	-0.062	-0.061	-0.061	-0.062	-0.054	-0.057
	(-1.14)	*(-1.14)*	*(-1.12)*	*(-1.13)*	*(-1.14)*	*(-0.98)*	*(-1.04)*
9	0.023	0.023	0.024	0.024	0.021	0.031	0.025
	(0.39)	*(0.39)*	*(0.41)*	*(0.41)*	*(0.36)*	*(0.53)*	*(0.42)*
10	0.010	0.009	0.010	0.010	0.008	0.017	0.01
	(0.19)	*(0.18)*	*(0.20)*	*(0.20)*	*(0.15)*	*(0.32)*	*(0.20)*

Table 1.8: Summary Statistics - Board Member Death and Random Dates

The table shows the sample statistics of the realised beta for the board member's death and the random dates. **Panel A** shows the sample statistics for the board member death sample from January 1, 2006 to December 31, 2014. I collect the death dates from BoardEx, and require that a board member - at the time of the death - had only one mandate. **Panel B** reports the sample statistics of the realised beta when the event dates and companies are selected randomly from January 1, 2010 to December 31, 2013. *Mean* is the mean; *Std* is the standard deviation; *Skew* is the skewness; *Q1* is the value of the first quartile; *Median* is the median; *Q3* is the value of the third quartile, and *No. Ann* is the total number of announcements per year.

Panel A: CEO Death

Year	***Mean***	***Std***	***Skew***	***Q1***	***Median***	***Q3***	***No. Ann***
2006	1.156	1.349	0.526	0.373	0.996	1.782	44
2007	0.911	1.071	0.404	0.322	0.823	1.464	46
2008	0.848	0.922	0.050	0.389	0.838	1.307	42
2009	0.982	0.941	0.503	0.458	0.898	1.466	34
2010	1.058	0.988	0.469	0.540	0.980	1.481	39
2011	1.092	1.087	0.248	0.528	0.963	1.657	31
2012	1.275	1.203	0.534	0.591	1.168	1.885	29
2013	0.956	1.221	0.356	0.350	0.934	1.519	30
2014	0.887	1.466	0.623	0.163	0.736	1.552	10
						Tot.	305

Panel B: Random Dates

Year	***Mean***	***Std***	***Skew***	***Q1***	***Median***	***Q3***	***No. Ann***
2010	1.089	0.858	0.234	0.604	1.050	1.547	105
2011	1.214	0.951	0.575	0.650	1.134	1.653	86
2012	1.083	1.165	0.380	0.457	0.958	1.661	91
2013	0.926	1.039	0.376	0.393	0.866	1.424	96
						Tot.	378

Table 1.9: Changes in Beta - Board Member Death

The table shows the results of Equation 1.4 estimated with 305 board member death announcements from January 1, 2006 to December 31, 2014. I collect the death dates from BoardEx, and require that a board member - at the time of the death - had only one mandate. The panel regression accounts for the realised beta of the company as a dependent variable calculated with sixteen intraday returns benchmarked against the S&P 500, which is regressed against the firm-year fix effect, as well as a dummy for each day in the estimation window (equal one if is the announcement day and zero otherwise). *Beta* represents the deviation of the systematic risk from long-run average in a given day (i.e., *Event Day*) over a twenty-one day estimation window, thus Event Day 0 is the announcement day. *No. Obs* represents the number of observations used to estimate the panel regression of daily realised beta, and $Adj.R^2$ denotes the adjusted R-squared. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm-year and days.

Event Day	***Beta***	***Event Day***	***Beta***	***Event Day***	***Beta***
-10	0.142^{*}	-3	-0.009	4	0.116
	(1.91)		*(-0.10)*		*(1.32)*
-9	-0.033	-2	-0.105	5	0.053
	(-0.39)		*(-1.39)*		*(0.67)*
-8	-0.050	-1	-0.065	6	0.012
	(-0.59)		*(-0.91)*		*(0.15)*
-7	-0.009	0	0.004	7	0.028
	(-0.10)		*(0.06)*		*(0.31)*
-6	0.166^{**}	1	-0.064	8	0.031
	(2.01)		*(-0.78)*		*(0.37)*
-5	0.051	2	0.035	9	0.036
	(0.48)		*(0.45)*		*(0.46)*
-4	0.041	3	0.046	10	0.030
	(0.42)		*(0.58)*		*(0.41)*
No.Obs	7,038				
$\boldsymbol{Adj.R^2}$	0.4885				

Table 1.10: Changes in Beta - Random Dates

The table shows the results of Equation 1.4 for the period starting from tJanuary 1, 2010 to December 31, 2013 for the twenty-one-day around 372 of random dates for random companies. Random dates are defined as days in which there are no identifiable company announcements. The panel regression accounts for the realised beta of the company as dependent variable calculated with sixteen intraday returns benchmarked against the S&P 500, which is regressed against the firm-year fix effect, as well as a dummy for each day in the estimation window (equal one if is the announcement day and zero otherwise). *Beta* represents the deviation of the systematic risk from long-run average in a given day (i.e., *Event Day*) over a twenty-one day estimation window, thus Event Day 0 is the announcement day. *No. Obs* represents the number of observations used to estimate the panel regression of daily realised beta, and *Adj.*R^2 denotes the adjusted R-squared. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm-year and days.

Event Day	***Beta***	***Event Day***	***Beta***	***Event Day***	***Beta***
-10	0.077	-3	-0.014	4	0.034
	(1.31)		*(-0.23)*		*(0.58)*
-9	0.052	-2	0.050	5	-0.055
	(0.97)		*(0.99)*		*(-0.91)*
-8	0.027	-1	0.019	6	0.024
	(0.47)		*(0.36)*		*(0.43)*
-7	0.051	0	0.084	7	0.032
	(0.92)		*(1.45)*		*(0.59)*
-6	0.074	1	0.008	8	-0.026
	(1.33)		*(0.14)*		*(-0.47)*
-5	0.038	2	0.057	9	0.023
	(0.66)		*(0.95)*		*(0.41)*
-4	0.037	3	-0.08	10	-0.002
	(0.62)		*(-1.43)*		*(-0.03)*
No.Obs	8,744				
Adj.R^2	0.6252				

Figure 1.1: Distribution of SEO

The figure plots the distribution of firm-level news writing about SEO from January 1, 2010 to December 31, 2013 at a monthly frequency. The event dates are from the published articles in the Dow Jones newswire for NYSE, AMEX and NASDAQ with a company relevance score of at least 80%. To be part of the sample a company must not lie on the bottom decile of the the distribution, when sorted according to the Amihud illiquidity ratio (Amihud, 2002), have a price greater than five dollar and not be classified in the Utilities or Banking sectors. These criteria return a total number of 530 firm-level news. The *x-axis* of the figure shows the year, and *y-axis* measures the number of news items in each month.

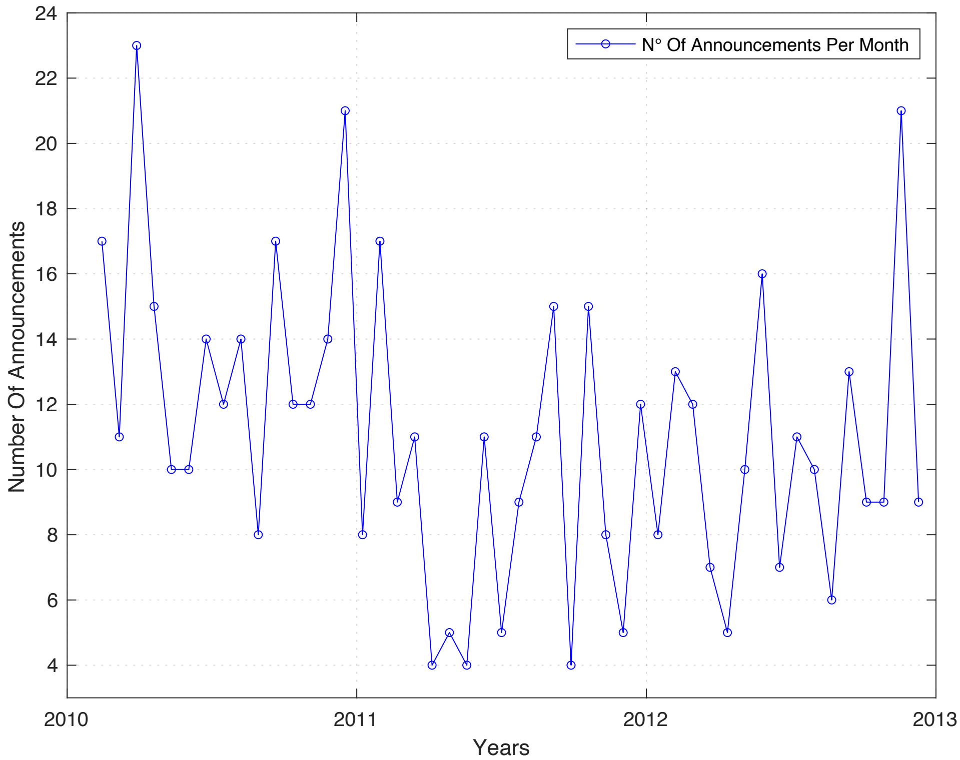


Figure 1.2: Distribution of SEO per Sector

The figure plots the distribution of firm-level news writing about SEO from January 1, 2010 to December 31, 2013 at a quarterly frequency for each individual sector. The event dates are from the published articles in the Dow Jones newswire for NYSE, AMEX and NASDAQ with a company relevance score of at least 80%. To be part of the sample a company must not lie on the bottom decile of the the distribution, when sorted according to the Amihud illiquidity ratio (Amihud, 2002), have a price greater than five dollar and not be classified in the Utilities or Banking sectors. These criteria return a total number of 530 firm-level news. The *x-axis* of the figure shows the quarters and years, and *y-axis* measures the number of news items in each quarter.

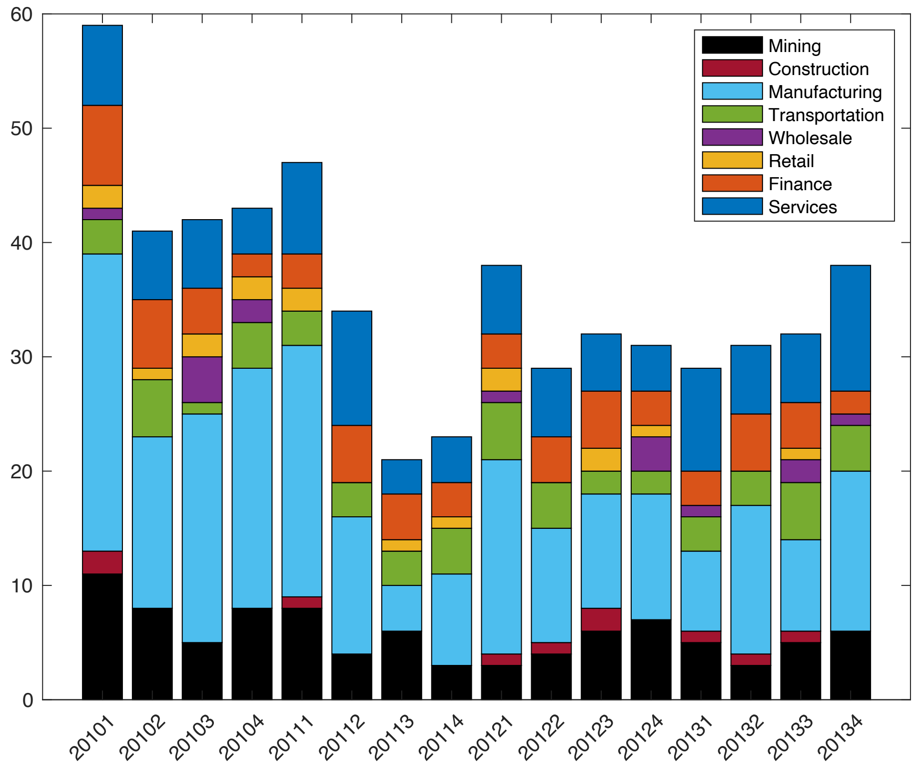


Figure 1.3: Return and Trading Volume Around SEO

The figure plots the daily open to close returns and volume for twenty-one days around firm-level news writing about SEO from January 1, 2010 to December 31, 2013. The daily returns are from the quote prices of the TAQ dataset and are calculated from the midpoint of the National Best Bid Offer (NBBO) from 9:45 a.m. to 4:00 p.m. and measured on the left-hand side of the figure. The blue line represents the cross-sectional average returns across all companies, while in red line represents the S&P 500 return. The dashed lines represents the 95% confidence bounds of the companies' returns. The histogram reports the average *Volume* across all companies in thousand, measured on the right-hand side of the figure. The *x-axis* at the bottom of the figure shows the days before and after the news day, thus *Event Date* equal 0 is the news date.

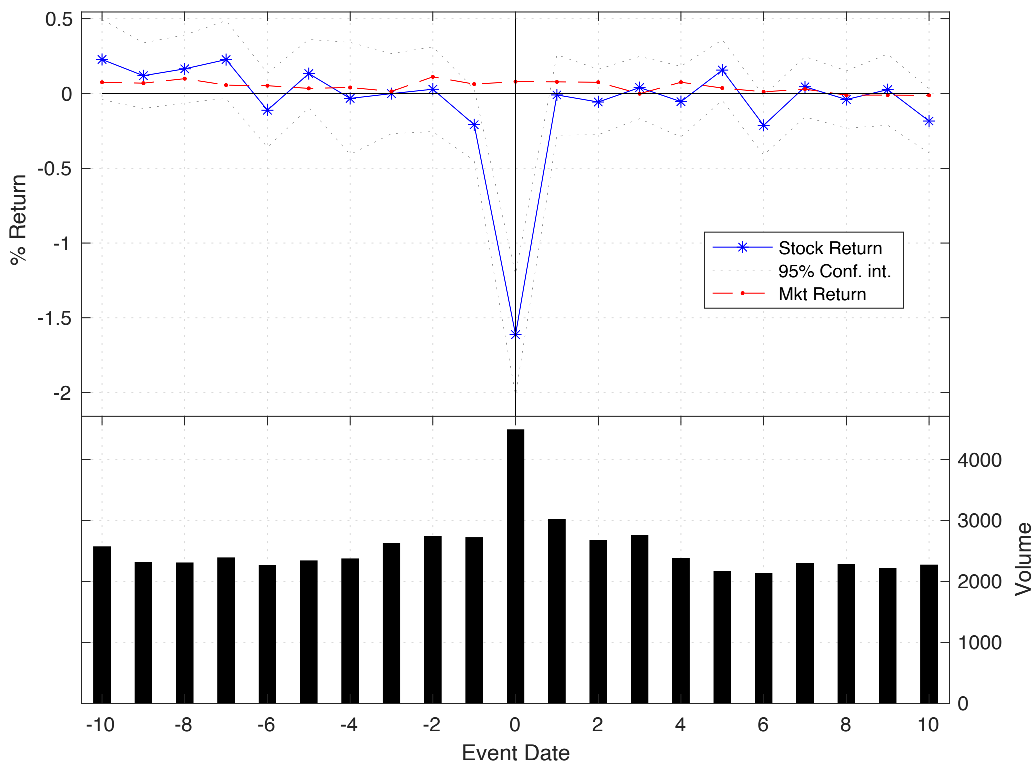


Figure 1.4: Changes in Beta - Market Sample

The figure plots the coefficients reported in Table 1.3, which are estimated through Equation 1.4 estimated with 530 events from January 1, 2010 to December 31, 2013. The blue line represents the change in realised beta from its long-run average calculated according to the S&P 500, while the dashed lines represents the 95% confidence bounds. The *x-axis* of the figure reports the days before and after the news day, thus *Event Date* equals 0 on the news date. The *y-axis* measures the percentage change of the beta form the long-run average.

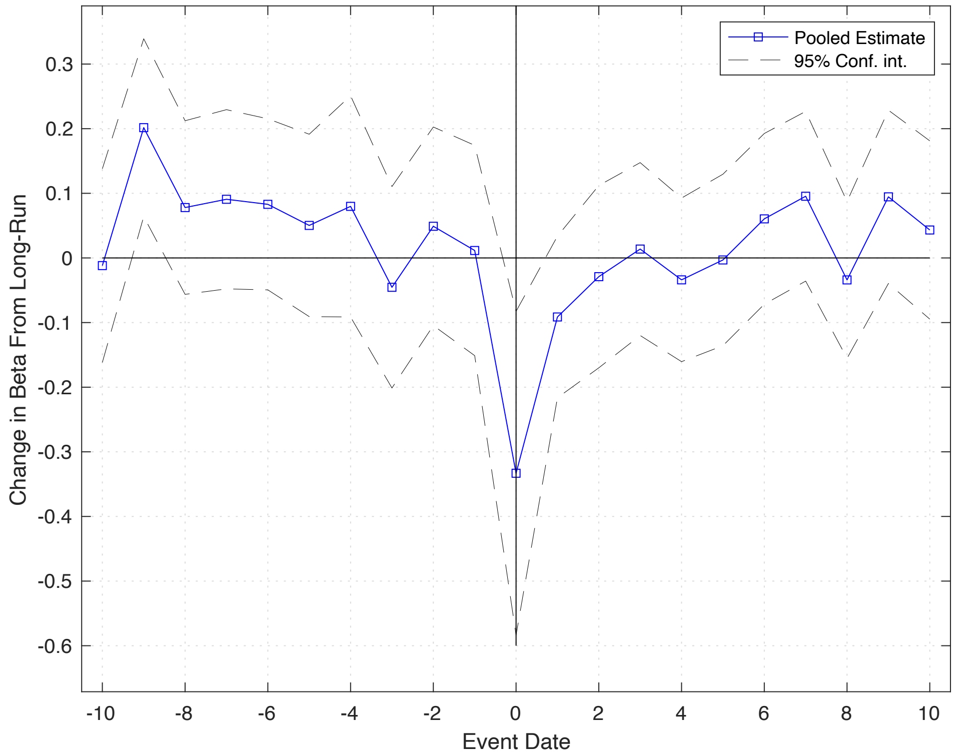


Figure 1.5: Changes in Beta - Sector Sample

The figure plots the coefficients reported in Table 1.4, which are estimated through Equation 1.4 estimated with 530 events from January 1, 2010 to December 31, 2013. The blue line represents the change in realised beta from its long-run average calculated according to the respective ETF sector's portfolios (see Appendix A.3), while the dashed lines represent the 95% confidence bounds. The *x-axis* of the figure reports the days before and after the news day, thus *Event Date* equals 0 on the news date. The *y-axis* measures the percentage change of the beta form the long-run average.

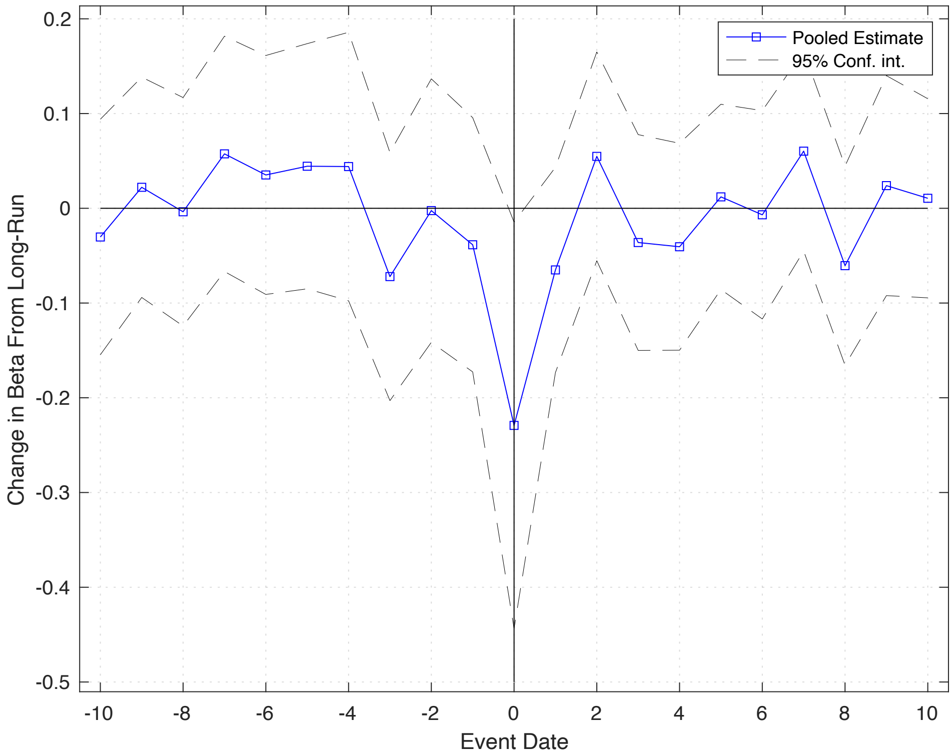


Figure 1.6: Sentiment Distribution

The figure plots the cross-sectional average sentiment across all companies over twenty-one days around 530 firm-level news writing about SEO from January 1, 2010 to December 31, 2013. In order to be part of the sample, a news must have at least a company relevance score of 80%. In the case of a day without identifiable news, the last available sentiment indicator is carried forward. The blue line shows the sentiment of the news on each day, while the dashed lines represent the 95% confidence bounds. The *x-axis* of the figure reports the days before and after the news day, thus *Event Date* equals 0 on the news date. The *y-axis* measures the sentiment.

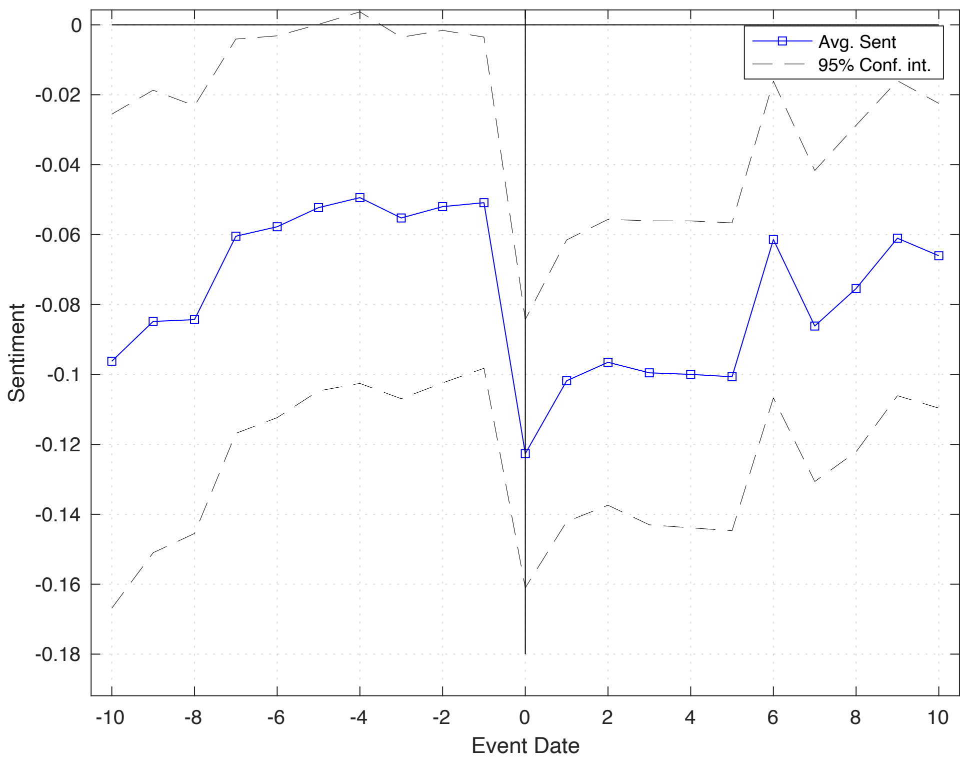


Figure 1.7: Distribution of Board Member Death

The figure plots the distribution of board member death announcements from January 1, 2006 to December 31, 2014 at a monthly frequency. I collect the death dates from BoardEx, and require that a board member - at the time of the death - had only one mandate. Furthermore, to be part of the sample a company must not lie on the bottom decile of the distribution, when sorted according to the Amihud illiquidity ratio (Amihud, 2002), have a price greater than five dollars and not be classified in the Utilities or Banking sector. I identify a total of 305 announcements made by public listed companies on the NYSE and AMEX and NASDAQ. The *x-axis* of the figure shows the year, and *y-axis* measures the number of news in each month.

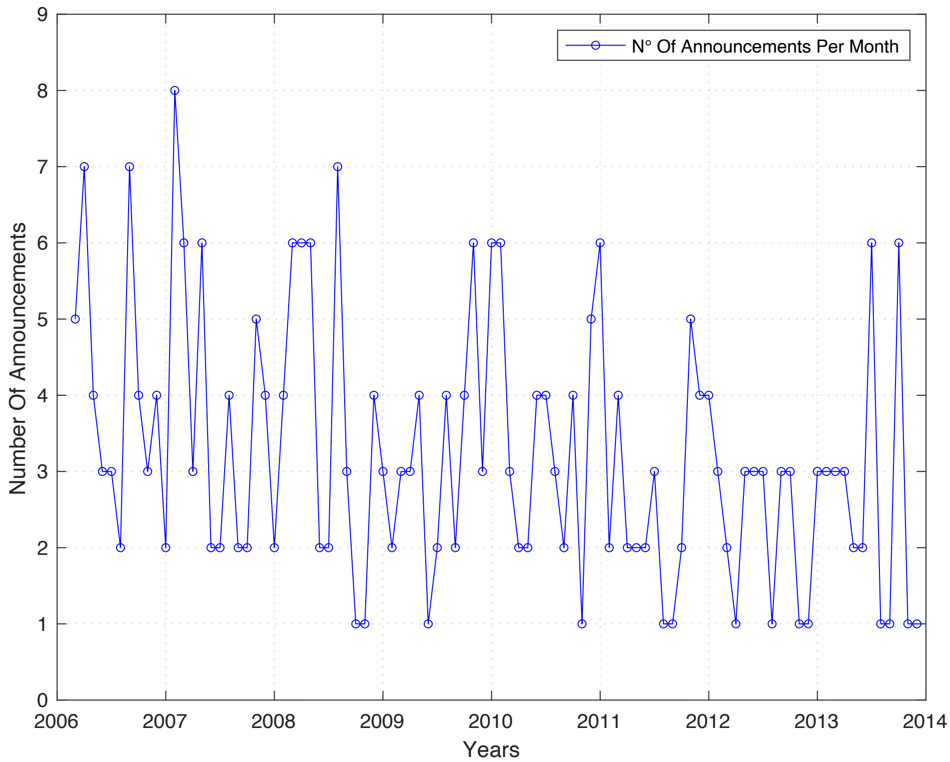


Figure 1.8: Changes in Beta - Board Member Death

The figure plots the coefficients reported in Table 1.9, which are estimated through Equation 1.4 estimated with 305 events from January ,1 2006 to December 31, 2014. The blue line represents the change in realised beta from its long-run average calculated according to the S&P 500, while the dashed lines represent the 95% confidence bounds. The *x-axis* of the figure reports the days before and after the news day, thus *Event Date* equals 0 on the news date. The *y-axis* measures the percentage change of the beta from the long-run average.

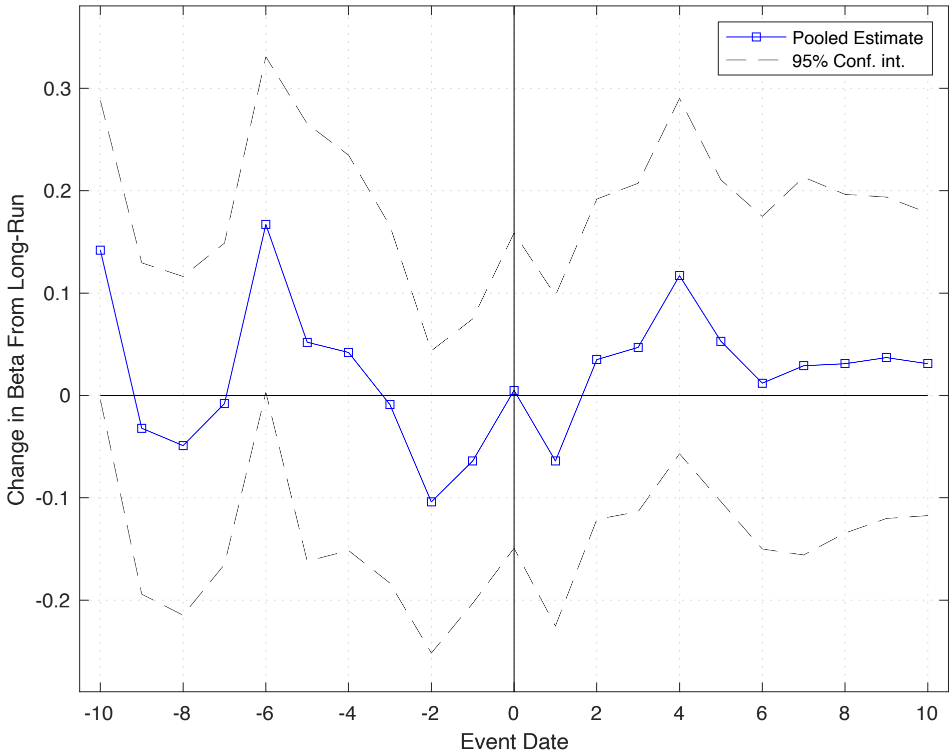


Figure 1.9: Distribution of Random Dates

The figure plots the distribution of random dates from January 1, 2006 to December 31, 2014 in a monthly frequency for a total of 378 events. These event dates and companies are randomly selected. The randomly selected companies are selected from the sample used in the main analysis of the paper and must not have an identifiable event over twenty-one days to be part of the sample. The *x-axis* of the figure reports the years, while the *y-axis* measures the number of events.

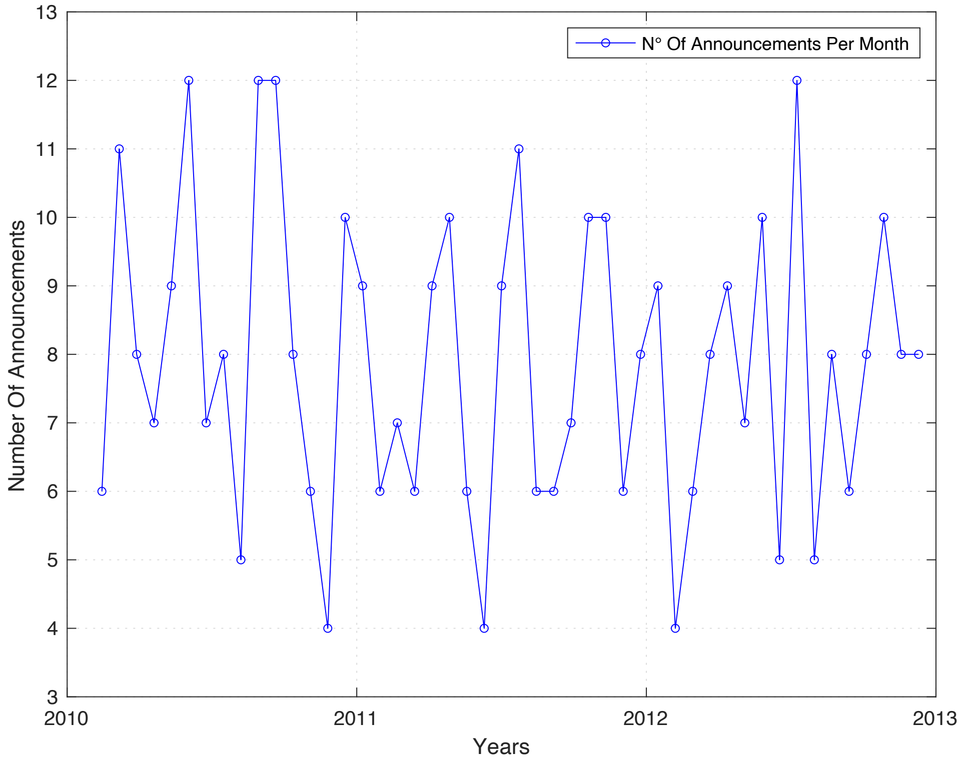


Figure 1.10: Changes in Beta - Random Dates

The figure plots the coefficients reported in Table 1.10, which are estimated through Equation 1.4 estimated with 378 events from January 1, 2010 to December 31, 2013. The blue line represents the change in the realised beta from its long-run average calculated according to the S&P 500, while the dashed lines represent the 95% confidence bounds. The *x-axis* of the figure reports the days before and after the news day, thus *Event Date* equals 0 on the news date. The *y-axis* measures the percentage change of the beta from the long-run average.

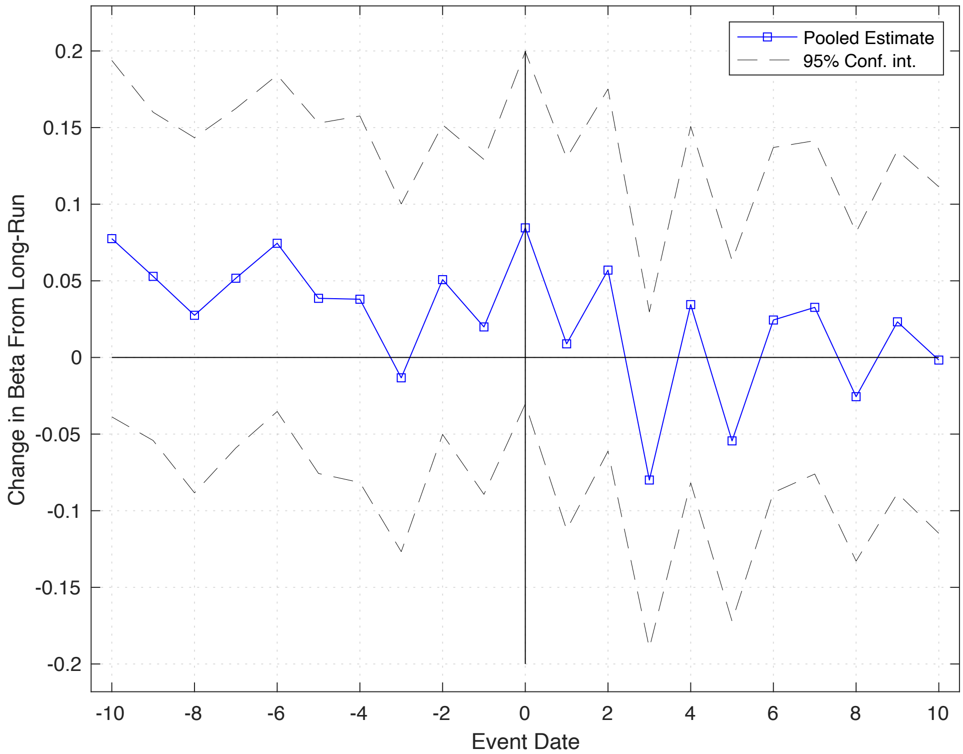
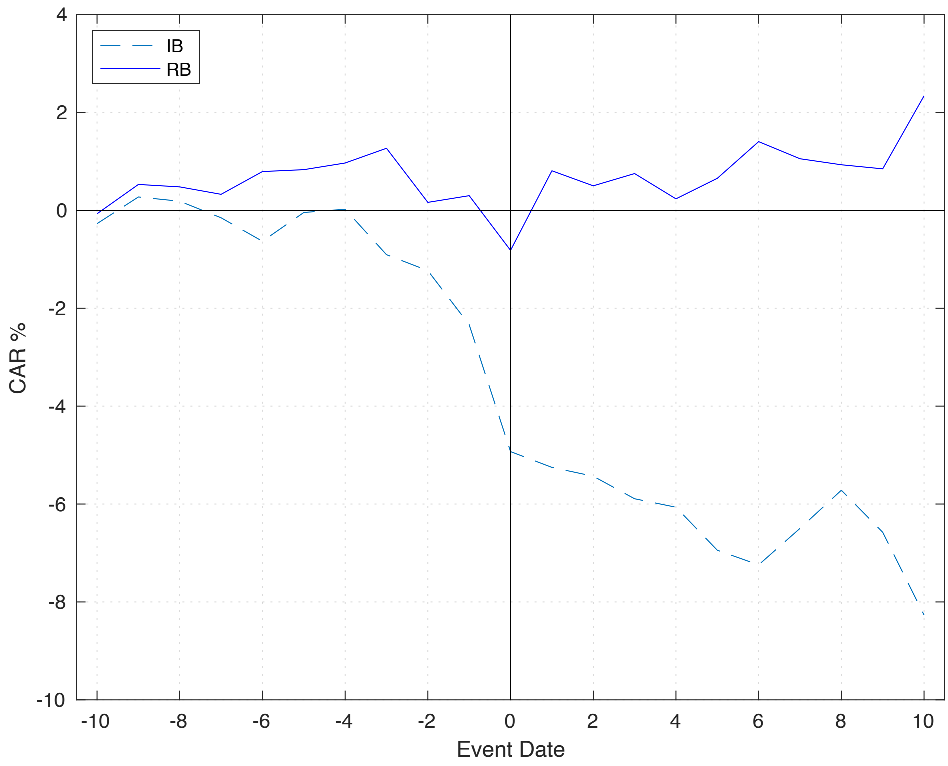


Figure 1.11: Cumulative Abnormal Return - Realised Beta vs. Integrated Beta

The graph plots the cumulative abnormal return (CAR) around 530 SEO news from January 1, 2010 to December 31, 2013. The green dashed line shows the CAR when the abnormal return is calculated with Equation 1.6 (integrated beta), while the blue line shows the CAR when the abnormal return is calculated according to Equation 1.8 (realised beta).



Chapter 2

Dissecting Firm-Level News: a New Measure to Capture the Time-Varying Risk of the Company

2.1 Introduction

Since the publication of Roll (1988) there has been a considerable amount of work linking public information to asset prices. This strand of the literature studies whether more information equates to more or less uncertainty (Morck et al., 2000; Teoh et al., 2009). To analyse this effect, researchers often compare the R-squared of the market model between days with and without public information. However, by identifying only relevant firm-level news, Boudoukh et al. (2019) argue that neither the presence, the quantity nor the timing of information resolve uncertainty, but rather the added information about firm-specific fundamentals. Through their approach, they demonstrate the portion of the return volatility that is influenced by news. In this paper, I extend their work by creating a novel methodology to derive a proxy variable from the linguistic description of firm-level news that does not require the selection of relevant news. I call this proxy, the Market Similarity (MS). I show that the MS explains the changes in the company's systematic risk, stock return volatility, and overall market risk. Finally, given the characteristic of the MS, I argue that this measure captures the degree of company uncertainty.

The MS quantifies the amount of idiosyncratic (company-specific) and systematic (market-wide) component in each firm-level news.[1] For instance, imagine that an investor reads a firm-level news, but the only information he/she obtains is about the market. Here, the news does not report any additional insights about the company but, instead,

[1]Hereafter, I will use firm-level news, news item or more simply news interchangeably.

describes only information related to the state of the market. The lack of company-specific information hinders the investor's ability to analyse the company accurately (Teoh et al., 2009). The uncertainty brought by the news is manifested through an increase in stock return volatility (West, 1988) and systematic risk (Coles et al., 1995). Furthermore, when a lot of companies capitalise the market-wide component in their stock returns (Roll, 1988), the diversification of the market portfolio decreases, leading to an increase of the market volatility. In this specific example, the news will have a counter-effect: it will increase the company uncertainty, instead of decreasing it. Similar to Zhang (2006), I define uncertainty as ambiguity with respect to the implications of new information for a firm's value.

Specifically, I investigate three predictions related to the market-wide component. First, when a considerable number of firms have news where the market-wide component dominates the company-specific one, the market risk increases. Second, the larger the market-wide component, the less information investors obtain about the company, and the higher the stock return volatility (West, 1988). Third, the larger the market-wide component, the more the company is related to the market and the higher its systematic risk. (Coles et al., 1995). Indeed, all these predictions can be reconciled with uncertainty, where the larger the market-wide component is, the higher the market and company uncertainty.

To quantify the market-wide component, I propose a new methodology where the narrative of each news item is compared to a hypothetical market news-index[2], in order to measure their text similarity. Intuitively, the part that is not similar to the news-index must be the company-specific one. If this methodology delivers the desired result, then the MS will enable investors to monitor their risk exposure in real time.

In order to create a methodology that approximates the composition of news, four conditions should be satisfied: (i) given the high sparsity and dimensionality of the written information, the methodology should distinguish "irrelevant" from "relevant" words, (ii) it has to infer precisely and consistently all information from the news; i.e. two firm-level news describing the same matter must be represented alike, across companies and over time, (iii) the extent to which the news reports the market-wide component has to be comparable across all news. Thus, the proxy variable has to vary between a defined interval and must be the same for two news describing different matters, but with similar proportions of market-wide information, and (iv) the methodology should interpret the

[2]The market news-index incorporates all news in the market in each period.

content of firm-level news as a human would and deliver interpretable results without the need of prior knowledge about the news (i.e., dictionaries). This last condition allows the methodology to be qualitatively evaluated, overcomes the "black-box" issue of machine learning models, is adaptive to change in the composition of the news and is applicable to various written information.

To meet condition (i), I propose a step-by-step procedure that satisfies all the text pre-processing requirements, and which is presented in Appendix B.1.[3] To meet condition (ii), I use Natural Language Processing (NLP) methodologies, which is a branch of Artificial Intelligence (AI). In more details, I use the Latent Dirichlet Allocation (LDA) (Blei et al., 2003) to infer 90 unlabelled topics composed by ten words each, which I interpret as the key information of the news. I then retain the two most discussed topics in the weekly collection of news for each company. Furthermore, I use a neural-network model; namely, Skip Gram (SG) (Mikolov et al., 2013a,b) to infer the meaning of each topic. To meet condition (iii), I create a news-index that groups all topics across all news in each week, which represents the market information. I then calculate the Cosine Similarity between the news and the news-index topics, which I will refer to as the MS. To check whether condition (iv) is satisfied, I propose several tests that are presented in Section 2.5.1.

The analysis accounts for all firms listed on the S&P 500 from January 1, 1990 to 31, December 2017, resulting in 1,071 unique companies. To be included in the sample, a company has to be listed for a minimum of thirty calendar days and have at least one news during the listing period. Following Hillert et al. (2014), I obtain the firm-level news from the LexisNexis and ProQuest archive, and focus only on news reported in the financial section. In total, I identify 458,636 firm-level news.

To test the first prediction, I calculate the cross-sectional mean of the MS in each week across all companies, and call this measure Average Market Similarity (AMS). Similarly to Manela and Moreira (2017), I analyse whether the AMS explains the variation in the Market Volatility Index (VIX). The VIX is calculated from the options' implied volatility of the S&P 500 constituent companies and is a recognised measure of market uncertainty (Whaley, 2009). At weekly frequency, I show that a 1% increase in the AMS leads to a 1.45% increase in the VIX. Furthermore, at monthly frequency, a 1% increase in the AMS leads to a 1.9% increase in the VIX. The same model specification shows that a 1% increase

[3]I propose a new stemmer, and lemmatiser. Code available upon request.

in the NVIX of Manela and Moreira (2017) is associated to a 0.94% increase in the VIX. These results indicate that the AMS is a good measure for market uncertainty, and it is also superior to the NVIX.

To test the second prediction, in each week, I sort companies in ascending order according to their MS, and create five equally-weighted quintiles - where quintile 1 has the lowest MS, while quintile 5 has the highest MS - to study the incremental effect of the MS. Teoh et al. (2009) show that investors strugle to analyse companies that incorporate less information about their fundamentals. The authors argue that those firms have high uncertainty about their fundamentals, which is reflected in a low R-squared. West (1988) shows that more information about future dividends leads to lower stock return volatility. If the MS captures the company uncertainty, and those companies have less information about future dividend, then Q1 should have a lower stock return volatility and higher R-squared compared to the other quantiles. I show that the R-squared of the market model monotonically decreases from Q1 to Q5, while the stock return volatility monotonically increases. Furthermore, through a Pooled OLS regression, I document that an increase by one standard deviation in the MS corresponds to an increase of 0.72% in the stock return volatility. These results indicate that investors' inability to correctly evaluate companies with high MS leads to higher return volatility. Hence, the higher the market-wide component, the higher the company uncertainty.

To test the third prediction, I use the coefficients of the market models (beta) estimated on each quintile and several other tests. Coles et al. (1995) theoretically demonstrate that high (low) informative stocks have lower (higher) beta. My results confirm their theoretical findings. Q1 has a beta of 0.967, while Q5 has a beta of 1.191. In an additional test, where I regress the realised beta on the MS, I show that a 1% increase in MS leads to a 0.148% increase in realised beta. Furthermore, through a Pooled OLS regression, I check the relationships between the market, the company returns and the MS. I regress the excess returns of the company on the excess returns of the market, the company MS and the interaction between the returns of the market and the company MS. The coefficient of the interaction term is 0.95, while the market coefficient is 0.49, both statistically significant at the 1% level. These results imply that, when uncertainty about the company is high, the market returns become more relevant. Altogether, the higher the MS the more the company is exposed to market variations, increasing its systematic risk.

This study contributes to the finance literature using advanced text analysis methodologies. Hanley and Hoberg (2018) argue that the greater the number of financial companies that discuss risk-related topics in their 10-k, the higher the systemic risk of the sector. Engle et al. (2020) show that by measuring the similarity between firm-level news and a climate vocabulary index, it is possible to hedge climate risk. The major difference of the methodology proposed in this paper and previous studies is that, I do not create any particular list of words. Thus, by limiting researcher interactions with the data, this methodology allows the written information to be more representative of the document. Furthermore, given the flexibility of the proposed methodology, it can be applied across a wide range of qualitative written information.

The evidence presented here sheds new light on the role of firm-level news on the uncertainty and risk of the company. After Roll (1988), who asserts that low R-squared seems "to imply the existence of either private information or else frenzy unrelated to concrete information", two different views have emerged about the meaning of low R-squared. One view argues that low R-squared implies more firm-specific information into stock prices (Morck et al., 2000). Instead, the alternative view asserts that companies with low R-squared are characterised by high uncertainty since they incorporate less information about their fundamentals (Teoh et al., 2009; Kelly, 2014). The results reported in this paper support the latter view. To the best of my knowledge, this paper is the first to analyse the relationship between the R-squared and a measure derived from the qualitative information of the firm-level news. This is an important step to address the debate between the two views. Moreover, Boudoukh et al. (2019) show that not all news is equally important, and demonstrate that prices move only in the presence of relevant news. I extend their work by showing that it is also about the composition of the news and not necessarily the news, per se.

This paper also provides new insights about the impact of firm-level news on the systematic risk. Patton and Verardo (2012) and Savor and Wilson (2016) show that around earnings announcements the systematic risk of the announcing company increases. Using firm-level news writing about secondary equity offering programs, Giannone (2021) demonstrates that the systematic risk of the company decreases. While the above literature uses news in an event study setting, this paper uses the narrative of news to measure the change in the systematic risk. Hence, it offers further evidence about the importance of firm-level news on the company's systematic risk. Altogether, this study provides a new approach to measure the relationship between the company and the market returns, which can be used by investors to assess the company and the market uncertainty. Furthermore, it empirically supports the prediction of Coles et al. (1995) about the relationship between uncertainty and the systematic risk.

This paper is organised as follows: Section 2.2 presents the relevant literature and the hypothesis tested in this study. Section 2.3 describes the methodology to derive the MS measure. Section 2.4 explains with the sample construction and reports some descriptive evidence for the firm-level new. Section 2.5 evaluates the methodology, and presents the results about the relationship between the Market Similarity and the market uncertainty. Section 2.6 shows that the Market Similarity proxies the change in the company's uncertainty, measured through stock volatility return, systematic risk as well as the R-squared of the market model. Finally, Section 2.7 provides the conclusions and remarks of the paper.

2.2 Current Literature and Hypothesis Development

This paper is related to the literature that investigates the implications of company's information on the stock return volatility and systematic risk and, by extension, on company uncertainty. I investigate the extent to which firm-level news provide information about either the firm or the market is reflected in the company's return volatility and systematic risk. Toward this end, I combine three different strands of the literature. The first examines the links between companies' information on their level of uncertainty. The second explores the use of text analysis to measure the company's stock return volatility; and, finally, studies that analyse the effect company information on the systematic risk.

The interplay between news and quantitative financial variables is a well established phenomenon. The pioneer work of Tetlock et al. (2008) argues that if the information in a company's filing is incomplete, investors rely on the linguistic description of the firm's profit-generating activity to evaluate the company. Following this publication, scholars started exploring different implications of words on stock risks and returns, by selecting specific terms in the various corpus (Loughran and McDonald, 2016). This approach can be referred to as the dictionary-based analysis, which assumes that the only relevant words are those in the specific dictionary and the context of the words is not important.

News items are also used to build event studies. Early works around earnings announcements (Savor and Wilson, 2016; Patton and Verardo, 2012) and secondary equity offerings (Giannone, 2021) show that, around the release of firm-level news, the systematic risk of the company deviates from its long-run average. Overall, this implies that, each firm-level news is composed of an idiosyncratic (company-specific) and systematic (market-wide) component. Following this, one could measure the extent to which each piece of news is related to the market to capture the change in the riskiness of the company. Assuming that firm-level news writes about a fixed number of topics, then the less (more) the news describes similar topics such as the market, the more (less) investors obtain accurate information about the firm, and the lower (higher) the risk of the firm.

To implement the idea above, I apply Natural Language Processing (NLP) methodologies (Jurafsky and Martin, 2008). One of the most used NLP methodologies is the Latent Dirichlet Allocation (LDA), which is a topic model. LDA returns a list of n words that represent an unlabelled topic.[4] Antweiler and Frank (2005) assume that each news item represents a company's event, and demonstrate that firm-level news topics explain jumps in stock prices and increases in trading volume. Bybee et al. (2020) show that, through topics inferred from all news of the Wall Street Journal (WSJ), it is possible to track economic activities and forecast macroeconomic variables. Finally, by identifying fourteen types of news through a third party text classification, and through a rule-based information extraction platform (e.g., The Stock Sonar) using an alternative approach, Boudoukh et al. (2019) classify news as unidentified and identified. They show that identified news is associated with high levels of stock return volatility when compared to unidentified news.

By using another NLP methodology, text-similarity, Hanley and Hoberg (2018) show that the more companies in the banking sector discuss risk themes, the higher the systemic risk of the sector. Toward this end, they select only the sections in the 10-k where the word "risk" appear, and summarise the sections through LDA. For the words identified by LDA, they create lists of related themes (i.e., vocabulary) through a Semantic Vector Analysis (SVA).[5] Finally, to measure the exposure of the bank to a specific theme, they calculate the Cosine Similarity between the vocabulary list associated with each SVA risk theme, and the raw text of each bank. Furthermore, Engle et al. (2020) show that by measuring

[4]For details about the LDA, I refer the reader to Section 2.3.2.

[5]SVA can be considered a similar model as the SG used in this paper and described in Section 2.3.3.

the similarity between firm-level news and a climate vocabulary index, it is possible to hedge against climate risk.

Following this literature, I summarise news in topics through LDA and interpret these as key information in the news. To understand how related the news is to the market, I create a news-index, which is composed of all topics inferred from the news. I interpret the news-index as the information about the market. To measure the extent to which firm-level news report information about the market, I infer the similarity between the topics of the news and the news-index. I will refer to this measure as the Market Similarity (MS).

There are two important implications of high level of MS for the risk of the company. First, high MS will increase the stock return volatility of the company: (i) an increase in the exposure of the company to the market will increase the total volatility of the company through the market volatility; and (ii) when firm-level news reveals less information about the company's idiosyncrasies, the investors' ability to rely on the linguistic description of the firm's profit-generating activity to evaluate the company is hindered, which increases the uncertainty about the company (Jiang et al., 2009). Second, high MS will increase the systematic risk of the company. If the firm-level news reports the same news as the overall market, the return comovement between the market and the company increases, which drives up the systematic risk. Stated differently, given the low information about the company, investors will gauge the company mostly through market information, increasing its exposure to market variations. The implications of the MS on stock return volatility and on the systematic risk can both be reconciled with the concept of uncertainty. Similarly to Zhang (2006), I define uncertainty as ambiguity with respect to the implications of new information for the firm's value.

H_1: The higher the similarity between firm-level news and the news-index, the higher the uncertainty about the company

This hypothesis implies that, the lower (higher) the MS, the more (less) investors gather information about the company, and the lower (higher) the uncertainty about the firm. I measure the effect of the MS according to the change in stock return volatility and the systematic risk. However, as an additional test to understand whether the MS meaningfully captures any degree of uncertainty regarding the company, I rely on what is asserted by Roll (1988) where, in his pioneering research he argues that low R-squared seems "to imply the existence of either private information or else occasional frenzy unrelated to concrete information" (page 566). Hence, if the MS quantifies the amount of information that is either related to the firm or to the market, then low (high) MS should be associated with high (low) R-squared.

Similarly, Teoh et al. (2009) show that the R-squared is a positive proxy for the quality of the information environment. That is, low R-squared firms have low information quality and high volatile earnings. They assert that low R-squared firms are characterised by high uncertainty about their future fundamentals, which, in turn, are more difficult to analyse by investors. West (1988) demonstrates through a theoretical model, that a low R-squared is related to less firm-level information and more noise returns. Kelly (2014) shows that firms with low R-squared have a lower quality of information environment since they have fewer institutional holdings and lower analyst coverage.

The above literature investigates the relationship between the R-squared and uncertainty through a cross-sectional analysis. In this paper, I focus on the time-series R-squared, since I am mostly interested in the effect of the MS over time rather than across companies.[6]

Moreover, previous literature also linked the time-varying systematic risk of the company to information disclosure. Xing and Yan (2019) measure the importance of accounting information quality on the systematic risk by estimating the Capital Market Pricing Model (CAPM) with daily observations. They demonstrate that the higher the information quality (i.e., low level of total accruals) the lower the systematic risk. Coles et al. (1995) theoretically show that high "information security" tends to have lower systematic risk, since it has a lower parameter uncertainty. Through a theoretical approach, Lambert et al. (2007) show that an increase in accounting information quality reduces the comovement of the firm's cash-flow with other firms in the economy, thereby reducing the company's cost of capital. Altogether, these studies support the idea that firms with high MS will also have a high systematic risk.

[6] Teoh et al. (2009) asserted that one can study and the uncertainty through the R-squared in a time-series setting is reasonable.

2.3 Methodology

Over the last decade, scholars have mostly focused on searching words according to predetermined lists in various texts to retrieve salient information about companies. The most significant drawback of such a methodology is the sensitivity of the results to the selected words' lists, since it assumes that words occur independently and that the only ones that matter are those in the list (Loughran and McDonald, 2016).

Even though the dictionary-based analysis produced relevant results, the information that can be inferred from text is bound by the type of dictionary. To expand the purpose of text analysis in finance, human interaction should be limited. Indeed, NLP methodologies provide unbounded possibilities to extract information from text with limited discretionary choices.[7]

For each company, I group firm-level news in each week across all sources from Wednesday to Wednesday.[8] This choice mitigates the effect of investor over-reactions documented over shorter periods (Tetlock, 2011), improves the pricing ability of the news (Heston and Sinha, 2017) and ensures that all printed information is reflected into stock prices (Dang et al., 2015).

In this section, I present how the news is elaborated and how the MS is constructed, while in Section 2.5.1 I show the ability of topics and topic similarity to measure the change in the company and market information environment. The remainder of this section is organised as follows: Section 2.3.1 delineates how the text is prepared for the entire analysis. Sections 2.3.2 and 2.3.3 show how topics and their meanings are inferred. Section 2.3.4 presents the methodology to construct the MS.

2.3.1 News Pre-Processing Step

Both LDA and SG rely on the distribution of words in a document. To improve the statistical accuracy of these two models, words (i.e., uni-grams) that lie on the two tails of

[7] It is worth highlighting that choice of the model and its hyperparameters used to infer the information from the document has implication on the results. However, by a less extent that a dictionary-based analysis.

[8] Unless specified, I will refer to firm-level news or simply news as the weekly collection of news for each company.

the distribution are eliminated.[9] On the right tail, there are words that have no or little lexical meaning (i.e., stop words) and are common across news, while on the left tail there are words that are unique to each company (i.e., products or companies' name). I then infer the roots of the words (i.e., lemmatisation), to create a comparable lexicon among words (i.e., making vs. made). This last step enables a unique representation of a word independently from its declination.

Furthermore, I group all news for each unique company and eliminate words that appear less than twenty-five times and in more than 40% of the groups. Hence, I retain 16,241 unique words, which is approximately the same number of words identified by Bybee et al. (2020). To account for the co-occurrence of words and their joint meaning (i.e., chief and executive), I infer the most frequent bi-grams through a maximum likelihood estimation, for a total of 3,247 bi-grams. The outcome of this pre-processing step is a set of words that collectively represent the content of the news, uniform the writing styles and reduce the sparsity of the news.

2.3.2 Determine Topics - Latent Dirichlet Allocation

I use LDA to reduce the dimensionality of the news into 90 topics composed of ten *Key Terms* each. Decomposing the raw text into topics magnifies the analysis of news and, at the same time, makes the entire study computational scalable. Thus, through the LDA, it is possible to summarise a thousands of words in each week into a few unlabelled topics. Collectively, these ten *Key Terms* give meaning to the topic. For example: *debt, bankruptcy, creditor, loan, pay, plan, file, file bankruptcy, finance, protection* refer to a company filing for bankruptcy. I interpret topics as the key information in the news.

The LDA assumes that each document is built as a distribution of hidden topics, and each topic is built as a distribution of words. A practical intuition is as follows. Imagine that a writer needs to describe a bankruptcy event. To do so, the person will draw upon some key terms from a distribution of words that are needed to describe such an event (e.g., bankruptcy, debt), which will be distributed across news writing about bankruptcy events. These terms will re-occur over the entire sample whenever the writer describes this type of event. The LDA identifies whenever the writer uses words from the same distribution of

[9] In this section, I provide only a high-level overview of the process implemented to get the study sample, for more details I refer the reader to Appendix B.1.

words (i.e. bankruptcy), and clusters them together. The result of such a clustering step is the ten words that build the unlabelled topic. Stated differently, to get the final list of ten *Key Terms*, the model starts by randomly allocating each unique term across the n number of topics specified by the researcher, and through the Gibbs-Sampler algorithm, it allocates terms that co-occur more often together. This notion is also reflected in the name of the model. The term Latent underlines the idea that topics are the hidden component of the text, which is modelled through the observed distribution of words. The second part of the name refers to the type of probability distribution employed. The Dirichlet component allows this model to be flexible regarding the number of topics to be identified. The Allocation part of the name indicates that the topics are then allocated among the documents under a probabilistic framework.

In more detail, the LDA is a probabilistic version of the Bag-Of-Words methodology.[10,11] This model is part of the unsupervised machine learning technique, where the user does not need to specify the response variable since the algorithm will aim to learn it. Even though the LDA is an unsupervised model, the researcher needs to specify the number of topics to be estimated as well as two hyperparameters. To determine the optimal number of topics, I rely on a coherence measure, which is estimated over the entire sample for a various number of topics (Mimno et al., 2011). Figure 2.1 shows that the U-Mass - which is a coherence measure - spikes at 80 and 120 number of topics, indicating that the best outcome of the model is observed with these two number of topics. Since this measure is entirely data-driven, I use this result to optimise the two hyperparameters α and β as well as the number of topics by qualitatively judging the results. The two hyperparameters govern the Dirichlet priors, the closer these are to zero, the fewer topics are identified in the documents (α) and the lower/minimum number of words that build a topic is (β). The best outcome is reached at 90 topics with α equal 1 and β equal 0.01.[12]

In Table 2.1, I report the full list of topics inferred over the entire dataset. Consistently withe extent research using the LDA, I give a *Topic Label* to the *Key Terms* for presentation reasons (Larsen and Thorsrud, 2017; Hanley and Hoberg, 2018; Bybee et al., 2020). The *Key Terms* are the output of the LDA and are completely data-driven, whereas the *Topic*

[10]Another intuition is to see the LDA as a Principal Component Analysis (PCA) for discrete variables.

[11]In this section, I provide the intuition of the model through an example. For a more formal explanation I refer the reader to the original papers of Blei et al. (2003).

[12]The number of topics is in line with Larsen and Thorsrud (2017), who identify 80 topics, and Bybee et al. (2020) who find the optima parameter as α equal 1 and β equal 1.

Label represents what I believe is a natural label for the given topic. Furthermore, I separate the topics into *Groups*, to identify the topics that refer to similar facts. However, *Topic Label* and *Groups* do not play any role in this study, but it is a convenient way to refer to the *Key Terms*.

An important feature of LDA is that, once the topics are learned, the researcher can determine the extent to which that each of these topics is discussed in the given news. That is, each news item is then represented by a topic's probability distribution. Bybee et al. (2020) refer to this probability as an attention measure and indicate that topics that represent the document with more than 30% probability are "pure" topics. For this reason, I retain only the two most probable topics. On average, the two most discussed topics account for more than 50% of the total topics' probability distribution of each news. Indeed, by retaining a higher number of topics, the news would be represented by topics that are discussed only marginally, misleading the inferred information from the news.

2.3.3 Topic Meaning - Skip-Gram

The output of the LDA describes how words are distributed over the news. However, incorporating such information directly into asset-pricing models can challenging. For example, what is the effect of topic x as opposed to topic y, when topic y is interpreted as good or bad? Event though some topics possess a directional impact (i.e., bankruptcy), most communicate matters that might have different implications over time and/or across companies. One way to overcome this issue is to assess the topics' sentiment (Larsen and Thorsrud, 2017). However, since one of the objectives of this paper is to use the least discretionary choices, I do not rely on this solution but, I propose a new approach.

To determine the relationship between topics, I use the SG neural-network technique, where words are represented through a network in an n dimensional vector (Mikolov et al., 2013a,b). The goal is to understand the meaning of a word by observing the distribution of the surrounding ones. Two words with a similar meaning are used in similar contexts, whereas words with different meanings are used in different contexts.[13] The statement of Firth (1957) summarises the overall idea of the SG: "*A word is characterised by the company it keeps*".

[13]In this section I present only a high-level overview of the model, for a more technical description, I refer the reader to the original papers of Mikolov et al. (2013a) and Mikolov et al. (2013b).

Also in this second model, the researcher needs to determine two parameters: window-size and the vector dimension. I set the vectors dimension to 100, considering precision and computational efficiency (Mikolov et al., 2013b). Furthermore, I set the window size to two. The window-size determines the number of neighbours words to account for when the vector representing each word is calculated. The smaller the window-size, the fewer the number of words used to optimised the vectors and the higher the similarity with neighbouring words. Conversely, the higher the window's size, the more generalised the concept of each word.[14]

The output of this model is a unique vector for each word that represents its relationship to all other words. Two words that often co-occur or are used in a similar context will have a similar vector, whereas words that never appear together, or appear in a different context, will have two independent vectors. In order to determine the meaning of the topics, I combine the vectors of the *Key Terms* of each topic in Table 2.1.

To train the SG, I use the same sample as the LDA, with a slight modification. That is, I train the SG on sentences, rather than on the entire news. This avoids the calculation of vectors between adjacent words in separate sentences. Moreover, to preserve the semantic of the words, I eliminate stop-words and replace uni-grams with respective bi-grams.

Through the vectors' representation of the topics, I calculate the similarity between topics through the Cosine Similarity:

$$S_{i,j} = \frac{\mathbf{i} \cdot \mathbf{j}}{||\mathbf{i}|| \, ||\mathbf{j}||}, \tag{2.1}$$

where i and j are the vectors representing the news topics of two companies, and $S_{i,j} \in [0, 1]$. That is, if two news items have the same (different) topics, the Cosine-Similarity will be one (zero), respectively.[15]

To understand whether the topics' similarity captures the semantic relationship between topics, Figure 2.2 reports a heat-map for each of the 90 *Topic Labels.* For presentation reasons, the figure displays only the *Group* of Table 2.1. Each of the squares in the figure

[14]For example, the word in the fifth position in the sentence is optimised against the previous and subsequent two. In the following iteration, the word in the sixth position is optimised against the previous two and subsequent two, and so on.

[15]A similarity of zero is returned only if two words -or topics- are orthogonal, which is unrealistic. Even though one would assume that two words could be orthogonal, considering that each similarity is calculated between two topics and each topic is composed of ten words, a similarity of zero will require that each single pair of words composing the topics have a similarity of zero.

depicts the similarity between two topics. The brighter (darker) the square, the more (less) similar the two topics are. If the *Topic Labels* are classified correctly in each *Group* and the SG infers the semantic similarity of the *Key Terms*, then the squares around the diagonal between *Group*'s names should be brighter. Indeed, this is the case. For example, around the diagonal of the *M&A* group the squares are brighter than anywhere else, indicating that all these topics are semantically related, while the same *M&A* group shows little similarity with groups such as *Food & Drinks* and *Air Services*.

In conclusion, the combination of the two models deliver, on the one side, human interpretable results while, on the other side, classify news as a human would possibly do. All of this is achieved with limited number of discretionary choices. Furthermore, this approach overcomes one of the limitations of LDA. Even though the LDA will return slightly different *Key Terms* for each topic when the model is estimated again, this will only have a marginal effect on the vector representation of the topic, leaving the similarity virtually the same. Stated differently, since the topic is represented by a vector composed of ten words, if the LDA returns slightly different words' representation for the topic, the vector representing the topic will mutate only marginally, leaving the similarity between topics nearly unchanged. This last consideration is of critical importance for replicating this methodology.

2.3.4 Market Similarity

The MS measures the similarity between the firm-level topics with the news-index. The more (less) similar the topics of the news with the news-index are, the higher (lower) the MS would be. To avoid a positive bias in the MS, when computing the MS for a company, I excluded the topics of that company from the news-index.

The calculation of the MS for each company requires the estimation of the similarity between circa 210 topics that build the news-index and the two topics of the company.[16] This procedure must be repeated as many times as the number of companies with a piece of news by the number of weeks in the sample, which is extremely computationally demanding. Instead, one can calculate the similarity between company i and all other companies that are part of the news-index, and then take the cross-sectional average of the similarities:

[16] On average there are 105 companies with news in each week and for each company I retain two topics.

$$MS_{i,t} = \frac{\sum_{j=1}^{N-1} S_{i,j,t}}{N-1}, \tag{2.2}$$

where $MS_{i,t}$ is the Market Similarity of company i at time t; $S_{i,j,t}$ is the Cosine Similarity between the topics of firm i with respect to company j as in Equation 2.1, and N is the number of companies with a piece of news at time t. Hence, the MS is an equally-weighted similarity between company i and all other companies in the market.

2.4 Sample Construction and Descriptive Statistics

The sample consists of all firms listed on the S&P 500 from January 1, 1990 to December 31, 2017. I select this sample since these companies are easy to track, receive enough coverage by the media and represent approximately three-quarters of the US market capitalisation (Tetlock et al., 2008). The S&P 500 constituents companies, their quarterly number of shares outstanding, the book values of equity are from Compustat, while stock price and volume data are from CRSP and the news is from the LexisNexis and ProQuest archive. I calculate the weekly return from Wednesday to Wednesday to avoid possible calendar anomalies.[17] For the same reason, a company is taken into and dropped from the analysis in the following and previous Wednesday of the original listing dates.

To be included in the analysis, a company has to be listed for a minimum of thirty calendar days and have at least one news over the listing period. However, if the only news is reported in the first week of the listing period, the company is excluded. Tetlock (2011) asserts that this news likely discusses the inclusion of the company in the index, which might distort the results. These criteria reduce the total number of firms that appear at list once in S&P 500 from 1,193 to 1,174 where 1,071 are unique.

I calculate the company book-to-market similarly Asness and Frazzini (2013). The size of the company is defined as the log transformation of the most recent quartet number of shares outstanding multiplied by the current stock price. The weekly turnover is the log transformation of the weekly average trading volume, divided by the current number

[17] From this moment onwards, unless specified, I refer to the weekly return or more simply to return as the Wednesday to Wednesday.

of shares outstanding. I calculate the sentiment of the news, for company i at time t:

$$Sent_{i,t} = \frac{\#Positive - \#Negative}{\#Positive + \#Negative}, \tag{2.3}$$

where $\#Positive$ and $\#Negative$ are the number of positive and negative words in the news, which are taken from the sentiment dictionary of Loughran-McDonald.[18]

Table 2.2 shows the sample statistics. On average, over the entire period there are 105 companies in each week, and each company has 2.85 news items, with an average sentiment of -0.323. The average MS is 0.653 with a standard deviation of 0.018 and positive skewness. The long right tale implies that, if any, the effect of the MS will not be symmetric on return volatility and systematic risk.

Table 2.3 reports the correlation between company level data and news variables (grey shaded). To account for the different number of companies and observations for each company, I calculate the Pooled correlation according to the Fisher Z transformation. The table shows that the MS is positively correlated with turnover and the number of news (the bigger the market-wide component, the higher the news circulation and the more investors trade), while it is negatively correlated with size (the bigger the companies the smaller the market-wide component in the news is) and with sentiment (the bigger the market-wide component the more negative the news is). Altogether, the correlation of the MS is consistent with the idea that this measure captures some sort of risk.

I gather the news from the LexisNexis archive, similarly to Fang and Peress (2009) and I focus on US newspapers with the highest weekday coverage of firm-level news (Hillert et al., 2014). More precisely, the sample considers news released by The New York Times, USA Today, Washington Post, St. Louis Post Dispatch, The Atlanta Journal-Constitution, The Philadelphia Inquirer, Pittsburgh Post Gazette, and Tulsa World. I also search in ProQuest for the WSJ as in Engelberg and Parsons (2011). Hence, the sample encompasses both, national (i.e., USA Today) and local (i.e., New York Times) newspapers. These sources are well known to be the financial news for institutional and retail investors (Tetlock, 2011).

To ensure that each news addresses a given company, I rely on the company relevance score ranging between 80% and 99% for the LexisNexis newspapers. This percentage is

[18] `https://sraf.nd.edu/textual-analysis/resources/`.

based on the frequency of the firm's name, weighted by its location in the news. In some instances, it also recognises the abbreviation of the company name as its primary name (i.e., IBM and International Business Machine). To exclude any divergence between the name of the company and its abbreviation, I also search for its ticker. Considering that tickers are reused, or parallel-used in different exchanges, I pair this identifier with the exchange symbol. All this information is manually tracked through Compustat and the company search function of LexisNexis. Since ProQuest does not offer a relevance score, I select only news with the name and/or ticker of the company in the headline or in the abstract.

Through this sampling methodology, I identify 610,016 firm-level news across the nine sources. I then eliminate news that are not reported in the business section (i.e., 126,282) and news that has less than 100 characters where at least 60% are alphabetic (i.e., 25,098). Hence, the total number of news used in the analysis is of 458,636 or 16,974 news per year across all companies. The final distribution of news is reported in Table 2.4. The two most popular sources are the WSJ and The New York Times that, together, account for approximately 50% of the sample, while the least popular source is The Philadelphia Inquirer that covers 5.43% of the sample. Except for the Pittsburgh Post Gazette (01.1993), The Atlanta Journal-Constitution (01.1991), The Philadelphia Inquirer (01.1994) and the Tulsa World (01.1996) all other sources start from January 1, 1990. Figure 2.3 plots the time-series distribution of each source. Except for the WSJ, all other sources show a decay in coverage over time, driving down the average number of news in the second half of the sample (from 2003 to 2017).

2.5 Decomposing Firm-Level News

In this section, I empirically demonstrate that by aggregating all topics of the news it is possible to measure the relevant market information, while the measure of similarity captures the change in the composition of news over time.

In section 2.5.1, I show the evolution of topics over time as well as an example that demonstrates how the similarity evolves around earnings announcements. Altogether, this provides an intuitive way to understand how the similarity between firm-level news and the news-index works. In section 2.5.2, I document one application of the MS to evaluate and forecast the VIX. This application demonstrates that the MS measures the market volatility, and tests the first prediction about the market-wide component.

2.5.1 Topic Evolution

Intuitively, firm-level news provides insightful information about the company, and only marginally describes thoughtful details about macroeconomic matters. Indeed, from Table 2.1 it emerges that topics associated with *Earnings* or *M&A* account for 3% and 2% of the total distribution, whereas topics associated with the *USA Economy* or *Labour* account for 0.94% and 0.84%, respectively. To understand how the groups' distribution evolves over time, Figure 2.4 plots the staked area at a monthly frequency from January 1, 1990 to December 31, 2017. The figure shows that the distribution of some groups is observed more intensively in certain months than others (e.g., *M&A* and *Management*). In contrast, others are evenly distributed throughout the sample (e.g., *USA Economy* and *Media*).

Figure 2.5 depicts the monthly evolution of six topics reported in Table 2.1, by taking into account the number of times each topic is observed in any given month as a percentage of the total number of topics. These six topics can be reconciled to well known financial events, which showcase the ability of firm-level news to describe the state of the economy.

Panel A of Figure 2.5 shows distribution of the *M&A* topic (*Key Terms: deal, merge, acquisition, combine, buy, acquire, analyst, big, purchase, transaction*), which gains importance from 1994 to 2000, and spikes in 2006 and 2016, which are the periods of the three M&A waves. In addition, Panel B of Figure 2.5 shows the distribution of the *Website* topic (*Key Terms: site, online, service, search, offer, advertise, people, consume, customer, commerce*), whose coverage spikes around the dot-.com bubble and then steadily increases from 2004 onwards, which is consistent with the increase in online shopping and transition into a more digitalised economy. Moreover, Panel C of Figure 2.5 reports the importance of the *Bailout* topic (*Key Terms: govern, bank, financial, capital, fund, money, bailout, loan, crisis, rescue*), which surges in 2008 where the Big Bank Bailout program took place. Panel C of Figure 2.5 shows the distribution of the *Poor Prospective* topic (*Key Terms: analyst, problem, loss, industry, stock, cut, fall, loose, decline, drop*), whose distribution matches most of the crises. It increases in 1991 for the recession period, in 2001 for the dot-.com bubble, and 2009 for the financial crisis. A crisis or a poor perspective period often leads to redundancy programs. Hence, Panel E of Figure 2.5 shows the distribution of the *Lay-Off* topic (*Key Terms: cut, employee, job, work, plan, layoff, announce, close, operation, expect*), which spikes around the same periods as the *Website*, *Bailout* and *Poor Prospective* topics. For comparison reasons, Panel F of Figure 2.5 plots the *Opinion* topic (*Key Terms: people, real, problem, good, question, big, happen, fact, think, bad*), where no particular distribution is observed, since the media always express their opinion along with the seeable facts. Altogether, the evidence reported in this paragraph suggests that by grouping firm-level news according to topics, it is possible to understand the state of the economy.

The next set of tests document the ability of topics to change in concomitance with the company information. To this end, I collect earnings announcements dates from I/B/E/S and check whether, around these dates, there is an increase in coverage on topics related to the *Earnings* group. Panel A of Figure 2.6 reports the number of earnings announcements and topics in each month from January 1, 2010 to December 31, 2017 across all companies. From the figure it emerges that, around earnings announcements the media intensively covers matters related to earnings. In addition, Panel B of Figure 2.6 plots the similarity between firm-level news and an index composed by the *Key Terms* of the *Earnings* group from January 1, 2010 to December 31, 2010. The figure illustrates that the similarity between the news and the index increases in months where earnings announcements take place.

Panel C of Figure 2.6 shows the cross-sectional similarity for more than 5,000 earnings announcements without missing observations in the nine weeks around the announcement date from January 1, 1990 to December 31, 2017. To avoid the time and company effect, I subtract from each individual company's similarity the time-series and individual company mean. The zero event date represents the earnings announcements week. The fiigure shows that the similarity starts to increase two weeks before and goes back to the pre-event level three weeks after the announcement date. That is, before the announcement, the media starts to describe the upcoming release, on the event date it writes about the ongoing situation, and afterwards it explains the results. The variation in the similarity is statistically significant at the 5% level throughout the events study, confirming the reliability of the similarity measure to precisely quantify the composition of the news.[19]

In conclusion, the topics reflect the change in the market and company information environment, while the topic similarity promptly and precisely quantifies the amount of information that is also part of the news-index. That is, it increases when the news reports the same information as the news-index and decreases afterwords, demonstrating that the methodology proposed in this paper correctly quantifies the change in the composition of news.

2.5.2 Topic Similarity and Market Uncertainty

If the MS measures the change in the company risk, then the aggregation of all MS across all companies should measure the change in the market risk. The average MS is defined as:

$$AMS_t = \frac{\sum_{i=1}^{N} MS_{i,t}}{N}, \tag{2.4}$$

where the AMS_t is the Average Market Similarity, MS is from Equation 2.2 and N is the number of companies with a piece of news at time t. Hence, the AMS describes how similar all news in the market are. Following the same intuition as the MS: the lower (higher) the AMS, the higher (lower) the market risk. One can also see the AMS as an indicator of portfolio diversification. As the number of companies reporting the market topics increases, the greater amount of market-level information in stocks' returns (Roll, 1988), the less the diversification of the market portfolio will be and the greater the market risk. Stated differently, when a substantial number of companies simultaneously discuss the same topics, the AMS will be high and the market will probably be in a recession period.

To test whether the AMS measures the market risk, I use the Market Volatility Index (VIX) reported by the Chicago Board Options Exchange (CBOE). The VIX is calculated from the options' implied volatility of the S&P 500 constituent companies, and it is a measure of market uncertainty. This is not the first time words are linked to the VIX. For instance, Manela and Moreira (2017) show that words in the abstract of the front page of

[19]Panel A of Figure 2.6 accounts only for a sub-period for the sake of clarity; Panel B of Figure 2.6 accounts only for one company because the time lapse across companies is not evenly spaced therefore it posits some challenge in measuring the cross-sectional similarity over time; Panel C of Figure 2.6 accounts only for eight weeks because the coverage for more extended periods shrink substantially the number of companies.

the WSJ predict the VIX. To do so, they build a factor named News Implied Volatility (NVIX), which reflects words related to economic disaster concerns.

Figure 2.7 shows the weekly VIX (blue line), AMS (grey line) along with its ten-weeks simple moving average (yellow line) from January 1, 1990 to December 31, 2017. The figure demonstrates that the AMS and the VIX move together, and the AMS increases in recession periods. Table 2.5 reports the correlation between news and market related variables. The AMS and its first lag is positively correlated with the current value of the VIX, which implies that the bigger the market-wide component, the higher the market volatility.

Table 2.6 shows the regression results where the VIX is regressed on the AMS and a series of controlling variables. The weekly regressions (left-hand side of the table) are estimated with 1,457 observations. In Model 1, the VIX is regressed on the AMS, which shows a coefficient of 1.454 (t-stat of 6.18). That is, a 1% increase in similarity leads to a 1.45% increase in the VIX. Model 2 adds the number of news and the sentiment. In this second instance, the AMS decreases to 1.012 (t-stat of 4.89). By adding the return of the S&P 500, in Model 3, the explanatory power of the AMS slightly decreases but without affecting its statistical significance. Thus, the AMS describes the change in the aggregate market risk above and beyond the news sentiment, the number of news and current market return. Since the VIX is a measure of market uncertainty, it is possible to assert that the AMS is a proxy for market uncertainty.

The right-hand side of Table 2.6 adds the NVIX factor of Manela and Moreira (2017). Because the time-series of the NVIX is available only at a monthly frequency, I calculate the AMS monthly average, resulting in 312 observations from January 1, 1990 to March 31, 2016. Model 4 checks whether the AMS remains statistically significant at a monthly frequency. Hence, its coefficient is equal to 1.897 (t-stat of 4.40). Model 5 regresses the VIX on the NVIX. The coefficient of the NVIX is 0.942 (t-stat of 14.84). That is, in a univariate case, the AMS explains the variation of the VIX twice as much the NVIX. Model 6 regresses the VIX on the AMS and NVIX. The coefficient of the AMS is 1.603 and the NVIX is equal to 0.942, both statistically significant at the 1% level. Model seven simultaneously controls for the AMS, NVIX and first lag of the VIX. The AMS is 0.807 (t-stat of 3.89), the NVIX is 0.518 (t-stat of 6.53) and the first lag of the VIX is 0.516 (t-stat of 10.89). These results show that the explanatory power of the AMS is higher at monthly frequency than at weekly. Most importantly, the coefficient of the AMS is larger than the NVIX and statistically significant also after controlling for first lag of the VIX.

To check whether the AMS possesses any forecasting ability for the VIX, I use four years of in-sample data and a one month rolling window to forecast the next month VIX from May 1994 to March 2016 through an OLS regression. Then, I benchmark the performance of the AMS to the NVIX since, to the best of my knowledge, it is the most similar variable to the AMS.

The first column of Panel A in Table 2.7 shows the mean squared error (MSE) and the mean absolute error (MAE) of the AR(1) model, which are 4.63 and 3.12, respectively. Column two reports the same results when the NVIX is added to the first lag of the VIX. This second specification leads to an MSE of 4.83 and MAE of 3.47. This implies that an econometrician that adds the NVIX to a simple AR(1) model would do worse compared to simply using the AR(1) model. Column three reports the results when the AMS is added to AR(1). The MSE corresponds to 4.66 and the MAE is equal to 3.17. Hence, the AMS delivers higher forecasting accuracy than NVIX according to both statistics, but slightly worse than a simple AR(1).

To formally test the difference in MSE and MAE across models, Panel B of Table 2.7 reports the p-values of the the Diebold-Mariano test (Diebold and Mariano, 1995). The first column compares the AR(1) with and without the AMS. According to both the MSE and MAE, the two models deliver the same forecasting accuracy. Untabulatd results show that by adding the AMS to the AR(1) model the forecasting accuracy is better in 131 times out of 263 forecasts. Column two compares the AR(1) model with and without the NVIX. According to the MSE the two forecasts deliver the same accuracy. However, the result of the MAE shows that the difference of the two forecasts is statistically significant at the 5% level. The third column compares the model with the AMS and NVIX. Again, the Diebold-Mariano test shows that the forecasting results of the two models are not statistically significant different from zero according to the MSE; however, they are statistically significant at 10% level when comparing the MAE of the two models. It is worth highlighting that the MAE penalises the effect of outliers compared to the MSE, which make the former a more robust statistic.

There is one important difference between the NVIX and AMS. The AMS is entirely driven by the narrative of news, whereas the words used by the NVIX are selected by training a Support Vector Regression (SVR) on the VIX. That is, the AMS does not require any training on the VIX, while the NVIX is an ad-doc measure for the VIX. Furthermore, untabulated results (available on request) show that the AMS remains positive and highly statistically significant also after controlling for the uncertainty index of Baker et al. (2016).

2.6 Results

The other two predictions investigated in this paper are that the higher the MS, the higher the company stock return volatility and systematic risk. As previously stated, I reconcile both effects of the MS with uncertainty. A preliminary test to investigate the implication of the MS on these two risk variables is to create portfolios based on MS, and analyse the portfolios' characteristics. There are three benefits of studying the portfolios' characteristics instead of analysing the single stock characteristics. First, it overcomes the issue of having missing observations in the time-series. Second, it is possible to study the incremental effect of the MS and, at the same time, avoid biased inferences from company-specific characteristics. Finally, it allows to study whether a spread exists between the first and the last portfolios.

If the MS measures the company's systematic risk and the stock return volatility, and can be reconciled with uncertainty, then a high level of MS will lead to high stock return volatility, high systematic risk, and low R-squared. In each week, I sort the companies in ascending order according to their MS, and create five equally-weighted portfolios (i.e., quintiles). Hence, the quintile with the lowest MS is Q1, while the one with the highest MS is Q5. Since the market has unit systematic risk, Q1 (Q5) should have a systematic risk lower (higher) than one. Intuitively, due to the methodology used to build the MS, one of the middle quintiles should possess similar characteristics to the market.[20] Furthermore, I create a long-short strategy, where I take a long position on Q5 and short position on Q1.

Table 2.8 shows the weekly annualised sample statistics of the quintiles. On average, the five quintiles are composed of 21 companies each. Except for Q4, the annualised weekly return monotonically increases from Q1 to Q5, ranging from 8.1% to 14.3%, respectively. The weekly return volatility also increases from Q1 to Q5, ranging from 0.177 to 0.279. These statistics are translated into a yearly Sharpe ratio of 0.413 for Q5, and 0.303 for

[20] All companies in the sample are from the S&P 500 while listed.

Q1, while the market Sharpe ratio is 0.320. Furthermore, the long-short strategy has an annualised return of 6.2% with a standard deviation of 0.216, which leads to a Sharpe ratio of 0.161.[21]

Overall, Table 2.8 offers an initial insight into the effect of the MS. Companies with a small (large) market-wide component have lower (higher) return volatility. This implies that companies with high MS are less informative about their fundamentals, thereby making their returns more noisy (West, 1988). From a portfolio theory point of view, companies that are more (less) related to the market will also be more (less) related to each other within each quintile, confirming that stocks with greater market information comove more closely (Roll, 1988), increasing the portfolio volatility.

2.6.1 Stock Return Volatility and Systematic Risk

Table 2.8 suggests that the stock return volatility increases according to the MS. To formally test the effect of the MS on the stock return volatility, I calculate the company monthly realised volatility. I then regress the realised volatility on the MS and a series of controlling variables through a Pooled OLS regression.[22,23] Table 2.9 confirms the relationship between the MS and the stock return volatility. Model 1 shows that an increase by one standard deviation in the MS corresponds to an increase of 0.72% in the stock return volatility, where this variation is statistically significant at the 1% level. In Model 2, I add the first lag of the realised volatility of the company and the current realised volatility of the market as controlling variables. The first lag of the realised volatility controls for the high persistence of the volatility, while the current volatility of the market controls for the market component of the stock return volatility. After controlling for these two variables the coefficient of the MS drops to 0.221, statistically significant at the 5% level. Furthermore, in Model 3, I add two additional controlling variables: size and turnover. According to the results of this last specification, the MS remains virtually the same. These results, along with those reported in Table 2.8 strongly support the second prediction of the paper,

[21] In Appendix B.2 Table B.1, I report the same evidence as in Table 2.8 but for the tertile sorted portfolios.

[22] The Pooled OLS regression overcomes the issue about the absence of observations when the companies do not have news. Such regression methodology was also used by Tetlock et al. (2008) when analysing the effect of news sentiment to predict earnings

[23] Whenever the model accounts for the first lag of the dependent variable, the standard error are checked with the Arellano and Bond (1991) methodology, to account for the feedback of the lagged variable in the dependent variable.

which states that the higher the MS the higher the stock return volatility.

From a systematic risk point of view, Panel A of Table 2.10 shows that Q1 has a systematic risk of 0.962 , Q2 of 1.027, Q3 of 1.089, Q4 of 1.144, and Q5 of 1.191, where all coefficients are statistically significant at the 1% level. The distribution of betas across the five quintiles implies that the effect of the MS is not symmetric, but it is skewed. Indeed, the only quintile with a beta less than the market is Q1, whereas all the remaining quintiles have a beta greater than one. This is also confirmed by the beta of the long-short strategy that is 0.224. The difference of 23% in the systematic risk between Q1 and Q5 suggests that the systematic risk of the companies increases according to the MS, which is consistent with the theoretical study of Coles et al. (1995) and empirical one of Xing and Yan (2019).

Panel B of Table 2.10 presents a formal statistical test of the difference in systematic risk between the market and the five quintiles.[24] The table shows that, except for Q2, all the other quintiles' systematic risk differences are statistically significant at the 1% level. Again, only Q1 has a beta less than one, and this difference is equal to 3%. As mentioned above, the fact that the second portfolio has the same systematic risk as the market is due to the methodology used to build the MS. Hence, companies with an average MS have news where the information about the market and the firm are evenly distributed. This composition of firm-level news leads to a systematic risk that is equal to the company long-run average.

Furthermore, I check whether the time-varying systematic risk of the company is correlated with the MS. I estimate a Pooled OLS with 52 weekly observations and one-week rolling window, where the excess return of each company is regressed on the market excess return, as suggested by Lewellen and Nagel (2006). The regression also accounts for the company and year fix-effect, with two-way clustered standard errors. For this analysis, I use the AMS instead of the MS since the beta of the Pooled OLS is the average beta across all companies. Table 2.11 reports the Pearson correlation between the rolling beta and the AMS. The table shows that the AMS and the time-varying beta are correlated by 0.256, statistically significant at the 1% level. Hence, when investors are not able to obtain enough company-specific information from the news, the systematic risk of the company increases.

[24] The t-statistic of the difference in beta is calculated as follows: $t\text{-}stat_i = \frac{\beta_i - \beta_m}{\sqrt{se_i^2 - se_m^2}}$, where β_i and β_m are the beta of the portfolio and the market, respectively. In addition se_i^2 and se_m^2 are their corresponding standard errors. I assume that the market beta is one and its standard errors zero.

That is, investors will gauge companies according to the market-wide component in the news, which increases the comovement of returns between the company and the market. This increase in comovement leads to an increase in the company's systematic risk.

To formally test whether the MS captures the variation in the systematic risk, I estimate the monthly realised beta using the definition of Barndorff-Nielsen and Shephard (2004) with daily data:

$$R\beta_{i,t}^{(S)} \equiv \frac{R\mathrm{Cov}_{i,m,t}^{(S)}}{RV_{m,t,k}^{(S)}} = \frac{\sum_{k=1}^{S} r_{i,t,k} r_{m,t,k}}{\sum_{k=1}^{S} r_{m,t,k}^{2}}, \tag{2.5}$$

where $r_{i,t,k}$ and $r_{m,t,k}$ are the log return calculated as $r_{i,t,k} = \log P_{i,t,k} - \log P_{i,t,k-1}$. That is, the return of stock i at time t computed with k^{th} daily prices P (the same applies to the return of the market $r_{m,t,k}$). The realised covariance $R\mathrm{Cov}_{i,m,t}^{(S)}$ and the realised volatility of the market $RV_{m,t,k}^{(S)}$ is calculated S times within the month. For a high level of frequency S the realised beta can be seen as a noisy, but unbiased estimate of the integrated beta.

I use the $R\beta_{i,t}^{(S)}$ as a dependent variable in a Pooled OLS regression, where the independent variable is the MS, plus a series of controlling variables and the company and year fix-effect. Model 1 of Table 2.12 shows that the coefficient of the MS is 0.148 (t-stat of 1.75). After controlling for the first lag of the realised beta the coefficient becomes 0.137 (t-stat of 2.13)[25]. I control for the first lag of the $R\beta_{i,t}$ since beta tends to be autocorrelated (Andersen et al., 2006). In Model 3, I also add the company turnover since the change in beta around events can be driven by the change in volume (Denis and Kadlec, 1994) while, in Model 4, I add the realised volatility of the market. After controlling for all these variables the MS remains positive and statistically significant at the 10% level. These results imply that, on average, a 1% increase in the MS leads to approximately 0.12% increase in the realised beta.

Furthermore, to investigate whether the MS links the market and company returns, I regress the excess return of the company on the market excess return, the MS and the interaction between the MS and market excess return, plus firm-level controlling variables. If the MS captures the relationship between the market and the company returns, then the interaction term should be positive and statistically significant. Table 2.13 confirms

[25]Residuals are also estimated with Arellano and Bond (1991) methodology.

this prediction. The interaction term and the market coefficients are always positive and statistically significant at the 1% level, even after controlling for size and book-to-market (Model 2) as well as the first lag of the returns and the turnover (Model 3). Interestingly, the coefficient of the market across the three specifications is 0.49, while the coefficient of the interaction term is 0.95. This implies that the information contained in the MS, along with the return of the market, is more relevant than the market alone to describe the return of the company.

In conclusion, this section confirms that the MS explains the variation in the stock return volatility (prediction two) and the systematic risk (prediction three). In the next section, I investigate whether the MS also captures the pattern in the R-squared, as documented in the current literature.

2.6.2 R-squared

To understand whether the MS captures uncertainty, I estimate a simple OLS regression where each of the quintiles excess return is regressed on the market excess return. Panel A of Table 2.10 reports the regression results. The R-squared monotonically decreases from Q1 to Q5, where the R-squared Q1 is 0.78 and Q5 is 0.48, implying a difference of 65%. Hence, the MS is able to group companies in a way that the relationship between uncertainty and R-squared, documented in the current literature, is satisfied.

Untabulated results show that when the market model is estimated with a portfolio where the constituent companies have at least a news item in the given week, the R-squared is 0.86, which is 4.7% lower than the same model estimated with a portfolio where the constituent companies do not have news items in the given week.[26] In other words, when companies are separated according to their MS, the R-squared drops by 88% when compared with companies without news, which is a much larger drop than when the companies are simply divided into news and no-news.[27]

The pattern in the R-squared presented in Panel A of Table 2.10 is consistent with Teoh et al. (2009), who show that low R-squared is attributed to high firm uncertainty. That is, stocks with low R-squared incorporate less information about fundamentals, and

[26] Appendix B.2 shows the results when the same model is estimated with companies without identifiable news items.

[27] Table B.1 in Appendix B.2 shows that the same relationship holds when companies are divided into tertiles.

are therefore more difficult to price. These results, support the idea that a high level of MS is translated into high company uncertainty. Stated differently, investors trade companies by relying on market information due to the absence of firm idiosyncratic information, which leads to an increase in company uncertainty.

2.6.3 Application

In this section, I show the empirical implication of the MS from a portfolio manager point of view. If companies with high MS are riskier than other companies, then an investor can use the MS to manage and better understand its risk exposure.

To illustrate the implications of the MS, let us assume that there are two portfolio managers that hold the S&P 500 and want to estimate the Value at Risk (VaR). Also, for simplicity, let us assume that the two managers invest $1,000 every week.[28] In each week, one manager will simply invest in the S&P 500, while the second will sell stocks that lie on Q5 and rebalance the S&P 500 such as to invest all the $1,000 on the remaining stocks.

I estimate the weekly VaR from January 1, 1992 to December 31, 2017, with 104 in sample observations and one-week rolling window. To compare the expected loss of the two mangers, I subtract the VaR of the S&P 500 from the VaR of the manger that implements the strategy. A positive value implies that the loss of the strategy is higher than the S&P 500, while a negative value implies that the loss of the S&P 500 is higher than the strategy. I repeat the same calculation for the 10%, 5% and 1% level of α and report the results in Panel A, B, and C of Figure 2.8, respectively. Across the three specifications, by selling the companies in Q5, the VaR improves in almost all weeks. Stated differently, over the entire period, the sum of the excess loss predicted by the VaR with α equal to 1% is around $5,000 more for the manager who does not use the insights provided by the MS.

2.7 Concluding Remarks

In this paper, I propose a methodology to derive a new proxy variable for company uncertainty, which I call the Market Similarity (MS). The MS quantifies the information in firm-level news into a company-specific or market-wide component. I show that the

[28]This later assumption allows the expected loss to be comparable across weeks and managers.

more the firm-level news presents facts about the market rather than the company, the higher the company stock return volatility and the systematic risk. I reconcile this effect to the notion of company uncertainty, since companies with high MS have low R-squared. Furthermore, when the MS is aggregated across all companies, this measure explains the variation and forecasts the VIX.

The MS represents a useful contemporary measure to quantify the degree of company uncertainty. To the best of my knowledge, contrary to previous works in this vein that focus only on particular subsets of information disclosure, this paper is the first to provide a measure of company uncertainty derived from all firm-level news. It will be of interest to test whether grouping companies according to their MS leads to the same conclusion about the variance-ratio demonstrated by Boudoukh et al. (2019). Hence, what is more important: the type of event reported in the news, or how the news is related to the market? Answering this question could shed further light on the importance of news on risk and asset pricing. It will tell us whether we should study only a subset of news as suggested by Boudoukh et al. (2019) or, instead, if we should focus on news based on the MS.

Another important consideration about this paper is that it is fully dedicated to study the time-series implications of the MS. Future works might study whether the MS is a useful measure to explain the cross-section of returns. Preliminary and untabulated results appear to confirm this idea. I suggest that future studies implementing this methodology in a cross-sectional setting to increase the cross-sectional dimension, which is one of the limitations of this work.

Table 2.1: Topic Key Terms

The table shows the topics identified by the LDA over 458,636 firm-level new. The news covers all the companies while listed in the S&P 500 from the beginning of January 1, 1990 to December 31, 2017 for nine US sources. *Key terms* represents the output of the LDA and is completely data-drive. *Topic Label* is manually tagged given the collective meaning of the key terms. *Group* is manually tagged and classifies the topic label according to taxonomy. Each topic label is paired with its percentage *%*, which is calculated as the total number of times that a topic label is observed over the entire period and across all companies divided by the sum of all topic label counts.

Group	***Topic Label***	***Key Terms***	***%***
USA Economy	Funds Rate	economy, rate, rise, report, increase, economist, inflation, growth, interest rate, decline	1,50
	Political Campaign	group, campaign, political, support, member, president, state, lead, public, community	0,89
	National Defence	contract, program, military, defense, win, govern, build, work, satellite, official	1,50
	Regulation	bill, legislation, state, law, industry, lawmaker, support, proposal, administration, govern	0,67
	Bill Auction	bond, revenue bond, debt security, general obligation, competitive, morgan, issue, auction bill, bill, rate	0,16
	International Trade	country, foreign, govern, trade, export, international, world, global, dollar, currency	0,93
Labor	Work Union	union, work, strike, contract, agree, pilot, pay, labor, plant, employee	0,87
	Work Life	people, work, day, family, call, home, start, run, life, late	1,80
	New Hire	work, employee, job, hire, manage, people, pay, employer, program, train	0,92
	Work Inequality	woman, list, rank, black, survey, minority, study, top, white, report	0,44
	Lay-Off	cut, employee, job, work, plan, layoff, announce, close, operation, expect	1,43
	Pension	tax, plan, pay, benefit, pension, retire, save, money, retiree, cost	0,60
Health Care	Health Insurance	hospital, health care, patient, doctor, medical, plan, cost, care, health, insure	0,98
	Pharm	drug, product, sale, pharmaceutical, medicine, generic, sell, device, treat, stent	1,33
	New Treatment	drug, patient, study, treat, test, disease, cancer, research, develop, vaccine	1,34
	Life Insurance	insurance, insure, policy, coverage , claim, life insurance, premium, loss, unit, pay	0,56
Technology	New Tech	technology, develop, work, research, industry, project, engineer, create, start, compute	1,25
	Data Protection	security, data, information, compute, hack, attack, mail, privacy, customer, number	0,61
	Software	software, compute, product, operate, technology, program, develop, personal compute, customer, run	1,44
	Mobile Device	iphone, device, phone, sell, tablet, music, ipod, mobile, product, smartphone	0,72
	Network	network, service, wireless, customer, technology, phone, provide, carry, access, data	1,15
	Computer	compute, program, screen, software, feature, work, file, version, video, machine	0,78
	Processor	chip, compute, machine, technology, personal compute, microprocessor, product, manufacture, processor, sell	1,29
	Monitor	design, light, color, test, glass, display, paint, home, small, inch	0,62
	Phone	long distance, service, cable, customer, network, local phone, phone, offer, competition, local	0,75
Energy	Oil	oil, natural gas, energy, gas, drill, production, pipeline, refine, oil gas, produce	2,30
	Utilities	utility, power, energy, electricity, plant, state, coal, natural gas, customer, power plant	1,30

Table 2.1: Cont. Topic Key Terms

Group	***Topic Label***	***Key Terms***	***%***
Food & Drink	Agriculture	farm, corn, crop, seed, plant, produce, food, product, grow, agricultural	0,54
	Bars	bottle, brand, drink, tobacco, cigarette, beverage, product, brew, soft drink, smoke	0,91
	Restaurant	food, restaurant, product, brand, coffee, cereal, consume, chain, sale, fast food	1,14
Air Services	Airline	airline, carry, fare, flight, route, fly, service, travel, passenger, seat	0,94
	Airport	travel, airline, flight, passenger, ticket, airport, mile, fly, hub, carry	0,51
	Plane Producer	plane, order, aircraft, jet, airline, engine, airplane, flight, fly, delivery	0,44
Veichles	Car Producer	car, auto, vehicle, sale, automaker, model, plant, deal, sell, production	0,80
	Car Parts	car, vehicle, drive, model, engine, tire, truck, auto, automaker, hybrid	0,51
	Transport	railroad, ship, rail, delivery, truck, train, package, freight, service, transport	0,67
	Car Safety	problem, recall, safety, report, test, battery, find, work, crash, issue	0,92
Media	Website	site, online, service, search, offer, advertise, people, consume, customer, commerce	1,02
	Costumer Service	customer, call, service, phone, plan, charge, number, long distance, carry, pay	0,59
	Newspaper	publish, newspaper, magazine, book, news, advertise, read, paper, report, medium	0,56
	Entertainment	cable, network, program, television, medium, movie, channel, broadcast, subscribe, video	1,14
	Marketing	advertise, agency, campaign, brand, commercial, consume, medium, account, spend, add	1,10
Retail	Goods Consumption	product, print, brand, paper, consume, sale, sell, film, diaper, consume product	0,80
	Increase Sale	sale, retail, store sale, rise, store, report, gain, increase, consume, result	1,04
	Store	store, retail, chain, sale, customer, sell, shop, shopper, depart store, merchandise	1,96
	Leisure Products	toy, child, brand, sell, game, shoe, product, clothe, line, sale	0,79
Debt Instrument	Downgrade	bond, debt, security, rating, issue, investor, downgrade, sell, rat, note	0,67
	Debt Offering	offer, initial public, common, note, file, sell, issue, rating, warrant, unit	0,31
Industry	Poor Prospective	analyst, problem, loss, industry, stock, cut, fall, lose, decline, drop	1,39
	Metal	steel, plant, manufacture, product, produce, supply, factory, production, chemical, aluminium	1,26
	Real Estate	build, city, develop, square foot, area, project, property, office, real estate, space	1,48
	Competition	big, industry, analyst, strategy, small, change, grow, growth, recent, competitor	2,54
Financial Issue	Mortgage	mortgage, loan, lend, borrow, house, home, foreclosure, bank, rate, homeowner	0,75
	Bankruptcy	debt, bankruptcy, creditor, loan, pay, plan, file, file bankruptcy, finance, protection	0,65
	Bailout	govern, bank, financial, capital, fund, money, bailout, loan, crisis, rescue	0,48
	Bank Crisis	bank, morgan, jpmorgan, loss, capital, trade, financial, risk, firm, financial crisis	0,66

Table 2.1: Cont. Topic Key Terms

Group	***Topic Label***	***Key Terms***	***%***
Financial Services	Funds	fund, invest, investor, mutual fund, manage, money, asset, stock, investment, hedge fund	0,73
	Deposit	bank, loan, branch, deposit, lend, asset, customer, consume, institution, small	1,11
	Credit Cards	card, credit card, bank, customer, fee, consume, pay, charge, account, credit	0,87
	Broker	firm, invest bank, client, bank, broker, analyst, brokerage, stock, research, security	0,59
	Derivatives	trade, exchange, price, stock, option, derivative, call, contract, investor, future	0,51
Legal	Fine	settle, pay, agree, state, case, lawsuit, claim, agree pay, fine, charge	1,13
	Patent	case, court, rule, patent, claim, lawsuit, lawyer, suit, judge, sue	1,66
	Regulation	rule, commission, decision, agency, regulator, state, issue, review, official, require	1,27
	Investigation	investigation, report, document, information, official, state, practice, employee, letter, investigate	1,31
	Antitrust	govern, case, antitrust, browse, operate, software, court, lawyer, judge, competition	0,31
	Fraud	charge, prosecutor, case, form, trial, fraud, plead guilty, lawyer, indict, govern	0,73
M&A	Acquisition	deal, merge, acquisition, combine, buy, acquire, analyst, big, purchase, transaction	1,75
	Buyback	dividend, stock, plan, shareholder, cash, rate, bill sell, announce, common stock, buyback	0,59
	Takeover	offer, bid, shareholder, deal, board, yesterday, stock, cash, takeover, price	1,18
	Sale Unit	sell, plan, buy, unit, sale, pay, agree, work, offer, state	1,52
	Transaction Agreement	sell, sale, plan, unit, deal, stake, asset, invest, acquisition, purchase	2,65
	Acquire Subsidiary	service, unit, provide, acquire, sell, buy, subsidiary, operation, agree, purchase	2,77
Management	Nomination	president, vice president, manage, director, join, senior vice, officer, executive vice, appoint, president chief	1,33
	Termination	executive, chief executive, president, chairman, leave, manage, board, retire, lead, officer	2,04
	Compensation	pay, bonus, executive, compensation, receive, option, stock option, employee, salary, award	0,72
	Voting	board, shareholder, director, vote, investor, meet, manage, proposal, board member, annual meet	0,90
	Insider Transection	sell, hold, buy, insider, director, common, total, exercise option, officer sell, dire hold	0,33
	Agreement	deal, talk, agree, executive, meet, discussion, negotiation, accord people, official, comment	1,07
Accounting	Auditor	account, report, financial, earn, audit, auditor, investor, file, transaction, book	0,57
	Net Income	price, cost, buy, discount, pay, increase, consume, sell, high, offer	0,83
Earnings	Positive Forecast	quarter, earn, revenue, sale, expect, analyst, growth, result, profit, increase	2,71
	Positive Release	earn, rise, profit, quarter, net income, sale, early, second quarter, revenue, fourth quarter	2,46
	Negative Release	quarter, earn, loss, result, charge, fourth quarter, early, sale, revenue, income	3,93
Market	Index	stock, rise, gain, close, fell, investor, point, trade, high, index	1,41
	Analyst	stock, investor, earn, price, buy, analyst, high, hold, growth, sell	2,10
	Sharp Movement	rise, fell, stock, gain, decline, drop, lead, climb, big, day	0,78
	Investor Reaction	rise, fell, stock, investor, report, high, point, gain, trade, close	0,95
Other	Opinion	people, real, problem, good, question, big, happen, fact, think, bad	1,54
	Education & Gamble	hotel, school, student, room, casino, program, property, manage, college, education	0,66

Table 2.2: Summary Statistics

The table shows the sample statistics from January 1, 1990 to December 31, 2017, where: *Mean* is the mean; *Std* is the standard deviation; *Skw* is the skewness; $10^t h$ and $90^t h$ are the tenth and ninetieth percentile of the distribution. *No. Comp* is the average number of companies that have at least one news item in a given week; *No. News* is the weekly average number of news items per company; *Sent* is the weekly average of sentiment across all firm-level news, and calculated as: $(\#Positive - \#Negative)/(\#Positive + \#Negative)$, where $\#Positive$ and $\#Negative$ are the number of positive and negative words from the sentiment dictionary of Loughran and McDonald (2011). MS is the average Market Similarity calculated as in Equation 2.2; *Size* is the market capitalisation of the companies reported in thousands, which is calculated as the latest quarterly number of shares outstanding from Compustat, multiplied by the current stock's price; *BM* is the book-to-market of the company calculated according to Asness and Frazzini (2013); *Turn* is the company turnover reported in thousands, which is calculated as the average weekly volume divided by the last available number of shares outstanding. The first and last percentile of the distributions are winsorised.

	Mean	***Std***	***Skw***	**10^{th}**	***Median***	**90^{th}**
No.Comp	105	26	0.028	71	105	138
No.News	2.850	0.485	0.377	2.231	2.840	3.476
Sent.	-0.323	0.073	-0.197	-0.424	-0.319	-0.235
MS	0.653	0.018	0.038	0.631	0.653	0.677
Size	20594.6	10896.2	0.4	5792.1	21007.9	38186.6
BM	16.067	83.871	5.575	0.366	0.451	0.613
Turn	73.301	750.222	0.012	13.185	7.683	3.187

Table 2.3: Correlation - Main Variables

The table shows the weekly correlation calculated according to the Z-Transformation from January 1, 1990 to December 31, 2017. *No. News* is the the weekly average number of news items per company; *Sent* is the weekly average of sentiment across all firm-level news, and calculated as: $(\#Positive - \#Negative)/(\#Positive + \#Negative)$, where $\#Positive$ and $\#Negative$ are the number of positive and negative words from the sentiment dictionary of Loughran and McDonald (2011); MS is the average Market Similarity calculated as in Equation 2.2; *Ret* and $Ret_{t\text{-}1}$ are the current and first lag of the returns of the company; *Size* is the *log* of the market capitalisation of the companies, which is calculated as the latest quarterly number of shares outstanding from Compustat, multiplied by the current stock's price; *BM* is the book-to-market of the company calculated according to Asness and Frazzini (2013); *Turn* is the company *log* of turnover calculated as the average weekly volume divided by the last available number of shares outstanding. The first and last percentile of the distributions are winsorised.

	No.News	***Sent***	***MS***	***Ret***	***$Ret_{t\text{-}1}$***	***Size***	***BM***	***Turn***
No.News	1	-0.089^{***}	0.090^{***}	-0.003	0.001	-0.002	-0.005^{**}	0.226^{***}
Sent		1	-0.016^{***}	0.044^{***}	0.032^{***}	0.043^{***}	-0.081^{***}	-0.045^{***}
MS			1	-0.004	-0.001	-0.019^{***}	-0.048^{***}	0.045^{***}
Ret				1	-0.038^{***}	0.058^{***}	-0.067^{***}	-0.031^{***}
$Ret_{t\text{-}1}$					1	0.052^{***}	-0.055^{***}	-0.061^{***}
Size						1	-0.423^{***}	0.050^{***}
BM							1	0.142^{***}
Turn								1

Table 2.4: Summary Statistics of Firm-Level News

The table shows the covered period of the newspapers, the total number of firm-level news of each source, the average number of firm-level news per year as well as the proportion in percentage of each newspaper in the sample. The weekday circulation is taken from BurrellesLuce using the Audit Bureau of Circulations as data source in June 2013 (`https://www.burrellesluce.com`). The firm-level news is obtained from LexisNexis, with a company relevance score of at least 80% and with a minimum of 100 characters where 60% have to be alphabetic, for the companies while listed on the S&P 500. The Wall Street Journal is obtained from ProQuest. All firm-level news reflects a piece of news reported in the financial section of the respective newspaper. According to the weekday circulation the importance of these newspapers in the U.S. are: The Wall Street Journal (1), USA Today (2), New York Times (3), Washington Post (8), The Philadelphia Inquirer (18), The Atlanta Journal-Constitution (25), Pittsburgh Post Gazette (37),St. Louis Post-Dispatch (39), Tulsa World (75).

Newspaper	***Covered Period***	***Total No. of Articles***	***Mean No. Articles per Year***	***% of the Total Sample***	***Weekday Circulation***
New York Times	01.1990 - 12.2017	106,961	3,820	23.32%	1,865,318
The Wall Street Journal	01.1990 - 12.2017	106,917	3,818	23.31%	2,378,827
The Washington Post	01.1990 - 12.2017	55,904	1,997	12.19%	474,767
St. Louis Post Dispatch	01.1990 - 12.2017	41,817	1,493	9.12%	167,199
The Atlanta Journal-Constitution	01.1991 - 12.2017	33,028	1,223	7.20%	231,094
Pittsburgh Post Gazette	01.1993 - 12.2017	31,737	1,269	6.92%	180,433
USA Today	01.1990 - 12.2017	29,900	1,068	6.52%	1,674,306
Tulsa World	01.1996 - 12.2017	27,480	1,249	5.99%	95,063
The Philadelphia Inquirer	01.1994 - 12.2017	24,892	1,037	5.43%	306,831
Total No. of Articles	-	458,636	16.974		

Table 2.5: Correlation - Average Market Similarity

The table shows the Pearson weekly correlations between news and market variables from January 1, 1990 to December 31, 2017 for a total of 1,459 observations. *No. News* is the the weekly average number of news per company; *Sent* is the weekly average of sentiment across all firm-level news, and calculated as: $(\#Positive - \#Negative)/(\#Positive + \#Negative)$, where $\#Positive$ and $\#Negative$ are the number of positive and negative words from the sentiment dictionary of Loughran and McDonald (2011); *AMS* is the Average Market Similarity calculated as in Equation 2.4, while $AMS_{t\text{-}1}$ is the first lag of the AMS; *VIX* is the weekly Market Volatility Index reported by the Chicago Board Options Exchange (CBOE), *Mkt* is the weekly return of the S&P 500. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively.

	No.News	***Sent***	***AMS***	***VIX***	***Mkt***	$\boldsymbol{AMS_{t\text{-}1}}$
No.News	1	-0.188***	0.222***	0.246***	-0.024***	0.136***
Sent		1	-0.056***	-0.434***	0.090***	-0.067***
AMS			1	0.254***	-0.019	0.457***
VIX				1	-0.228***	0.200***
Mkt					1	-0.030
$\boldsymbol{AMS_{t\text{-}1}}$						1

Table 2.6: Regression - Average Market Similarity, VIX and NVIX

The table presents the results of the following OLS regression:

$$VIX_t = \alpha + \beta AMS_t + \gamma' \mathbf{X}_t + \epsilon_t,$$

where *VIX* is the Market Volatility Index (VIX) reported by the Chicago Board Options Exchange (CBOE), *AMS* is the Average Market Similarity calculated as in Equation 2.4 at weekly frequency (left-hand side of the table) and at monthly frequency (right-hand side of the table). The matrix of controlling variables $\mathbf{X}_t$ includes: *Sent* is the weekly average of sentiment across all firm-level news, and calculated as: $(\#Positive - \#Negative)/(\#Positive + \#Negative)$, where $\#Positive$ and $\#Negative$ are the number of positive and negative words from the sentiment dictionary of Loughran and McDonald (2011); *No. News* is the total number of news items in each week across all companies, *Mkt* is the weekly return of the S&P 500, *NVIX* is the monthly NVIX from Manela and Moreira (2017), $VIX_{t\text{-}1}$ is the first lag of the monthly VIX. *Alpha (Bps)* is the regression intercept in basis points, *Adj.*R^2 is the Adjusted R-squared, *No. Obs* is the number of observations used in the model. The left-hand side of the table covers the period from January 1, 1990 to December 31, 2017, while the right-hand side of the table covers the period from February 1, 1992 to March 31, 2016. The standard errors of the regressions are corrected for heteroskedasticity and with the Newey-West procedure with one lag and the t-statistics are reported between brackets. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively.

	Weekly			***Monthly***			
	1	2	3	4	5	6	7
Alpha (Bps)	19.343***	19.341***	19.450***	19.867***	-3.502***	-3.183**	-3.245***
	(69.20)	*(79.84)*	*(80.87)*	*(35.83)*	*(-2.46)*	*(-2.21)*	*(-2.62)*
AMS	1.454***	1.031***	1.012***	1.897***		1.603***	0.807***
	(6.18)	*(4.98)*	*(4.89)*	*(4.40)*		*(5.61)*	*(3.89)*
Sent		-3.130***	-3.000***				
		(-10.79)	*(-10.71)*				
No.News		1.112***	1.106***				
		(4.38)	*(4.60)*				
Mkt			-67.372***				
			(-4.68)				
NVIX					0.942***	0.930***	0.518***
					(14.84)	*(14.83)*	*(6.53)*
$\boldsymbol{VIX_{t\text{-}1}}$							0.516***
							(10.98)
No. Obs	1,460	1,460	1,460	312	312	312	312
Adj. $\boldsymbol{R^2}$	0.0345	0.2328	0.2676	0.0638	0.6306	0.676	0.8071
F-Test	0.001	0.001	0.001	0.001	0.001	0.001	0.001

Table 2.7: Contribution of the AMS and NVIX on the VIX Forecasting

The table shows the forecasting results of the Market Volatility Index (VIX) reported by the Chicago Board Options Exchange (CBOE). On the left-hand side of the two panels the mean squared error (*MSE*) and mean absolute error (*MAE*) are reported. The forecasting models account for four years of in-sample data and one month rolling-window to forecast the next month VIX through an OLS regression, from May 1, 1994 to March 31, 2016. **Panel A** shows the forecasting results, where column one shows the results of the AR(1) (Model 1), column two adds the NVIX of Manela and Moreira (2017) to the first lag of the VIX (Model 2), column three adds the average market similarity (AMS) to the first lag of the VIX (Model 3). **Panel B** reports the p-values of the Diebold-Mariano test (Diebold and Mariano, 1995). The first row indicates the forecasting models that are compared. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively.

Panel A: Mean Squared Error and Mean Absolute Error

	1	**2**	**3**
MSE	4.63	4.84	4.66
MAE	3.12	3.48	3.17

Panel B: Diebold-Mariano Test

	1 - 3	**1 - 2**	**2 - 3**
MSE	0.371	0.445	0.564
MAE	0.255	0.010**	0.029*

Table 2.8: Market Similarity Sorted Quintiles - Sample Statistics

The table shows the annualised weekly sample statistics from January 1, 1990 to December 31, 2017 of five equally-weighted quintile portfolios sorted in ascending order according to their MS in each week, where *Q1* is composed by companies with the lowest MS and the *Q5* is composed by companies with the highest MS, *Q5-Q1* is an investment strategy that goes long on Q5 and short on Q1, and *Mkt* are the statistics of the S&P 500. *Ret* is the average returns of the portfolios; *Std* is the standard deviation of returns; *SR* is the Sharpe ratio; *No. Comp* is the average number of companies in each portfolio; *Size* is the market capitalisation of the companies in thousands, which is calculated as the latest quarterly number of shares outstanding from Compustat, multiplied by the current stock's price, reported in in thousands; *Turn* is the turnover in thousands, which is calculated as the average weekly volume divided by the last available number of shares outstanding.

	Q1	***Q2***	***Q3***	***Q4***	***Q5***	***Q5-Q1***	***Mkt***
Ret	0.081	0.084	0.107	0.075	0.143	0.062	0.079
Std	0.177	0.195	0.212	0.219	0.279	0.216	0.162
SR	0.303	0.292	0.376	0.218	0.413	0.161	0.320
No.Comp	21	21	21	21	21	42	500
Size	50.627	47.222	47.103	47.117	47.646	-	-
Turn	70.726	139.610	218.992	356.392	134.199	-	-

Table 2.9: Regression - Market Similarity on Stock Volatility

The table presents the results from January 1, 1990 to December 31, 2017 of the following Pooled OLS regression:

$$RV_{i,t}^{(S)} = \alpha + \beta MS_{i,t} + \gamma' \mathbf{X}_{i,t} + \epsilon_{i,t},$$

where $RV_{i,t}^{(S)}$ is the monthly realised volatility of company i at month t calculated with (S) daily observations; $MS_{i,t}$ is the market similarity, calculated as in Equation 2.2. $\mathbf{X}_{i,t}$ is a matrix including the year and company fix-effect, as well as: $RV_{i,t\text{-}1}$ and RV_m are the first lag of the company and the contemporaneous market realised volatility; *Size* is the *log* of the market capitalisation of the companies, which is calculated as the latest quarterly number of shares outstanding from Compustat, multiplied by the current stock's price; *Turn* is the *log* of turnover calculated as the average weekly volume divided by the last available number of shares outstanding. *Alpha* is the regression intercept; *No. Obs* is the number of observations used in the model, R^2 is the model *R*-squared. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm and year.

	1	**2**	**3**
Alpha	13.921***	-5.886*	-12.050***
	(18.09)	*(-1.92)*	*(-4.21)*
MS	0.722***	0.221**	0.217**
	(3.81)	*(2.15)*	*(2.13)*
$\boldsymbol{RV_{i,t\text{-}1}}$		0.373***	0.370**
		(4.80)	*(4.78)*
$\boldsymbol{RV_m}$		0.785***	0.776***
		(4.56)	*(4.51)*
Size			0.109
			(0.927)
Turn			0.695***
			(3.72)
No.Obs	23,404	23,404	23,404
$\boldsymbol{R^2}$	0.0063	0.5961	0.5976
F-Test	0.001	0.001	0.001
Year-Fix	Yes	Yes	Yes
Comp-Fix	Yes	Yes	Yes

Table 2.10: Market Similarity Sorted Quintiles - Systematic Risk and R-squared

The table presents the results from January 1, 1990 to December 31, 2017 of the market model estimated for each of the quintile portfolio:

$$r_t = \alpha + \beta r_{m,t} + \epsilon_t,$$

where r_t is the weekly excess return at time t of the five equally-weighted quintile portfolios sorted in each week in ascending order according to their Market Similarity (*Q1*, *Q2*, *Q3*, *Q4* and *Q5*), and for the strategy (*Q5-Q1*) that takes a long position on Q5 and a short position on Q1, $r_{m,t}$ is the excess market return from the Fama-French website, and α is the model intercept (not reported). **Panel A** shows the systematic risk (β) of the quintiles, and the t-statistics between brackets calculated with standard errors corrected for heteroskedasticity and with the Newey-West procedure with one lag. *No. Obs* is the number of weekly observations used to estimate the model, and R^2 is the the R-squared of the model. **Panel B** shows the same coefficients β as in Panel A, and the difference between each of the quintiles coefficient and the beta of the market, where the market is assumed to have beta equal to one and standard error equal to zero; the t-statistics of the difference in beta are reported between brackets. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively.

Panel A: Quintiles Systematic Risk and R-squared

	Q1	***Q2***	***Q3***	***Q4***	***Q5***	***Q5-Q1***
$\boldsymbol{\beta}$	0.967***	1.027***	1.089***	1.144***	1.191***	0.224***
	(49.32)	*(35.67)*	*(44.53)*	*(41.49)*	*(28.52)*	*(5.07)*
No.Obs	1,460	1,460	1,460	1,460	1,460	1,460
$\boldsymbol{R^2}$	0.7875	0.7288	0.6913	0.7185	0.4782	0.0283

Panel B: Difference Systematic Risk - Quintiles vs. Market

	Q1	***Q2***	***Q3***	***Q4***	***Q5***	***Q5-Q1***
$\boldsymbol{\beta}$	0.967	1.027	1.089	1.144	1.191	0.224
Diff	-0.033***	0.027	0.089***	0.144***	0.191***	-0.776***
	(-2.88)	*(1.62)*	*(4.85)*	*(7.27)*	*(6.35)*	*(-27.26)*

Table 2.11: Correlation - Average Market Similarity and Systematic Risk

The table shows the Pearson correlation between the systematic risk (*beta*), the regression intercept (*Alpha*) and the Average Market Similarity (*AMS*) calculated according to Equation 2.4. The systematic risk and intercept are estimated with a Pooled OLS regression from January 1, 1991 to December 31, 2017:

$$r_{i,t} = \alpha + \beta r_{m,t} + \epsilon_{i,t},$$

where $r_{i,t}$ is the weekly excess return of company i at time t and $r_{m,t}$ is the excess return of the market from the Fama-French website. The model is estimated with 52 weeks observations and one week rolling window. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively.

	Alpha	***Beta***	***AMS***
Alpha	1	-0.267***	0.018
Beta		1	0.256***
AMS			1

Table 2.12: Regression - Market Similarity on Realised Beta

The table shows the results of a Pooled OLS from January 1, 1990 to December 31, 2017:

$$R\beta^{(S)}_{i,t} = \alpha + \beta MS_{i,t} + \gamma' \mathbf{X}_{i,t} + \epsilon_{i,t},$$

$R\beta^{(S)}_{i,t}$ is the realised beta of company i at month t calculated with (S) daily observations as in Equation 2.5, $MS_{i,t}$ is the Market Similarity of company i calculated as in Equation 2.2. $\mathbf{X}_{i,t}$ is a matrix including the year and firm fix-effect, as well as: $R\beta_{i,t\text{-}1}$ is the first lag of the realised beta; *Turn* is the company *log* of turnover calculated as the average weekly volume divided by the last available number of shares outstanding; RV_m is market realised volatility. *Alpha* is the regression intercept; *No. Obs* is the number of observations used in the model; R^2 is the model R-squared. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm and year.

	1	**2**	**3**	**4**
Alpha	0.937***	0.549***	-0.597***	-0.622***
	(15.04)	*(9.33)*	*(-5.47)*	*(-5.48)*
MS	0.148*	0.137**	0.112*	0.117*
	(1.75)	*(2.13)*	*(1.68)*	*(1.77)*
$\boldsymbol{R\beta_{i,t\text{-}1}}$		0.382***	0.340***	0.339***
		(13.92)	*(13.13)*	*(13.21)*
Turn			0.155***	0.158***
			(13.43)	*(13.37)*
$\boldsymbol{RV_m}$				-2.988*
				(-1.64)
No.Obs	24,930	24,930	24,930	24,930
$\boldsymbol{R^2}$	0.0106	0.1783	0.2070	0.2087
F-Test	0.001	0.001	0.001	0.001
Year-Fix	Yes	Yes	Yes	Yes
Comp-Fix	Yes	Yes	Yes	Yes

Table 2.13: Regression - Interaction Between Market Return and Market Similarity

The table shows the results of a Pooled OLS from January 1, 1990 to December 31, 2017:

$$r_{i,t} = \alpha + \beta r_{m,t} + \zeta MS_{i,t} + \psi r_{m,t} * MS_{i,t} + \gamma' \mathbf{X}_{i,t} + \epsilon_{i,t},$$

where $r_{i,t}$ is the weekly excess return of company i at time t; $MS_{i,t}$ is the Market similarity of company i calculated as in Equation 2.2, $r_{m,t}$ is the excess return of the market from the Fama-French website. $\mathbf{X}_{i,t}$ is a matrix including the year and company fix-effect, as well as: *Size* is the *log* of the market capitalisation of the companies, which is calculated as the latest quarterly number of shares outstanding from Compustat, multiplied by the current stock's price; *BM* is the book-to-market of the company calculated according to Asness and Frazzini (2013); *Turn* is the company *log* of turnover calculated as the average weekly volume divided by the last available number of shares outstanding. *Alpha (Bps)* is the regression intercept in basis points, *No. Obs* is the number of observations used in the model, R^2 is the model *R*-squared. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm and year.

	1	**2**	**3**
Alpha (Bps)	-0.002	-1.412***	-2.511***
	(-0.95)	*(-4.20)*	*(-3.16)*
Mkt	0.497***	0.493***	0.493***
	(3.26)	*(3.22)*	*(3.24)*
MS	0.002	0.002	0.002
	(0.58)	*(0.55)*	*(0.50)*
Mkt*MS	0.954***	0.953***	0.952***
	(3.12)	*(3.58)*	*(3.59)*
Size		0.001***	0.002***
		(4.88)	*(5.12)*
BM		-0.001***	-0.001***
		(-2.90)	*(-2.86)*
$\boldsymbol{Ret_{i,t\text{-}1}}$			-0.017***
			(-2.61)
Turn			0.001
			(1.30)
No.Obs.	153,307	153,075	153,075
$\boldsymbol{R^2}$	0.1139	0.1148	0.1149
F-Test	0.001	0.001	0.001
Year-Fix	Yes	Yes	Yes
Comp-Fix	Yes	Yes	Yes

Figure 2.1: Topics Coherence Measure

The figure shows the coherence measured by the U-Mass. The higher the coherence level, the better the quality of topics. The coherence measure is calculated by splitting the dataset in two: training and test. The training part is used to calculate the topics' distribution and the test part is an unseen sub-sample where the log-likelihood is estimated as follows:

$$UMass(w_i, w_j) = log\frac{D(w_i, w_j) + 1}{D(w_i)},$$

where $D(w_i, w_j)$ is the number of times that word i and j are observed together in one document and $D(w_i)$ is the number of times that the word i is observed in the document. This equation is estimated for different numbers of topics, that is from 30 to 140 topics with an interval of 10 topics measured in the *x-axis*. Instead, the *y-axis* measures the U-Mass score.

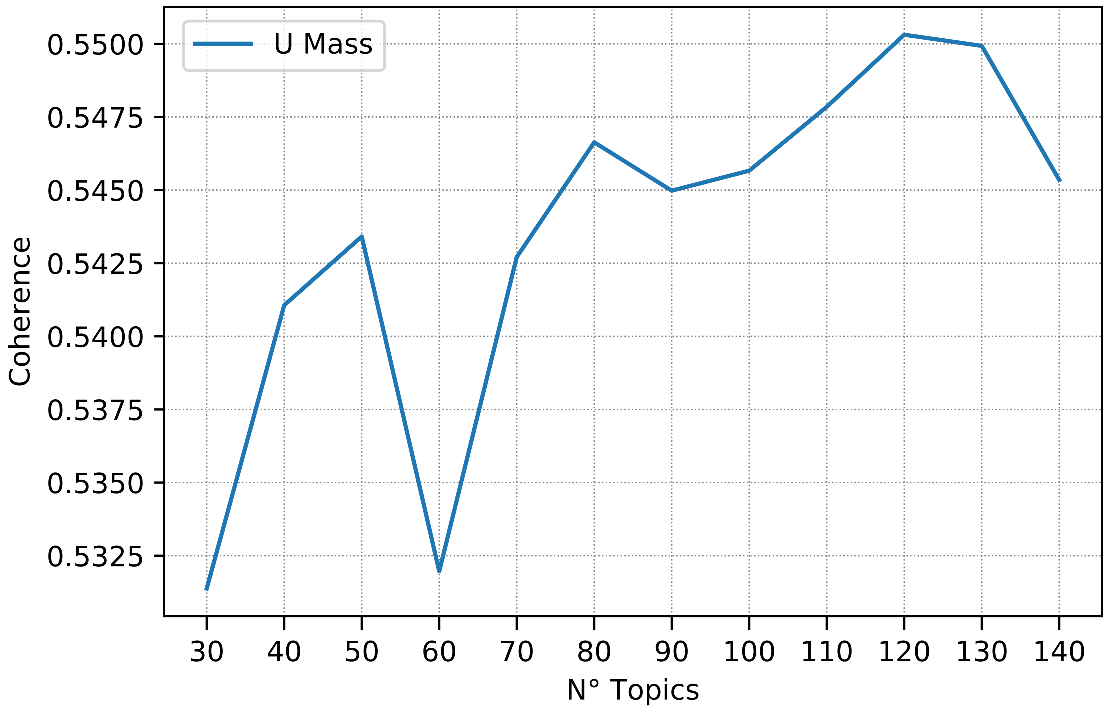


Figure 2.2: Heatmap - Topics Similarity

The figure shows the similarity between the topics reported in Table 2.1. The squares represent the similarity calculated according to the Cosine Similarity of the vector representation of the Key Terms as in Equation 2.1. The brighter (darker) the square, the more (less) similar the two topics are. The label of the figure stands for the topics' groups which are manually assessed and based on the taxonomy of the topics.

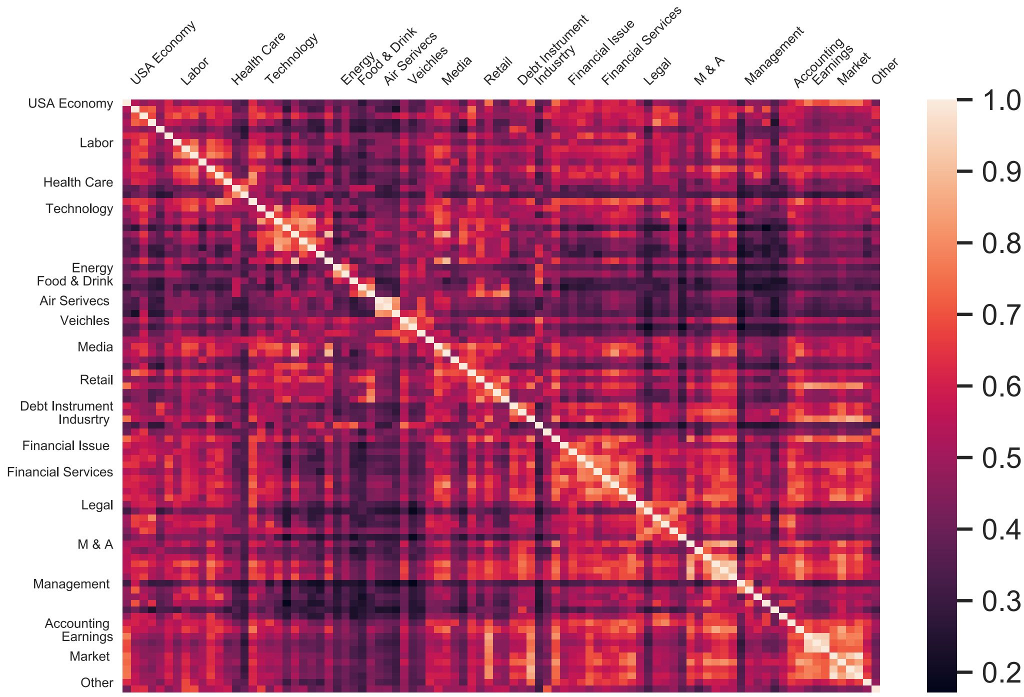


Figure 2.3: Distribution of Firm-level News

The figure shows the distribution of the yearly number of firm-level news from January 1, 1990 to December 31, 2017 for the companies while listed on the S&P 500. The firm-level news is obtained from LexisNexis, except for The Wall Street Journal that is obtained from ProQuest with a relevance score of at least 80%. All firm-level news is from the financial section of the respective journal with a minimum of 100 characters and 60% of alphabetic characters. The sample accounts for firm-level news of the following newspapers: The New York Times (NYT), USA Today (UST), Washington Post (WaP), St. Louis Post Dispatch (SLP), The Atlanta Journal-Constitution (AjC), The Philadelphia Inquirer (PhI), Pittsburgh Post Gazette (PPG), Tulsa World (TuW) and The Wall Street Journal (WSJ). The black-dashed line is the yearly mean across all sources and companies included in the sample.

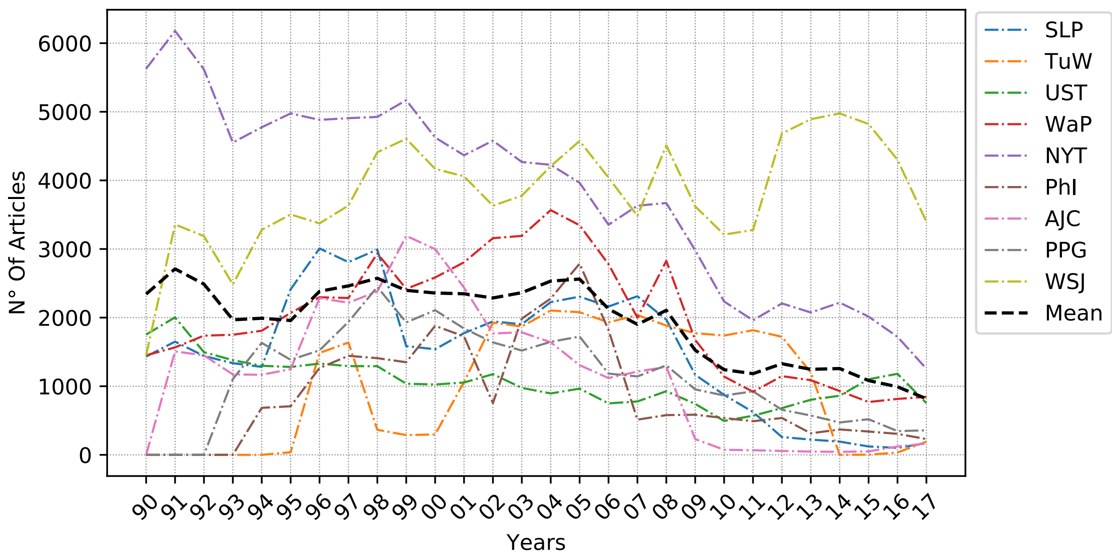


Figure 2.4: Sacked Area Plot - Distribution of Topics' Groups

The figure shows the monthly stacked area of all groups reported in Table 2.1 from January 1, 1990 to December 31, 2017. In each month the proportion of each individual group is calculated in respect of all groups, which reflects the high of the coloured area of each group. Consequently, the high of the figure sum to one in each period. Each colour represents an individual group label reported on the left hand side of the figure.

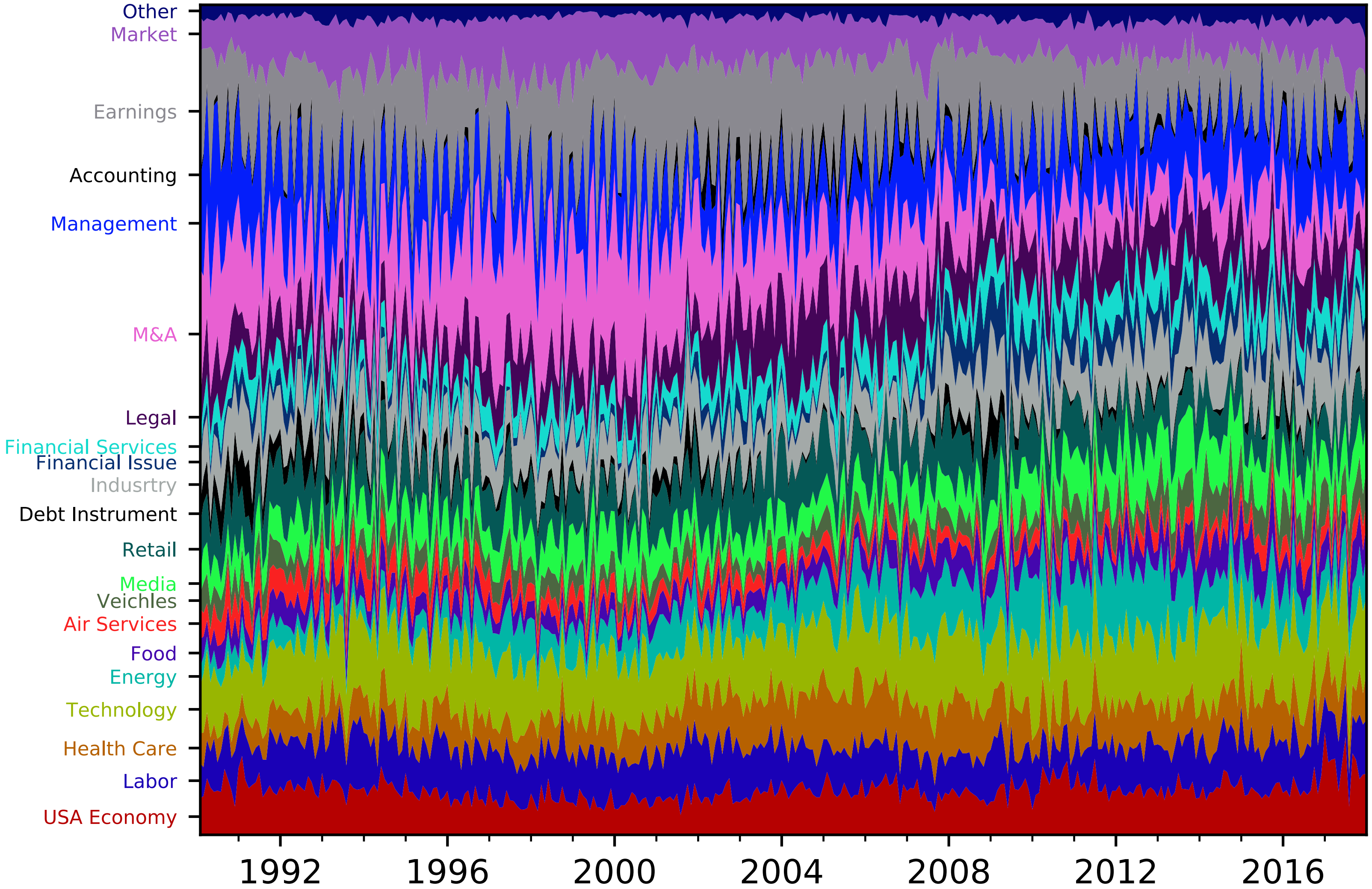


Figure 2.5: Distribution of Selected Topics

The figures plot the monthly time-series distribution of the six topics' labels reported in Table 2.1. The black line represents the number of times a given topic is observed in a the firm-level news in each week as a percentage of all topics across all companies. The resulting percentages are then summed in each calendar month.

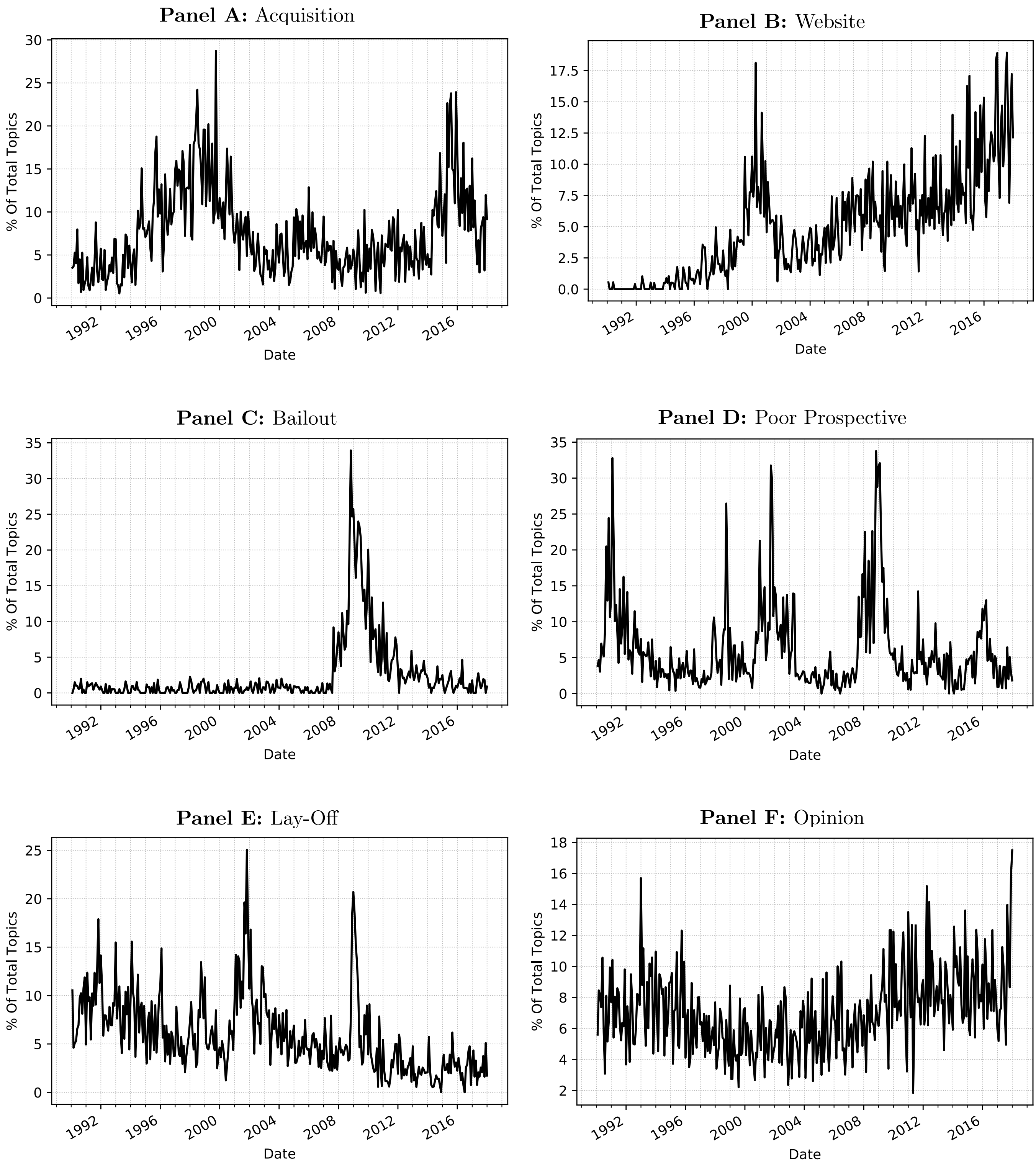
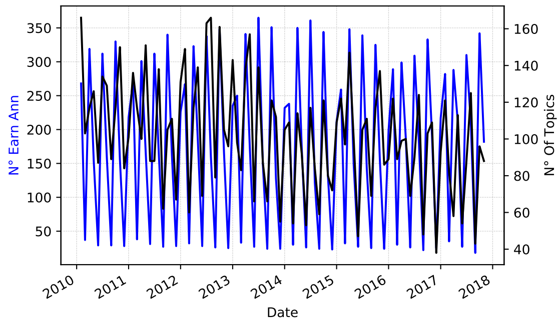


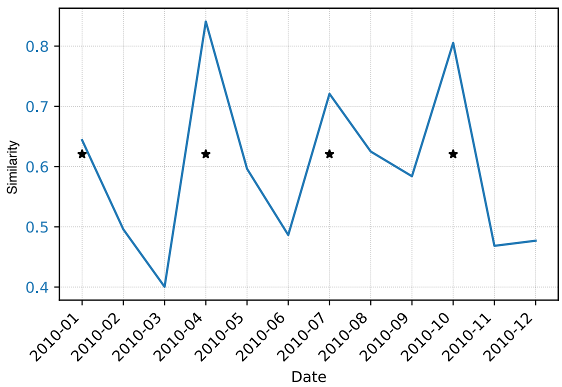
Figure 2.6: Earnings Announcements' Topics and Topics Similarity

The figure shows the relation between the earnings topic's group reported in Table 2.1 and the actual earnings announcements dates from I/B/E/S. **Panel A** shows the number of earnings announcements (*blue line*) across all companies from January 1, 2010 to December 31, 2017 and the number of times that any of the topics related to the Earnings group are observed (*black line*). **Panel B** shows how the similarity (*blue line*) between the monthly topics inferred from firm-level news and the Earnings group calculated according to Equation 2.1, where the *black stars* represent the earnings announcements dates. **Panel C** shows the similarity between the topics of the firm-level news and the Earnings Group according to Equation 2.1 across all companies for nine weeks around the earnings announcement dates (Event Date 0). The *blue line* shows the average similarity minus the cross-sectional and time-series average, while the *black dashed lines* represent the 95% confidence bounds.

Panel A: Number of Earnings Announcements and Topics



Panel B: Similarity of News and Earnings Group



Panel C: Cross-Sectional Similarity Around Earnings Announcements

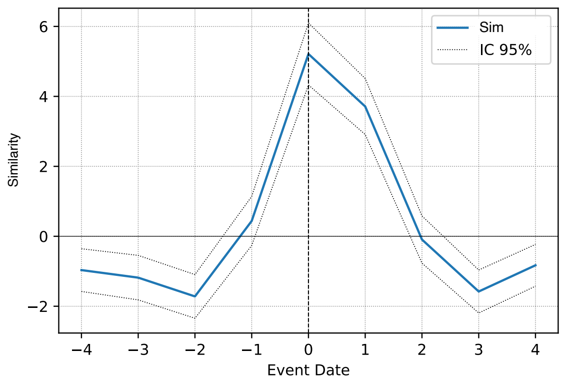


Figure 2.7: Average Market Similarity and VIX Level

The figure shows the level of the Market Volatility Index (VIX) reported by the Chicago Board Options Exchange (CBOE), the Average Market Similarity (*AMS*) calculated as in Equation 2.4 and its 10 weeks Simple Moving Average (*10 SMA*) from January 1, 1990 to December 31, 2017 in a weekly frequency. The AMS is reported in *grey*, the 10 weeks SMA in *yellow* and are measured on the left-hand side of the figure (*Distance*). The VIX in reported in *blue* and is measured on the right-hand side of the figure.

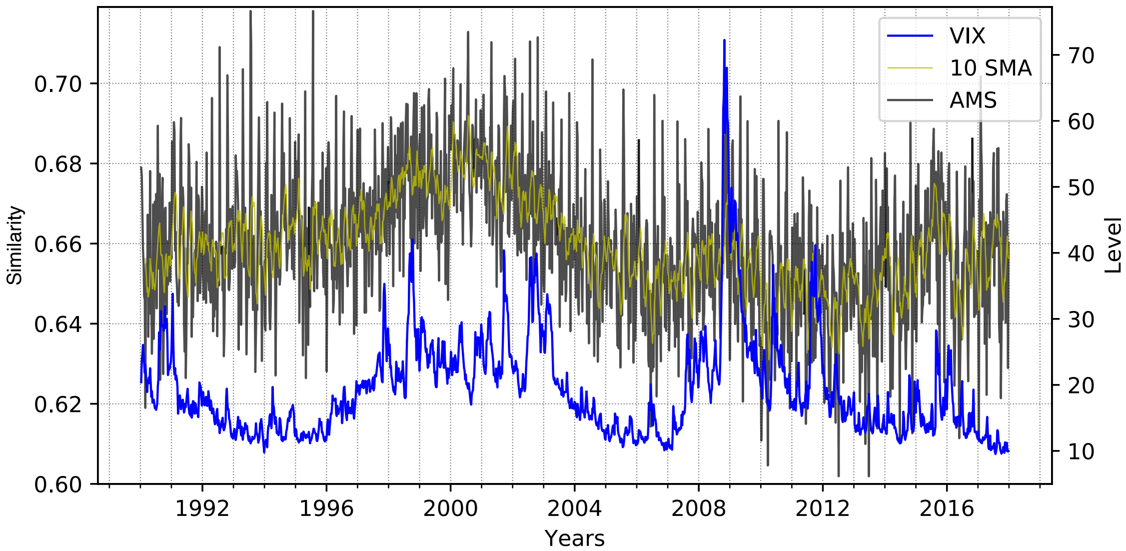
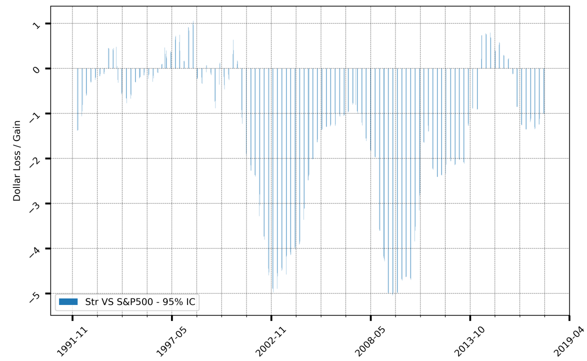


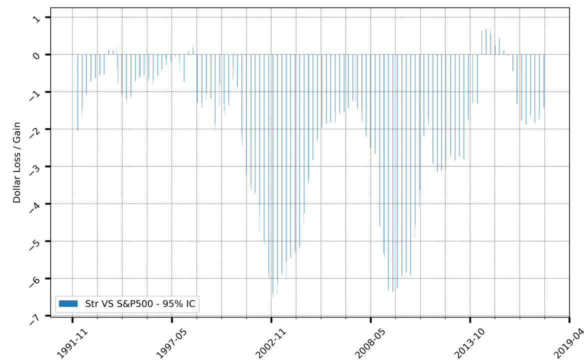
Figure 2.8: Value at Risk - Implication of High Market Similarity

The figure shows the expected loss calculated with the Value at Risk (VaR) with α equal to 10%, 5% and 1%, respectively. The figure reports the expected loss difference between the S&P 500 and the S&P 500 where the companies with high MS (i.e., Q5) are short in each week. The VaR is calculated by assuming that in each week $1,000 is invested in both portfolios, calculated with 104 observations and one-week rolling window.

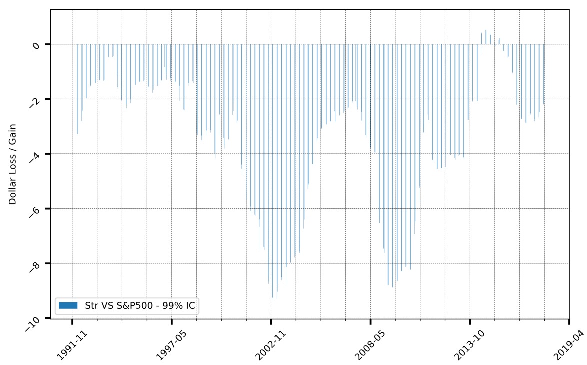
Panel A: VaR α 10%



Panel B: VaR α 5%



Panel C: VaR α 1%



Chapter 3

Are the Primary Dealers of the New York Fed Really Special?

3.1 Introduction

Intermediary asset pricing provides a novel perspective on the role of financial intermediaries as marginal investors in the major asset markets. The intermediary asset pricing models of He and Krishnamurthy (2012, 2013) and (Brunnermeier and Sannikov, 2014) provide the foundations and key determinants of the pricing kernel of financial intermediaries. On the empirical side, Adrian et al. (2014) propose an intermediary pricing kernel with broker-dealer leverage shocks as a single risk factor. In a similar vein, He et al. (2017, HKM) advocate the use of shocks to the equity capital ratio of primary dealers (PDs) in a two-factor model along with the market factor. HKM provide extensive evidence for significant risk premia and explanatory power of this intermediary capital ratio factor across seven asset classes and contrast their findings with those in Adrian et al. (2014). Overall, the paper offers strong empirical support to the capital ratio of PDs as a priced factor across asset classes with no evidence of pricing for the widely used market factor. This leads HKM to conclude that PDs are special marginal investors with very different characteristics relative to the non-primary broker-dealer sector.

This evidence raises two questions. First, can HKM's aggregate capital ratio and intermediary risk factors be replicated? HKM provide the final series that are used in their empirical analysis, but they do not make the generating code publicly available. In fact, it is not clear how HKM manually match the New York Fed's (NY Fed) PDs to their domestic and foreign ultimate (publicly traded) parent companies. Gospodinov and Robotti (2021) show that there is nothing special about NY Fed's PDs when using HKM's data and robust statistical methods. In contrast, in this paper we are interested in understanding whether HKM's capital ratio and intermediary risk factors are correctly computed. The NY Fed's website is of little help here because it provides a list of domestic and foreign PDs but no information on their ultimate parent companies. Since HKM identify PDs with their ultimate parents, it is of crucial importance to get the ultimate parents right in the analysis. Second, are PDs special broker-dealers as argued by HKM and can they be considered the relevant marginal investors in multi-asset markets? In particular, we are interested in determining whether (i) the intermediary factor of HKM is priced in the cross-section of asset returns; (ii) there is an announcement effect in terms of risk and return from becoming active NY Fed's PDs; and (iii) the performance of the NY Fed's PDs deviates substantially from that of the non-primary broker-dealer sector. If the answer to this question is affirmative, this would elevate the status of the intermediary capital factor in enhancing our understanding about the underlying drivers of the cross-section of asset returns. We provide evidence that sheds light on these questions.

Our sample construction delivers aggregate capital ratio and intermediary risk factor series that differ substantially from those in HKM. While it is difficult to attribute the differences in the series to one particular reason, we suspect that HKM performed a heavy and ad-hoc trimming of the list of ultimate parent companies to obtain their capital ratio and risk factors. When we eliminate the Japanese parent companies (and a few more selected companies) from the sample, we obtain virtually the same results as in HKM. Ultimately, this discrepancy issue could be solved if the NY Fed were to make publicly available the list of ultimate parent companies of its PDs. The rest of the analysis is conducted using our newly built and updated sample. While it may be prestigious for a broker-dealer to become a New York Fed's PD, we find no price reaction around the day in which a broker-dealer becomes a PD. In essence, there is no evidence of substantial changes in risk and return for those dealers that become active PDs. Finally, we examine whether PDs are special relative to non-primary broker-dealers by employing a difference-in difference (diff-in-diff) approach with PDs in the treatment and broker-dealers in the control groups, respectively. Again, we find no meaningful differences between treatment and control groups, which implies that there may be nothing too special about NY Fed's PDs relative to the wider financial sector. NY Fed's PDs are certainly large finance intermediaries but they do not seem to differ substantially in performance from the non-primary sector based on their risk/return profile and several other characteristics.

The rest of the paper is organized as follows. Section 3.2 deals with the sample construction and reports some descriptive evidence for the aggregate capital ratio and characteristics of the NY Fed's PDs. Section 3.3 describes our empirical strategy with particular emphasis on the cross-sectional asset pricing, event-study, and performance analyses of the domestic and foreign PDs. In Section 3.4, we provide some preliminary insights on PDs and the performance of HKM's intermediary capital model vs. the capital asset pricing model (CAPM). The main results of our analysis can be found in Section 3.5. Section 3.6 summarises our main conclusions. Additional material is provided in the appendix.

3.2 Sample Construction and Descriptive Statistics

We obtain the historical list of PDs from 1970 to 2014 as well as all the relevant information to update the list up to 2020 from the NY Fed's website.[1] Consistent with HKM, we construct the aggregate capital ratio and capital risk factors for the intermediary sector by matching the NY Fed's PD list with Center for Research in Security Prices (CRSP)/Compustat (for the domestic sample) and Datastream (for the foreign sample) data on their publicly traded holding companies. Following the authors approach, we manually match the PDs reported by the NY Fed to their ultimate parent companies that are characterised by unique global company keys (GVKEY) and permanent numbers (PERMNO) for the domestic and mnemonics (MNEM) for the foreign samples, respectively. We then categorise each PD according to its ultimate parent's headquarters. Specifically, a PD is classified as foreign if it is controlled by a company outside of the US and domestic if it is controlled by a US company. Furthermore, for each identified publicly-listed company, we track its mergers, acquisitions, and name change over time to ensure that we employ the most relevant and updated information about the ultimate parent company.

In Appendix C, we report the main results of this replication exercise and compare our newly constructed series with those in HKM. Without delving into details, we are able to obtain correlations of around 0.97 between our and HKM's series.[2] However, it should be noted that these high correlations are observed only when some companies are excluded

[1]The original list is available at the following website `https://www.newyorkfed.org/markets/primarydealers`.

[2]HKM's series are available at `http://apps.olin.wustl.edu/faculty/manela/data.html`.

from the sample (i.e., all of the Japanese companies and some other selected firms). Unless specified otherwise, throughout the paper we employ our own series (that do not exclude any companies) and have correlations of around 0.55 with those of HKM. Tables 3.1 and 3.2 display the full list of domestic and foreign ultimate parent companies for all of the NY Fed's PDs.

From 1970 to 2020 there are 177 PDs, where 74 are foreign and controlled by 33 distinct parent companies and 80 are domestic and controlled by 47 distinct public holding companies. The remaining 23 PDs are privately held. It is worth emphasising that a few ultimate parents control more than one PD over some particular time periods. To understand the geographic distribution of the parent companies over time, Figure 3.1 plots the number of PDs that appear at least once in a given quarter.

We group the PDs based on the currency in which the stock of the ultimate parent is traded. In total, there are seven different geographic areas outside of the US: Japan (JPY), Great Britain (GBP), Euro Area (EUR), Switzerland (CHF), Australia (AUD), Canada (CAD), and Hong Kong (HKD). Before 1975 there are only domestic PDs, from 1975 to 1983 British companies also become active PDs, and from 1990 onwards PDs from other regions appear in the sample. The unconditional average number of PDs per quarter is 22.7, where 8.1 are foreign and 14.6 are domestic. The higher number of domestic companies is driven by the first half of the sample, where the only foreign companies are those from Great Britain. Over the 1990 to 2020 sample period, the average number of PDs is 25.6, where 14.3 are foreign and 11.3 are domestic.

Moreover, we analyse the standard industry classification (SIC) of the ultimate parent companies over time. Figure 3.2 shows the distribution of the ultimate parents over time based on SIC codes. Among the twelve different SIC that appear at least once in the sample, the most frequent ones are by far (i) security brokers, dealers, and flotation companies (6211) and (ii) commercial banks (6020).[3] Except for two companies that are classified as industrial conglomerates (9997) and department stores (5311), the remaining institutions are classified as finance, insurance, and real estate (6000-6799).

[3]The SIC codes that appear at least once in the sample are (i) department stores (5311); (ii) commercial banks (6020); (iii) personal credit institutions (6141); (iv) mortgage bankers and loan correspondents (6162); (v) finance services (6199); (vi) security and commodity brokers (6200); (vii) security brokers, dealers, and flotation companies (6211); (viii) investment advice (6282); (ix) life insurance (6311); (x) fire, marine, and casualty insurance (6331); (xi) investors, not elsewhere classified (6799); and (xii) industrial conglomerates (9997).

To determine wether there is a difference between the domestic and foreign parent companies, we perform some accounting and ratio analyses over the 1970 to 1990, 1990 to 2020, and 1970 to 2020 sample periods. Panel A of Table 3.3 displays the asset, debt, equity and market capitalisation positions of the domestic and foreign parent companies as well as of the two combined (USD, in millions).

The foreign PDs have more assets and book debt but a smaller market cap compared to the domestic PDs in the first two periods. In terms of book equity, the foreign companies have more equity in the first period and less in the second. Panel B of Table 3.3 displays the debt-to-asset, debt-to-equity, and book-to-market ratios. The foreign PDs have higher debt-to-equity and debt-to-asset in the second period, while in the first period the two ratios are nearly identical across foreign and domestic PDs. However, for the entire period, the debt-to-equity of foreign PDs is twice as large, and the debt-to-asset is about 5% higher than the one of the domestic PDs. Overall, foreign PDs are leveraging more their equity and assets. Furthermore, the domestic PDs have a slightly higher book-to-market ratio compared to the foreign counterparts.

We also briefly investigate how special the domestic PDs are compared to the financial sector in the US (i.e., SIC codes 6000 to 6799). Panel C of Table 3.3 reports the same statistics as in Panel A of Table 3.3 for the financial sector by excluding the parent companies of the NY Fed's PDs. By comparing Panels A and C of Table 3.3, it emerges that on average the domestic PDs are larger than the other financial firms. Finally, Panel D of Table 3.3 displays various average characteristic ratios between the domestic PDs and the financial sector in the US. On average, for all the four statistics that we consider and over the entire period 1970-2020, the domestic PDs have characteristic ratios that are 50 to 65 times larger than those of the average US financial sector.

To summarise, our sample construction delivers vastly different series relative to HKM. We suspect that HKM performed a heavy and ad-hoc trimming of the list of ultimate parent companies to obtain their ratio and risk factor series. Based on our sample construction, foreign PDs are slightly more important than the domestic ones in terms of number and concentration of their ultimate parent companies. Moreover, the foreign PDs seems to be riskier than their domestic counterparts and their stock seems to be slightly more overvalued. Finally, the domestic PDs are much bigger than the average financial institution in the US.

3.3 Empirical Strategy

Our empirical analysis is divided into three parts. First, given the difference between our series and those proposed by HKM[4], we replicate HKM's main cross-sectional analysis using their empirical methods and the newly proposed methodological insights of Gospodinov and Robotti (2021). Second, we investigate market reactions around the NY Fed's announcement of a new PD. Finally, we analyse the performance of the domestic NY Fed's PDs relative to the US broker-dealer sector.

Following HKM, the aggregate capital ratio is defined as

$$\eta_t = \frac{\sum_{i=1}^{n} MktCap_{i,t}}{\sum_{i=1}^{n} (MktCap_{i,t} + BookDebt_{i,t})}, \tag{3.1}$$

where n is the total number of domestic and foreign ultimate parent companies, t denotes calendar time (end of quarter), $MktCap_{i,t}$ is the closing price in the current calendar quarter multiplied by the most recent number of shares outstanding, and $BookDebt_{i,t}$ are the most recent total assets minus the most recent common equity available at the end of each calendar quarter. Over periods of distress, the market value of a company's stock decreases and its book debt increases, thus leading to a contraction in the capital ratio. Equivalently, a procyclical capital ratio translates into a countercyclical intermediary leverage, which has been widely documented in the literature. Then, HKM's nontraded capital risk factor is computed as

$$\eta_t^{\Delta} = \frac{\epsilon_t}{\eta_{t-1}}, \tag{3.2}$$

where ϵ_t is the residual from the following AR(1) specification for η :

$$\eta_t = \rho_0 + \rho\eta_{t-1} + \epsilon_t. \tag{3.3}$$

In the following analysis, we denote η_t^{Δ} by $CPTL$. HKM also propose a traded intermediary capital factor that is computed as

$$r_{p,t} = \sum_{i=1}^{n} MktCap_{i,t} * r_{i,t}, \tag{3.4}$$

where $r_{i,t}$ is the return of PD i at the end of the current calendar quarter t, and $r_{p,t}$

[4]See Appendix C.

represents the value-weighted portfolio return. This traded factor is denoted by $CPTLT$ in the following analysis. We then apply the traditional two-pass methodology of Fama and MacBeth (1973) to investigate the robustness of these nontraded and traded capital factors in cross-sectional asset pricing. In particular, we focus on the price of multivariate beta risk and on the price of covariance risk in the subsequent empirical analysis. We refer the readers to Gospodinov and Robotti (2021) for the relevant methodological details.

In the second part of our empirical analysis, we will investigate how investors react to the NY Fed's announcement of a new PD. If PDs are special entities, investors should closely monitor the Fed's announcement of a new PD, and we should observe a price reaction around the announcement day. To fulfil this task, we rely on the methodology proposed by Patton and Verardo (2012), and we analyse the stock price reaction of the ultimate parent companies five days around the announcement day. We consider the following specification:

$$\begin{aligned} r_{i,t} = {} & \delta_{-5} I_{i,t+5} + \cdots + \delta_0 I_{i,t} + \cdots + \delta_5 I_{i,t-5} \\ & + \bar{\beta}_{i,1} D_{i,1} + \bar{\beta}_{i,2} D_{i,2} + \bar{\beta}_{i,3} D_{i,3} + \gamma' \mathbf{X}_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{3.5}$$

where $r_{i,t}$ is the time t dollar return of company i that is about to become a PD on the event day. $I_{i,t}$ are dummy variables defined over a 10-day event window around the day that a broker-dealer becomes a PD. Specifically, $I_{i,t} = 1$ if day t is an announcement day for company i, and $I_{i,t} = 0$ otherwise. Moreover, $D_{i,1}$, $D_{i,2}$, and $D_{i,3}$ are, in order, firm, year, and industry fixed effects. We also include a series of control variables in our specification, through $\mathbf{X}_{i,t}$. We control for various firm-specific characteristics: the first lag of the returns (to account for the mild autocorrelation in equity returns), market capitalisation, book-to-market ratio, assets, and book equity (to account for company-specific characteristics).[5] To detect whether the PD announcements lead to a statistically significant change in returns over the estimation window, we analyze the size and significance of the δ_j coefficient estimates for each day in the estimation window (δ_{-5} refers to the first day, and δ_5 to the last day in the window). In brief, δ_j captures the daily deviations of returns from their long-run average. If investors interpret news positively, then the δ estimates will be positive and statistically significant. In contrast, if investors perceive the news as bad, the estimates will be negative and statistically significant. If the estimates are not statistically significant,

[5]The standard error are checked with the Arellano and Bond (1991) methodology, to account for the feedback of the lagged variable in the dependent variable, when controlling for the first lag of the returns.

the obvious interpretation is that prices do not deviate from their long-run average and investors do not react to this new information. Furthermore, by looking at the timing, it is possible to understand whether investors can predict the PD events and/or have a delayed reaction to the PD announcements. We cluster the standard errors by firms and days, as suggested by Petersen (2009). Such a procedure yields standard errors that are robust to heteroskedasticity and arbitrary within-cluster correlation.

Finally, we investigate whether New York Fed's PDs are special relative to non-primary dealers, as represented by the broker-dealer sector as a whole. We employ a diff-in-diff approach and only consider domestic parent companies in the analysis (due to data limitations in the identification of a control group). We run the following regression model for each of the ultimate parent companies of the various PDs (treatment group):

$$r_{i,t} = \alpha + \beta PD_i + \zeta Time_t + \psi PD_i * Time_t + \epsilon_{it}, \tag{3.6}$$

by using an equally-weighted portfolio return of all of the US broker-dealers (SIC codes 6211 and 6221) as our control group. To avoid that a company is contemporaneously in the treatment and control group, we exclude broker-dealers while they act as New York Fed's PDs in the computation of the equally-weighted portfolio. In the above specification, $r_{i,t}$ represents the daily returns of the control and treatment groups one year before and one year after a financial intermediary becomes a PD. The dummy variable PD_i takes on the value of one if i is a PD in the treatment group and of zero if i represents the control group. Moreover, $Time_t$ is another dummy variable that takes the value of one after a company becomes a PD and zero before the event (this dummy variable does not change across treatment and control groups). Finally, $PD_i * Time_t$ is the interaction term between the two indicators. The intuition behind this formulation is that if becoming a PD leads to any performance improvement (deterioration), then the coefficient estimate associated with the interaction term should be positive (negative) and statistically significant.

3.4 Preliminary Insights on the New York Fed's Primary Dealers

In this section, we provide some preliminary insights on the NY Fed's PDs and report some descriptive statistics of the series used in the empirical analysis. Figure 3.3 plots the capital ratio as defined in Equation (3.1) (at a quarterly frequency) along with the National Bureau of Economic Research (NBER) recession periods.

The figure shows that the aggregate capital ratio decreases and reaches its troughs around periods of market distress. Furthermore, around 1985 the capital ratio exhibits a threefold increase and then shrinks again towards the average starting from the early 2000s. These two shifts are merely driven by the inclusion and exclusion of Japanese and industrial conglomerate PDs. Figure 3.4 plots $CPTLT$, the value-weighted portfolio return described in Equation (3.4). Again, portfolio returns are lower around recessions and tend to increase afterwards.

In a cross-sectional setting, HKM show that $CPTLT$ is a priced source of risk in a wide range of asset classes.[6] In addition, Gospodinov and Robotti (2021) find that the correlation between HKM's traded capital factor and the excess value-weighted return on the US stock market is about 0.84, with $CPTLT$ having approximately the same mean as the market factor and a substantially higher standard deviation. We repeat the same exercises here using our newly created sample. Panel A of Table 3.4 leads to similar conclusions.

The correlation between the market and $CPTLT$ is about 0.69, while the $CPTLT$ factor commands a statistically insignificant risk premium. Panels B and C of Table 3.4 also confirm the results of Gospodinov and Robotti (2021) when it comes to the Sharpe ratio performance of the two-factor (HKM) and single-factor (HKMSF) models of HKM. Neither of these models is superior to the CAPM based on the squared Sharpe ratio comparisons in the table.[7]

To shed further light on $CPTLT$, we compare $CPTLT$ with IXG. IXG is an iShares ETF that invests in the financial sector worldwide.[8] The IXG website provides a list of

[6] We refer the readers to their Table 17 for further details.

[7] See Barillas et al. (2020) for an explanation of the nested and non-nested model comparison tests.

[8] Details on the IXG ETF can be found at the following link: `https://www.ishares.com/us/products/239742/ishares-global-financials-etf`.

the constituents and their geographical locations. Interestingly, all of the NY Fed's PDs are also part of this ETF. So, if PDs are special and unique relative to the non-primary broker-dealer sector, they should be even more unique when comparing them with the whole financial sector worldwide. This is not what we find. Figure 3.5 plots $CPTLT$ and IXG from 2007:Q2 to 2019:Q3. The two series seem to almost perfectly match each other, something that is also reflected in the very high correlation of the two series (0.89). $CPTLT$ appears to be only slightly more volatile than IXG (0.17 vs. 0.13).

Next, we analyze whether becoming a NY Fed's PD entails any substantial changes in volatility and cumulative returns for the company. Towards this end, we measure the average monthly realised volatilities and cumulative returns of the ultimate parent companies from 12 months before until 12 months after they become PDs. Changing name or merging with other companies do not represent an event in our analysis. Events occur only when the NY Fed announces that a broker-dealer has become a PD, and the ultimate parent was not already identified as a holding company of a PD. Furthermore, we require that the ultimate parent that is about to become the controlling company of a PD has one year of observations before the event date. This leaves us with a total of 74 companies in the sample. Figure 3.6 shows the monthly average realised volatility (RV) across all PDs over a two-year window.

RV is calculated by summing the daily squared returns in each calendar month for each PD. The zero value on the x axis denotes the month in which the company becomes a PD. The dashed lines denote the 95% confidence bounds for average RV. The figure highlights a slightly higher RV before the event date. The average RV is 0.012 before the event date and 0.010 after. However, bootstrap experiments reveal that this difference is not statistically significant.

Figure 3.7 plots the cumulative returns 12 months before until 12 months after a company becomes a PD. The cumulative return reflects an initial investment of 100$ at t-13 for each distinct PD that is then compounded over time using monthly returns. At each time period t, we calculate the cross-sectional average across all PDs and the 95% confidence bounds. The plot displays a slight increase in cumulative returns from the period before to the period after the month of the event. The average cumulative return is around \$101 at the month of the event and about \$118 at the end of the twelve months after the event date. Moreover, the average difference between before and after the event is approximately \$12. The larger confidence bounds after the event date indicate that this difference is highly volatile across PDs and suggest that becoming a PD does not lead to any considerable performance improvement. Such conclusion is also supported by bootstrap tests on the difference in average cumulative returns before and after the event takes place. In conclusion, this preliminary evidence all points to the same direction. NY Fed's PDs do not appear to be special in the sense that the performance of these companies before and after becoming PDs is not substantially different, neither from a risk nor from a return perspective.

3.5 Main Results

In this section, we present the main results of our empirical analysis. In Subsection 3.5.1, we analyse the cross-sectional performance of the nontraded capital risk factor at a quarterly frequency by employing HKM's methods.[9] Subsection 3.5.2 presents the results of an event study around the announcement of a new PD to determine whether there is a positive announcement effect on the company's stock from becoming PDs. Finally, Subsection 3.5.3 displays the results of a diff-in-diff exercise to evaluate whether PDs are special entities relative to the broker-dealer sector. In the last two subsections, the event of a company becoming a PD is defined only when the ultimate parent company is not already a PD in the previous period. This implies that we do not consider mergers and acquisitions and name changes as events.

3.5.1 Cross-Sectional Pricing

This subsection reports cross-sectional asset-pricing results for HKM's two-factor model with market and nontraded capital (Table 3.5) as well as their corresponding single-factor specifications (Table 3.6). Below each risk premium estimate, in round brackets we report the t-statistic under correctly specified models (t-stat$_c$).[10] To be consistent with HKM, we do not adjust for serial correlation in the computation of the t-statistics. Doing so would render the standard errors of the estimates even larger. We also include the ordinary least

[9] The results based on the traded capital risk factor and a monthly frequency are qualitatively similar and available on request.

[10] The t-statistics under correctly specified models are the standard generalized method of moments (GMM) t-statistics under conditional heteroskedasticity. (See also Jagannathan and Wang (1998).)

squares (OLS) cross-sectional regression (CSR) R^2 in the table.

In Panel A of Table 3.5, we report OLS cross-sectional asset-pricing tests for HKM's two-factor model. Based on t-stat$_c$ and a 5% significance level of the test, Panel A of Table 3.5 shows that we cannot reject the null of a zero risk premium for HKM's capital risk factor in all seven asset classes. Moreover, when considering an unbalanced panel of all asset excess returns (the All column), the overall capital risk premium is 10.30% per quarter with a t-statistic of 1.23. Consistent with the findings of Lewellen et al. (2010) and Kleibergen and Zhan (2015), the OLS CSR R^2s are found to be unrealistically large for many asset classes. In addition, the estimated prices of multivariate beta risk for capital are not very sensible. For example, for sovereign bonds and credit default swaps, the capital risk premium flips sign and becomes negative. In contrast market risk appears to be priced for US bonds, options, credit default swaps, and foreign exchange.

As pointed out by Gospodinov and Robotti (2021), it is incorrect to focus on the price of multivariate beta risk if the factors are correlated and the goal is to determine whether a given factor adds to the explanatory power of the model. In our setting, the correlation between the market factor and *CPTL* is about 55%, and therefore we also compute the prices of covariance risk and report them in Panel B of Table 3.5. Panel B leads to the same conclusions as for Panel A. These results strongly suggest that the PD factor is not priced across a wide range of alternative asset classes.

In Table 3.6, we also consider two single-factor specifications: the CAPM and HKM's single-factor model (HKMSF). Since we are now analyzing single-factor models, focusing on betas or covariances is equivalent, and we choose to report results for the price of beta risk. Comparing Panels A and B again indicates that the market factor works much better than intermediary capital in pricing test assets that are notoriously difficult to explain. Overall, based on HKM's empirical methods, this exercise highlights the poor performance of intermediary capital in cross-sectional asset pricing. In contrast, market risk appears to be a much better candidate risk factor when interest lies in explaining these cross-sections of asset returns.

3.5.2 Event Study Analysis

In this subsection, we conduct an event study analysis at a daily frequency for the domestic (USA), foreign (For), and whole (All) samples. Splitting the sample according to the geographic location of the ultimate parent company allows us to investigate whether there exists a difference in price reaction across PDs that have headquarters in the US and abroad. Moreover, the full-sample analysis allows us to draw more general conclusions about the investors' perception of a PD announcement by looking at the price changes in the reported event window.

Table 3.7 displays our findings. "Day" in the table indicates the number of days before and after announcement day, where "Day 0" is the event day. Starting from the full-sample results, the only statistically significant coefficient estimate can be found five days after the announcement day. The estimate is negative and significant at the 99% significance level, which implies that the market reacts only five days after the PD announcements. The coefficient estimate is negative (-0.23) and implies that the prices of the ultimate parent companies' stocks decrease by 0.23% five days after the announcement date. These results are based on 580,749 daily observations and the adjusted R^2 of the regression is 0.14. For the US sub-sample, prices significantly deviate from their long-run average two days before the announcement date, on the announcement day, and five days after the announcement day. In all of the cases, the coefficient estimates are negative and statistically significant at least at a 95% significance level. This model is estimated based on 224,092 daily observations and delivers an adjusted R^2 of 0.20. Finally, the foreign sample only delivers two statistically significant coefficient estimates; a positive one five days before the announcement day that is statistically significant at the 90% level, and a negative one two days after the event that is statistically significant at the 99% level. The latter model is estimated based on 356,657 daily observations and entails an adjusted R^2 of 0.003.

Overall, these results suggest that investors negatively perceive the NY Fed announcement of a new PD. Even though the foreign sample contains more observations than the US sample, investors seem to pay more attention to the US companies than to the foreign ones. This conclusion is based on the number of observations, statistical significance and size of the coefficient estimates, and the adjusted R^2 measure. However, across the two sub-samples, we do not observe consistent and reliable patterns for price changes. This might indicate that the inference and conclusions presented above are unreliable. In fact, when adding a few control variables (lagged returns, market capitalisation, book-to-market ratio, assets, and book equity) through the $\mathbf{X}_{i,t}$ term in Equation (3.5), untabulated results (available on request) indicate that none of the coefficient estimates are statistically significant any longer. Stated differently, the announcement of a new NY Fed's PD does not seem to generate any relevant price reaction.

3.5.3 Performance

In this subsection, we report results for the diff-in-diff specification discussed earlier. The treatment group includes all of the US PDs, while the control group is represented by an equally-weighted portfolio where the constituent companies are the US broker-dealers (SIC codes 6211 and 6221).

Figure 3.8 plots the results of the analysis. The two horizontal dashed lines denote statistical significance at the 95% confidence level. The black dots indicate the size of the coefficient estimates (multiplied by 100), and the red bars represent the t-statistics associated with the parameter estimates. Across the 29 companies in the sample, only one coefficient estimate is statistically significant at the 95% level. The sign of this estimate is negative (-0.001) and suggests that the performance of the company deteriorates compared to the performance of the treatment group after the PD announcement is made. The remaining 28 coefficient estimates are not statistically significant even at the 90% significance level. Overall, our results suggest that the NY Fed's PDs are not special relative to the other US non-primary broker-dealers.

3.6 Concluding Remarks

A common thread in the recent intermediary asset pricing literature is the relentless search for risk factors with robust pricing performance across a wide variety of asset classes. HKM claim that robust pricing can be achieved if we are willing to consider traded and nontraded capital risk factors based on the stock and debt of domestic and foreign ultimate parent companies whose subsidiaries are NY Fed's primary dealers. They identify the marginal investors in many asset classes as the NY Fed's primary dealers and claim that they are special relative to the broker-dealer sector.

We re-examine HKM's empirical evidence by first forming our own capital ratio and risk factors. Our series are very different from those that HKM made publicly available, especially for the middle part of the sample. Based on our newly built data, we document a total lack of pricing performance for the intermediary capital factors even when using HKM's empirical methodology. In addition, we employ an event-study analysis to show that becoming primary dealers of the NY Fed does not generate any relevant price effect for those companies. Therefore, while it is certainly prestigious to become a NY Fed's primary dealer, the market does not seem to reward this type of announcement. Finally, we rely on a difference-in-difference approach to conclude that the domestic NY Fed's primary dealers are not special relative to the broker-dealers in the US.

Table 3.1: List of Primary Dealers (Domestic Parent Companies)

The table shows the ultimate publicly-listed domestic holding company (Parent Company) for each NY Fed's primary dealer over time. We obtain the list of primary dealers from the NY Fed's website for HKM's 1970-2014 sample period, and we employ the information available on the same website to update the list until 2020. We include the name of the primary dealer and its ultimate parent company along with its global company key (GVKEY), as reported by Compustat. The date range highlights the period in which a company was a NY Fed's primary dealer.

Primary Dealer	From	Thru	Parent Company	Gvkey
BA SECURITIES, INC.	18/04/1994	30/09/1997	BANKAMERICA CORP-OLD	2024
BANC OF AMERICA SECURITIES LLC	17/05/1999	01/11/2010	BANK OF AMERICA CORP	7647
BANC ONE CAPITAL MARKETS, INC	01/04/1999	01/08/2004	BANK ONE CORP	1998
BANCAMERICA ROBERTSON STEPHEN	01/10/1997	31/08/1998	BANKAMERICA CORP-OLD	2024
BANCAMERICA SECURITIES, INC.	01/09/1998	30/09/1998	BANKAMERICA CORP-OLD	2024
BANK OF AMERICA NT & SA	17/11/1971	15/04/1994	BANKAMERICA CORP-OLD	2024
BANKERS TRUST	19/05/1960	07/07/1989	BANKERS TRUST CORP	2029
BEAR,STEARNS & CO., INC.	10/06/1981	01/10/2008	BEAR STEARNS COMPANIES INC	11818
BLYTH EASTMAN DILLON CAPITAL MARKETS	05/12/1974	31/12/1979	INA CORP	5843
BNY SECURITIES, INC.	01/08/1989	09/08/1990	BANK OF NEW YORK MELLON CORP	2019
BofA SECURITIES	13/05/2019	Current Dealer	BANK OF AMERICA CORP	7647
BT ALEX. BROWN INCORPORATED	23/10/1997	04/06/1999	BANKERS TRUST CORP	2029
BT SECURITIES CORPORATION	10/07/1989	22/10/1997	BANKERS TRUST CORP	2029
CHASE MANHATTAN CAPITAL MARKETS CORP	01/07/1987	19/12/1988	CHASE MANHATTAN CORP -OLD	2943
CHASE MANHATTAN GOV'T SECURITIES	15/07/1970	30/06/1987	CHASE MANHATTAN CORP -OLD	2943
CHASE SECURITIES, INC	01/04/1996	30/04/2001	JPMORGAN CHASE & CO	2968
CHASE SECURITIES, INC	20/12/1988	31/03/1996	CHASE MANHATTAN CORP -OLD	2943
CHEMICAL	19/05/1960	31/03/1989	JPMORGAN CHASE & CO	2968
CHEMICAL SECURITIES INC	01/01/1992	31/03/1996	JPMORGAN CHASE & CO	2968
CHEMICAL SECURITIES, INC.	01/04/1989	31/12/1991	JPMORGAN CHASE & CO	2968
CITIBANK	15/06/1961	13/04/1989	CITICORP	3066
CITICORP SECURITIES MARKETS, INC.	14/04/1989	14/07/1993	CITICORP	3066
CITICORP SECURITIES, INC.	15/07/1993	30/11/1998	CITICORP	3066
CITIGROUP GLOBAL MARKETS INC.	07/04/2003	Current Dealer	CITIGROUP INC	3243
CONTINENTAL BANK, NATIONAL ASSOC.	15/12/1988	30/08/1991	CONTINENTAL BANK CORP	3463
CONTINENTAL ILL.	19/05/1960	14/12/1988	CONTINENTAL BANK CORP	3463
COUNTRYWIDE SECURITES CORPORATION	15/01/2004	15/07/2008	COUNTRYWIDE FINANCIAL CORP	3555
DEAN WITTER REYNOLDS INC.	02/11/1977	30/04/1998	DEAN WITTER REYNOLDS ORG INC	3823
DEAN WITTER REYNOLDS INC.	02/11/1977	30/04/1998	SEARS HOLDINGS CORP	6307
DEAN WITTER REYNOLDS INC.	02/11/1977	30/04/1998	DEAN WITTER DISCOVER & CO	27867
DILLON, READ & CO., INC.	24/06/1988	02/09/1997	TRAVELERS CORP	10705
DISCOUNT CORPORATION OF NEW YORK	19/05/1960	10/08/1993	DISCOUNT CORP NY/DEL	3689
DLJ SECURITIES CORPORATION	06/03/1974	31/12/2000	DONALDSON LUFKIN & JENRETTE	4037
FIRST CHICAGO	19/05/1960	01/01/1990	FIRST CHICAGO CORP	4689
FIRST CHICAGO CAPITAL MARKETS	02/01/1990	31/03/1999	FIRST CHICAGO NBD CORP	7650
FIRST INTERSTATE	31/07/1964	31/10/1986	FIRST INTERSTATE BNCP	4710
FIRST INTERSTATE CAPITAL MARKETS,INC	03/11/1986	17/06/1988	FIRST INTERSTATE BNCP	4710
FIRST N/B OF BOSTON	21/03/1983	17/11/1985	BANKBOSTON CORP	2014
FIRST PENNCO SEC. INC.	07/03/1974	27/08/1980	FIRST PENNSYLVANIA CORP	4732
GOLDMAN, SACHS & CO.	04/12/1974	Current Dealer	GOLDMAN SACHS GROUP INC	114628
HARRIS TRUST	15/07/1965	31/08/1988	HARRIS BANKCORP INC	5491
HUTTON	02/11/1977	31/12/1987	HUTTON (E.F.) GROUP	5793
IRVING SECURITIES, INC.	19/05/1960	31/07/1989	IRVING BANK CORP	6186
J.P. MORGAN SECURITIES INC.	01/05/2001	01/09/2010	JPMORGAN CHASE & CO	2968
J.P. MORGAN SECURITIES LLC	01/09/2010	Current Dealer	JPMORGAN CHASE & CO	2968
J.P.MORGAN SECURITIES,INC.	19/05/1960	30/04/2001	MORGAN (J P) & CO	7562
JEFFERIES & COMPANY, INC.	18/06/2009	01/03/2013	JEFFERIES FINANCIAL GRP INC	6682
JEFFERIES LLC	01/03/2013	Current Dealer	JEFFERIES FINANCIAL GRP INC	6682
KIDDER, PEABODY & CO., INCORPORATED	07/02/1979	30/12/1994	GENERAL ELECTRIC	5047
L.F.ROTHSCHILD & CO., INC.	18/05/1987	17/01/1989	ROTHSCHILD (LF) HOLDINGS INC	12223
L.F.ROTHSCHILD,UNTERBERG,TOWBIN	11/12/1986	15/05/1987	ROTHSCHILD (LF) HOLDINGS INC	12223
LEHMAN	25/11/1976	31/12/1987	AMERICAN EXPRESS CO	1447
LEHMAN BROTHERS INC.	31/08/1995	22/09/2008	LEHMAN BROTHERS HOLDINGS INC	30128
LEHMAN GOVERNMENT SECURITIES INC	01/08/1990	30/08/1995	AMERICAN EXPRESS CO	1447
MANUFACTURERS HANOVER	31/08/1983	29/07/1988	MANUFACTURERS HANOVER CORP	7003
MANUFACTURERS HANOVER SECURITIES COR	01/08/1988	31/12/1991	MANUFACTURERS HANOVER CORP	7003
MERRILL LYNCH GOVERNMENT SEC. INC.	19/05/1960	11/02/2009	MERRILL LYNCH & CO INC	7267
MERRILL LYNCH, PIERCE, FENNER & SMITH INCORPORATED	01/11/2010	13/05/2019	BANK OF AMERICA CORP	7647
MF GLOBAL	02/02/2011	31/10/2011	MF GLOBAL HOLDINGS LTD	177745
MORGAN STANLEY & CO. INCORPORATED	01/02/1978	31/05/2011	MORGAN STANLEY	12124
MORGAN STANLEY & CO. LLC	31/05/2011	Current Dealer	MORGAN STANLEY	12124
NATIONSBANC CAPITAL MARKETS, INC.	01/10/1993	30/09/1997	BANK OF AMERICA CORP	7647
NATIONSBANC MONTGOMERY SECURITIES, INC	01/10/1997	30/09/1998	BANK OF AMERICA CORP	7647
NATIONSBANC MONTGOMERY SECURITIES, LLC	01/10/1998	16/05/1999	BANK OF AMERICA CORP	7647
NATIONSBANK OF NORTH CAROLINA, N.A.	06/07/1993	30/09/1993	BANK OF AMERICA CORP	7647
NORTHERN TRUST	08/08/1973	29/05/1986	NORTHERN TRUST CORP	7982
NUVEEN GOV'T SEC. INC.	18/11/1971	27/08/1980	ST PAUL COS	9380
PAINE WEBBER INCORPORATED	25/11/1976	04/12/2000	PAINE WEBBER GROUP	8299
PAINE, WEBBER, JACKSON & CURTIS INC.	22/06/1972	27/06/1973	PAINE WEBBER GROUP	8299
PRUDENTIAL SECURITIES INCORPO	25/02/1991	01/12/2000	PRUDENTIAL FINANCIAL INC	143356
PRUDENTIAL-BACHE	29/10/1975	24/02/1991	BACHE GROUP INC	1967
SALOMON SMITH BARNEY INC.	01/09/1998	06/04/2003	CITIGROUP INC	3243
SECURITY PACIFIC NATIONAL BANK	11/12/1986	17/01/1991	SECURITY PACIFIC CORP	9577
SHEARSON LEHMAN	01/01/1988	31/07/1990	AMERICAN EXPRESS CO	1447
SMITH BARNEY INC.	01/06/1994	31/08/1998	CITIGROUP INC	3243
SMITH BARNEY SHEARSON INC.	02/08/1993	31/05/1994	CITIGROUP INC	3243
SMITH BARNEY, HARRIS UPHAM & CO.,INC	22/08/1979	01/08/1993	CITIGROUP INC	3243
THE FIRST BOSTON CORPORATION	19/05/1960	11/10/1993	FIRST BOSTON INC	4684
WELLS FARGO SECURITIES, LLC	18/04/2016	Current Dealer	WELLS FARGO & CO	8007
ZIONS FIRST NATIONAL BANK	11/08/1993	31/03/2002	ZIONS BANCORPORATION NA	11687

Table 3.2: List of Primary Dealers (Foreign Parent Companies)

The table shows the ultimate publicly-listed foreign holding company (Parent Company) for each NY Fed's primary dealer over time. We obtain the list of primary dealers from the NY Fed's website for HKM's 1970-2014 sample period, and we employ the information available on the same website to update the list until 2020. We include the name of the primary dealer and its ultimate parent company along with its unique mnemonic code (MNEM), as reported by Datastream. The date range highlights the period in which a company was a NY Fed's primary dealer. Finally, we report the currency in which each company's stock is traded (Country column).

Primary Dealer	From	Thru	Parent Company	Country	Mnem
ABN AMRO BANK, N.V., NY BR	09/12/2002	15/09/2006	ABN AMRO HOLDING	NLD	H:AAB
ABN AMRO INCORPORATED	29/09/1998	08/12/2002	ABN AMRO HOLDING	NLD	H:AAB
AUBREY G. LANSTON & CO., INC.	19/05/1960	17/04/2000	INDUSTRIAL BANK OF JAPAN LTD	JPN	J:IK@N
BANK OF NOVA SCOTIA, NEW YORK AGENCY	04/10/2011	Current Dealer	BANK OF NOVA SCOTIA	CAN	C:BNS
BARCLAYS CAPITAL INC.	01/04/1998	Current Dealer	BARCLAYS	GBR	BARC
BARCLAYS DE ZOETE WEDD GSI	07/12/1989	01/03/1990	BARCLAYS	GBR	BARC
BARCLAYS DE ZOETE WEDD SECURITIES IN	02/03/1990	30/06/1996	BARCLAYS	GBR	BARC
BMO CAPITAL MARKETS CORP.	04/10/2011	Current Dealer	BANK OF MONTREAL	CAN	C:BMO
BMO NESBITT BURNS CORP.	15/02/2000	31/03/2002	BANK OF MONTREAL	CAN	C:BMO
BNP PARIBAS SECURITIES CORP.	15/09/2000	Current Dealer	BNP PARIBAS	FRA	F:BNP
BZW SECURITIES INC.	01/07/1996	31/03/1998	BARCLAYS	GBR	BARC
CARROLL MCENTEE & MCGINLEY INC.	29/09/1976	06/05/1994	HSBC HLDGS PLC	HK	K:HSBC
CIBC OPPENHEIMER CORP.	04/12/1997	02/05/1999	CANADIAN IMPERIAL BANK	CAN	C:CM
CIBC WOOD GUNDY SECURITIES CO	27/03/1996	03/12/1997	CANADIAN IMPERIAL BANK	CAN	C:CM
CIBC WORLD MARKETS CORP.	03/05/1999	08/02/2007	CANADIAN IMPERIAL BANK	CAN	C:CM
COUNTY NATWEST GOV. SEC., INC.	29/09/1988	13/01/1989	NATL WESTMINSTER BANK	GBR	NWB
CREDIT SUISSE 1ST BOSTON LLC	17/01/2003	16/01/2006	CREDIT SUISSE GROUP	CHE	S:CSGN
CREDIT SUISSE AG, NEW YORK BRANCH	13/11/2017	Current Dealer	CREDIT SUISSE GROUP	CHE	S:CSGN
CREDIT SUISSE FIRST BOSTON CO	16/12/1996	16/01/2003	CREDIT SUISSE GROUP	CHE	S:CSGN
CREDIT SUISSE SECURITIES (USA) LLC	16/01/2006	13/11/2017	CREDIT SUISSE GROUP	CHE	S:CSGN
CS FIRST BOSTON CORPORATION	12/10/1993	15/12/1996	CREDIT SUISSE GROUP	CHE	S:CSGN
DAIWA CAPITAL MARKETS AMERICA INC.	01/04/2010	Current Dealer	DAIWA SECURITIES GROUP INC	JPN	J:DS@N
DAIWA SECURITIES AMERICA INC.	11/12/1986	01/04/2010	DAIWA SECURITIES GROUP INC	JPN	J:DS@N
DEUTSCH BANC ALEX. BROWN INC.	12/01/2001	29/03/2002	DEUTSCHE BANK	DEU	D:DBK
DEUTSCHE BANK GSI	13/12/1990	30/09/1993	DEUTSCHE BANK	DEU	D:DBK
DEUTSCHE BANK SECURITIES CORPORATION	01/10/1993	31/10/1995	DEUTSCHE BANK	DEU	D:DBK
DEUTSCHE BANK SECURITIES INC.	01/06/1998	11/01/2001	DEUTSCHE BANK	DEU	D:DBK
DEUTSCHE BANK SECURITIES INC.	30/03/2002	Current Dealer	DEUTSCHE BANK	DEU	D:DBK
DEUTSCHE MORGAN GRENFELL/C.J.	01/11/1995	29/05/1998	DEUTSCHE BANK	DEU	D:DBK
DLJ SECURITIES CORPORATION	06/03/1974	31/12/2000	AXA SA	FRA	F:MIDI
DRESDNER KLEINWORT BENSON NOR	08/05/1997	29/04/2001	DRESDNER BANK	DEU	D:DRB
DRESDNER KLEINWORT SECURITIES LLC	18/09/2006	26/06/2009	ALLIANZ	DEU	D:ALV
DRESDNER KLEINWORT WASSERSTEIN SECURITIES LLC	30/04/2001	18/09/2006	ALLIANZ	DEU	D:ALV
EASTBRIDGE CAPITAL INC.	18/06/1992	29/05/1998	NIPPON CREDIT BANK LTD	JPN	J:NPCB
FUJI SECURITIES INC.	28/12/1989	31/03/2002	FUJI BANK LTD	JPN	J:FB@N
GREENWICH CAPITAL MARKETS, INC.	31/07/1984	01/04/2009	NATL WESTMINSTER BANK	GBR	NWB
GREENWICH CAPITAL MARKETS, INC.	31/07/1984	01/04/2009	ROYAL BANK OF SCOTLAND GROUP	GBR	RBS
GREENWICH CAPITAL MARKETS, INC.	31/07/1984	01/04/2009	LONG TERM CREDIT BANK OF JAPAN	JPN	J:LTCR
HARRIS GOVERNMENT SECURITIES	01/09/1988	30/12/1992	BANK OF MONTREAL	CAN	C:BMO
HARRIS NESBITT THOMSON SEC. INC.	08/09/1993	31/05/1995	BANK OF MONTREAL	CAN	C:BMO
HARRIS-NESBITT THOMSON SEC., INC.	31/12/1992	07/09/1993	BANK OF MONTREAL	CAN	C:BMO
HSBC SECURITIES (USA) INC.	01/06/1999	Current Dealer	HSBC HLDGS PLC	HK	K:HSBC
HSBC SECURITIES, INC.	09/05/1994	31/05/1999	HSBC HLDGS PLC	HK	K:HSBC
KLEINWORT BENSON GOV'T SEC., INC.	13/02/1980	27/12/1989	KLEINWORT BENSON GROUP PLC	GBR	KBL
LLOYDS GOV'T SECURITIES, INC.	22/12/1987	28/04/1989	LLOYDS BANKING GROUP PLC	GBR	LLOY
MIDLAND-MONTAGU GOV. SEC.,INC.	13/08/1975	26/07/1990	MIDLAND BANK	GBR	MID
MIZUHO SECURITIES USA INC.	01/04/2002	Current Dealer	MIZUHO FINANCIAL GROUP INC	JPN	J:MIZH
NESBITT BURNS SECURITIES INC.	01/06/1995	14/02/2000	BANK OF MONTREAL	CAN	C:BMO
NOMURA SECURITIES INTERNATIONAL,INC	11/12/1986	30/11/2007	NOMURA HOLDINGS INC	JPN	J:NM@N
NOMURA SECURITIES INTERNATIONAL,INC	27/07/2009	Current Dealer	NOMURA HOLDINGS INC	JPN	J:NM@N
PARIBAS CORPORATION	01/05/1997	14/09/2000	BNP PARIBAS	FRA	F:BNP
RBC CAPITAL MARKETS CORPORATION	08/07/2009	01/11/2010	ROYAL BANK OF CANADA	CAN	C:RY
RBC CAPITAL MARKETS, LLC	01/11/2010	Current Dealer	ROYAL BANK OF CANADA	CAN	C:RY
RBS SECURITIES INC.	01/04/2009	Current Dealer	ROYAL BANK OF SCOTLAND GROUP	GBR	RBS
S.G. WARBURG & CO., INC.	24/06/1988	26/07/1995	WARBURG SG GROUP	GBR	WARB
SANWA SECURITIES (USA) CO., L	01/01/1994	20/07/1998	SANWA BANK LTD	JPN	J:SA@N
SANWA-BGK SECURITIES CO., L.P.	20/06/1988	31/12/1993	SANWA BANK LTD	JPN	J:SA@N
SBC CAPITAL MARKETS INC.	03/01/1995	02/06/1996	SWISS BANK CO	CHE	S:SBVN
SBC GOVERNMENT SECURITIES, INC.	29/03/1990	02/01/1995	SWISS BANK CO	CHE	S:SBVN
SBC WARBURG DILLON READ INC.	03/09/1997	28/06/1998	SWISS BANK CO	CHE	S:SBVN
SBC WARBURG INC.	03/06/1996	02/09/1997	SWISS BANK CO	CHE	S:SBVN
SG AMERICAS SECURITIES, LLC	02/02/2011	07/12/2015	SOCIETE GENERALE GROUP	FRA	F:SGE
SG COWEN SECURITIES CORP.	01/07/1999	31/10/2001	SOCIETE GENERALE GROUP	FRA	F:SGE
SOCIETE GENERALE, NEW YORK BRANCH	07/12/2015	Current Dealer	SOCIETE GENERALE GROUP	FRA	F:SGE
TD SECURITIES (USA) LLC	11/02/2014	Current Dealer	TORONTO DOMINION BANK	CAN	C:TD
THE NIKKO SECURITIES CO. INT'	22/12/1987	03/01/1999	NIKKO CORDIAL CORP	JPN	J:NK@N
UBS SECURITIES INC.	07/12/1989	29/02/1996	UBS GROUP AG	CHE	S:UBSG
UBS SECURITIES LLC	01/03/1996	28/06/1998	UBS GROUP AG	CHE	S:UBSG
UBS SECURITIES LLC.	09/06/2003	Current Dealer	UBS GROUP AG	CHE	S:UBSG
UBS WARBURG LLC.	01/05/2000	08/06/2003	UBS GROUP AG	CHE	S:UBSG
WARBURG DILLON READ LLC.	29/06/1998	28/04/2000	UBS GROUP AG	CHE	S:UBSG
WERTHEIM SCHRODER & CO., INC.	24/06/1988	08/11/1990	SCHRODERS PLC	GBP	SDR
WESTPAC POLLOCK GOV'T SECURITIES INC	04/02/1987	27/06/1990	WESTPAC BANKING	AUS	A:WBCX
YAMAICHI INT'L (AMERICA), INC.	29/09/1988	04/12/1997	YAMAICHI SECURITIES CO LTD	JPN	J:YG@N

Table 3.3: Primary Dealers (Quarterly Average Sample Statistics)

The table displays sample statistics for domestic (USA) and foreign (Foreign) ultimate parent companies of the NY Fed's primary dealers (USD in millions). The various statistics are calculated by taking the cross-sectional average of the characteristics of the parent companies of active primary dealers in each quarter and then by averaging over each sub-period (1970-1990, 1990-2020, and 1970-2020). The accounting data for the domestic sample is from CRSP and Compustat, while for the foreign sample the data is from Datastrem. Total in Panel A refers to the sample of foreign and domestic companies. The ratios in Panel B are calculated based on the same data as in Panel A. Panel C reports sample statistics for the entire US financial sector by excluding the domestic parent companies whose subsidiaries were operating as primary dealers during each sample period. Panel D employs simple ratios to compare the entire US financial sector with the domestic parent companies of the primary dealer subsidiaries reported in Panel A.

(a) Accounting Sample Statistics

	Asset			**Debt**			**Equity**			**Mkt. Cap.**		
	Total	USA	Foreign	Total	USA	Foreign	Total	USA	Foreign	Total	USA	Foreign
1970-1990	101.960	35.138	66.822	97.012	33.439	63.572	4.936	1.686	3.250	9.485	2.098	7.387
1990-2020	1,673.505	792.240	881.265	1,574.793	728.960	845.833	97.960	62.528	35.432	116.516	72.509	44.007
1970-2020	1,137.326	480.033	657.293	1,072.859	442.147	630.711	64.020	37.439	26.582	77.437	43.474	33.936

(b) Financial Ratios

	Debt-To-Asset		**Debt-To-Equity**		**Book-To-Market**	
	USA	Foreign	USA	Foreign	USA	Foreign
1970-1990	0.952	0.951	19.829	19.563	0.804	0.440
1990-2020	0.920	0.960	11.658	23.872	0.862	0.805
1970-2020	0.921	0.960	11.810	23.727	0.861	0.783

(c) US Financial Sector (Without Domestic Primary Dealers)

	Asset	**Debt**	**Equity**	**Mkt. Cap.**
1970-1990	2.772	2.562	0.195	0.209
1990-2020	12.110	10.793	0.831	1.316
1970-2020	8.279	7.417	0.570	0.862

(d) Ratio of Domestic Primary Dealers VS Financial Sector

	Asset	**Debt**	**Equity**	**Mkt. Cap.**
1970-1990	12.676	13.051	8.660	10.015
1990-2020	65.422	67.539	75.260	55.079
1970-2020	57.982	59.617	65.704	50.415

Table 3.4: Summary Statistics and Sharpe Ratio Analysis

Panel A reports factor means (Fac. Mean), standard deviations (Fac. SD), and correlation (Fac. Corr). In Panel B, we report bias-adjusted squared Sharpe ratios (Sh^2) for the CAPM, the two-factor model of HKM (HKM), and the single-factor model of HKM (HKMSF). Panel C is for differences in bias-adjusted sample squared Sharpe ratios between models. MKT and $CPTLT$ denote the market and traded capital factors, respectively. Panels A through C are based on $T = 172$ quarterly observations from 1970:Q1 to 2012:Q4. The p-values are in square brackets

(a) Summary Statistics

	MKT		**MRF**
Fac. Mean	0.015		0.013
	[0.030]		[0.279]
Fac. SD	0.091		0.155
Fac. Corr		0.691	

(b) Squared Sharpe Ratios

	CAPM	**MRF**	**MRFSF**
Sh^2	0.021	0.017	0.001
	[0.056]	[0.229]	[0.698]

(c) Squared Sharpe Ratio Comparisons

	MRF	**MRFSF**
$CAPM$	0.004	0.020
	[0.557]	[0.299]
MRF		0.016
		[0.053]

Table 3.5: OLS Cross-Sectional Asset-Pricing Tests by Asset Class

The table presents the OLS estimates of the prices of beta and covariance risks for HKM's two-factor model. *MKT* and *CPTL* denote the market and nontraded capital risk factors, respectively. *INT* is the cross-sectional intercept estimate. For each parameter estimate, we report the t-ratio under correctly specified models (t-stat$_c$) in round brackets. In addition, we present the sample OLS cross-sectional R^2 (R^2). The sample periods for equities (*FF25*), government and corporate bonds (*US bonds*), sovereign bonds (*Sov. bonds*), options (*Options*), credit default swaps (*CDS*), commodities (*Commod.*), and foreign exchange (*FX*) are 1970:Q1-2012:Q4, 1975:Q1-2011:Q4, 1995:Q1-2011:Q1, 1986:Q2-2011:Q4, 2001:Q2-2012:Q4, 1986:Q4-2012:Q4, and 1976:Q2-2009:Q4, respectively. N and T represent the number of assets and time-series observations, respectively.

(a) Price of beta risk

	FF25	US Bonds	Sov. Bonds	Option	CDS	Commod.	FX	All
CPTL	12.92	2.15	-35.74	8.59	-17.8	1.68	20.04	10.30
t-$stat_c$	(1.83)	(0.21)	(-0.58)	(1.29)	(-1.77)	(0.75)	(0.87)	(1.23)
MKT	1.19	3.42	9.92	8.20	6.39	-0.95	14.51	1.99
t-$stat_c$	(0.74)	(2.15)	(0.94)	(2.88)	(2.17)	(-0.56)	(2.59)	(1.16)
INT	0.59	0.40	-1.40	-5.74	-0.26	0.40	-2.03	-0.32
t-$stat_c$	(0.38)	(3.80)	(-0.32)	(-2.67)	(-2.07)	(0.66)	(-1.54)	(-0.62)
R^2	0.50	0.81	0.90	0.91	0.83	0.03	0.51	0.53
N	25	20	6	18	20	23	12	124
T	172	148	65	103	47	105	135	172

(b) Price of Covariance Risk

	FF25	US Bonds	Sov. Bonds	Option	CDS	Commod.	FX	All
CPTL	10.02	-0.08	-39.95	3.01	-22.1	1.63	8.14	7.40
t-$stat_c$	(1.81)	(-0.01)	(-0.66)	(0.70)	(-2.02)	(0.95)	(0.50)	(1.14)
MKT	-5.38	4.49	35.26	9.06	24.5	-2.05	15.21	-2.63
t-$stat_c$	(-1.68)	(0.60)	(0.72)	(1.91)	(2.88)	(-0.78)	(1.51)	(-0.57)

Table 3.6: OLS Cross-Sectional Asset-Pricing Tests by Asset Class (Price of Beta Risk in Single-Factor Models)

The table presents the OLS estimates of the prices of beta risks for the CAPM (Panel A) and HKM's single-factor model (HKMSF, Panel B). MKT and $CPTL$ denote the market and nontraded capital risk factors, respectively. INT is the cross-sectional intercept estimate. For each parameter estimate, we report the t-ratio under correctly specified models (t-stat$_c$) in round brackets. In addition, we present the sample OLS cross-sectional R^2 (R^2). The sample periods for equities (*FF25*), government and corporate bonds (*US bonds*), sovereign bonds (*Sov. bonds*), options (*Options*), credit default swaps (*CDS*), commodities (*Commod.*), and foreign exchange (*FX*) are 1970:Q1-2012:Q4, 1975:Q1-2011:Q4, 1995:Q1-2011:Q1, 1986:Q2-2011:Q4, 2001:Q2-2012:Q4, 1986:Q4-2012:Q4, and 1976:Q2-2009:Q4, respectively. N and T represent the number of assets and time-series observations, respectively.

(a) CAPM

	FF25	US Bonds	Sov. Bonds	Option	CDS	Commod.	FX	All
MKT	-0.99	3.41	4.21	8.60	5.85	-0.88	12.48	1.77
t-stat$_c$	(-0.80)	(2.97)	(1.74)	(2.91)	(2.85)	(-0.53)	(2.93)	(1.04)
INT	3.27	0.40	0.56	-5.98	-0.34	0.52	-1.74	0.09
t-stat$_c$	(3.12)	(1.86)	(0.88)	(-2.72)	(-3.54)	(0.84)	(-1.91)	(0.16)
R^2	0.08	0.81	0.69	0.91	0.61	0.01	0.49	0.32
N	25	20	6	18	20	23	12	124
T	172	148	65	103	47	105	135	172

(b) HKMSF

	FF25	US Bonds	Sov. Bonds	Option	CDS	Commod.	FX	All
CPTL	1.70	8.48	8.25	27.08	10.65	0.89	40.20	5.68
t-stat$_c$	(0.43)	(2.32)	(1.32)	(1.85)	(1.78)	(0.40)	(1.45)	(1.02)
INT	1.49	0.33	0.95	-4.62	-0.25	0.28	-2.22	-0.14
t-stat$_c$	(1.22)	(1.35)	(1.37)	(-1.88)	(-3.28)	(0.48)	(-1.26)	(-0.26)
R^2	0.02	0.79	0.61	0.91	0.42	0.00	0.27	0.38

Table 3.7: Price Reaction Around the NY Fed's Announcement of a New Primary Dealer

The table displays the daily estimation results based on Equation (3.5) from January 1970 until December 2020. We consider the 10 days that surround the NY Fed's announcement of a new primary dealer. We have 70 such events. The returns on the ultimate parent companies are regressed against company, year, and industry fixed effects, and a series of dummy variables for each day in the estimation window. These dummy variables are equal to one when it is announcement day and zero otherwise. Beta represents the deviation of the ultimate parent company's return from its long-run average on a given day. Day 0 represents the announcement day. The asterisks *, **, *** denote statistical significance at the 90%, 95%, and 99% significance levels, respectively. The *t*-statistics (in brackets) are computed by using robust standard errors clustered by firms and days. N. Obs. represents the number of observations used to estimate the panel regressions, and Adj. R^2 denotes the adjusted R-squared.

	Beta				**Beta**				**Beta**		
Day	**All**	**USA**	**For**	**Day**	**All**	**USA**	**For**	**Day**	**All**	**USA**	**For**
-5	0.117	-0.002	0.006*	-1	0.178	0.278	0.003	3	0.111	0.127	0.001
	(0.78)	*(-0.63)*	*(1.75)*		*(1.33)*	*(1.22)*	*(0.93)*		*(0.65)*	*(0.39)*	*(0.23)*
-4	0.352	0.219	0.002	0	1.301	-0.386**	-0.001	4	-0.169	-0.142	-0.001
	(0.84)	*(0.52)*	*(1.09)*		*(1.12)*	*(-2.19)*	*(-0.50)*		*(-1.19)*	*(-0.51)*	*(-0.27)*
-3	0.198	0.487	0.000	1	0.060	-0.256	0.004	5	-0.232***	-0.521***	0.002
	(0.97)	*(1.51)*	*(0.08)*		*(0.37)*	*(-0.90)*	*(1.06)*		*(-3.00)*	*(-3.80)*	*(0.88)*
-2	-0.159	-0.449**	0.000	2	0.008	0.165	-0.010***				
	(-1.58)	*(-2.35)*	*(-0.07)*		*(0.07)*	*(0.69)*	*(-2.95)*				
N°Obs.	580,749	224,092	356,657								
Adj.R^2	0.140	0.203	0.003								

Figure 3.1: Number of Primary Dealers per Year per Geographic Area

The figure shows the distribution of the primary dealers over time based on the geographic area of the ultimate parent company. The list of primary dealers is from the NY Fed's website, while the ultimate parent companies are manually matched. The sample covers the period from 1970 to 2020 for a total of 177 primary dealers, where 74 ultimate parent companies are located outside the US and 80 of them are from the US. The countries are identified by the currency of the traded stock's price: United States (USA), Japan (JPY), Great Britain (GBP), Euro Area (EUR), Switzerland (CHF), Australia (AUD), Canada (CAD), and Hong Kong (HKD).

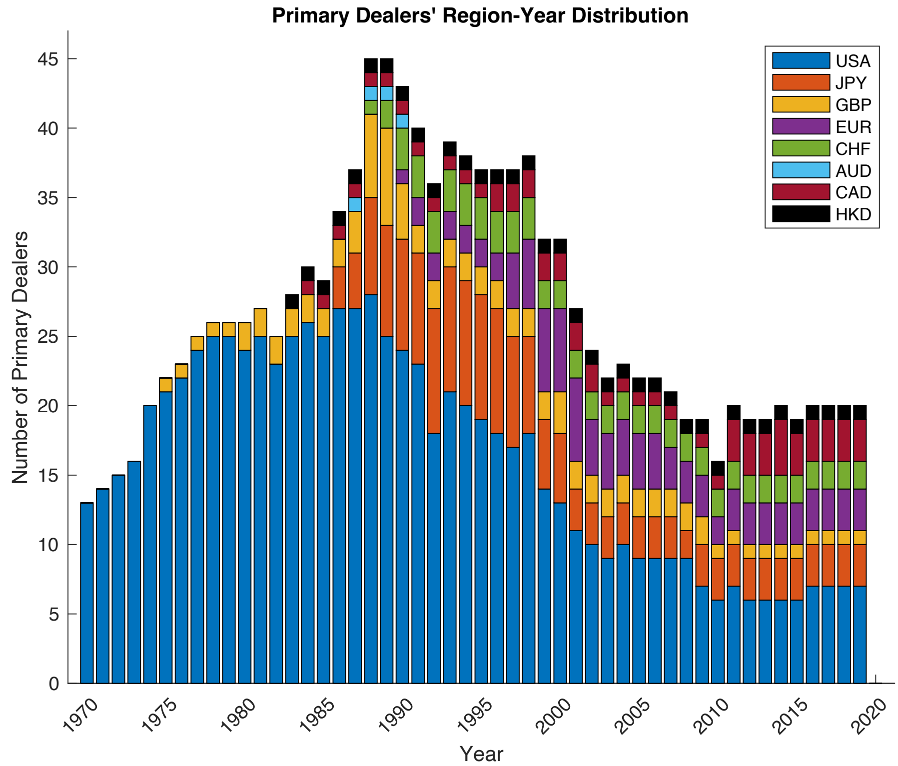


Figure 3.2: Number of Primary Dealers per Year per Standard Industry Classification Code

The figure shows the distribution of the primary dealers over time based on the standard industry classification (SIC) of the ultimate parent company. The list of primary dealers is from the NY Fed's website, while the ultimate parent companies are manually matched. The sample covers the period from 1970 to 2020. The SIC classification is as follows: department stores (5311); commercial banks (6020); personal credit institutions (6141); mortgage bankers and loan correspondents (6162); finance services (6199); security and commodity brokers (6200); security brokers, dealers, and flotation companies (6211); investment advice (6282); life insurance (6311); fire, marine, and casualty insurance (6331); investors, not elsewhere classified (6799); industrial conglomerates (9997).

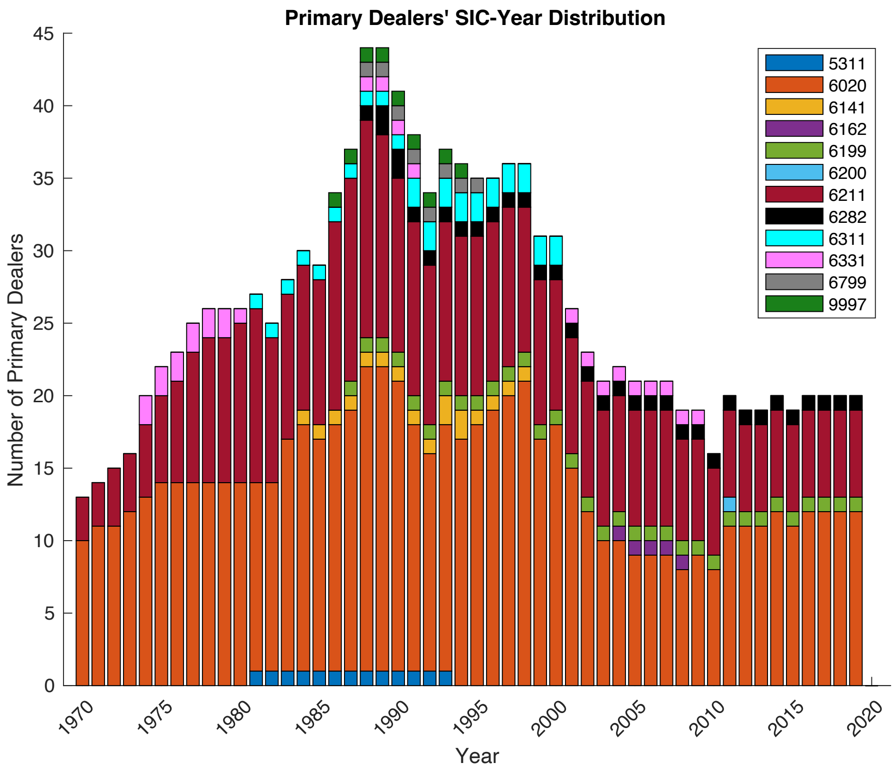


Figure 3.3: Intermediary Capital Ratio

The figure plots the capital ratio of the NY Fed's primary dealers (as identified by their ultimate parent companies) at a quarterly frequency from 1970:Q1 to 2020:Q4. The shaded regions in the chart represent the NBER recession periods.

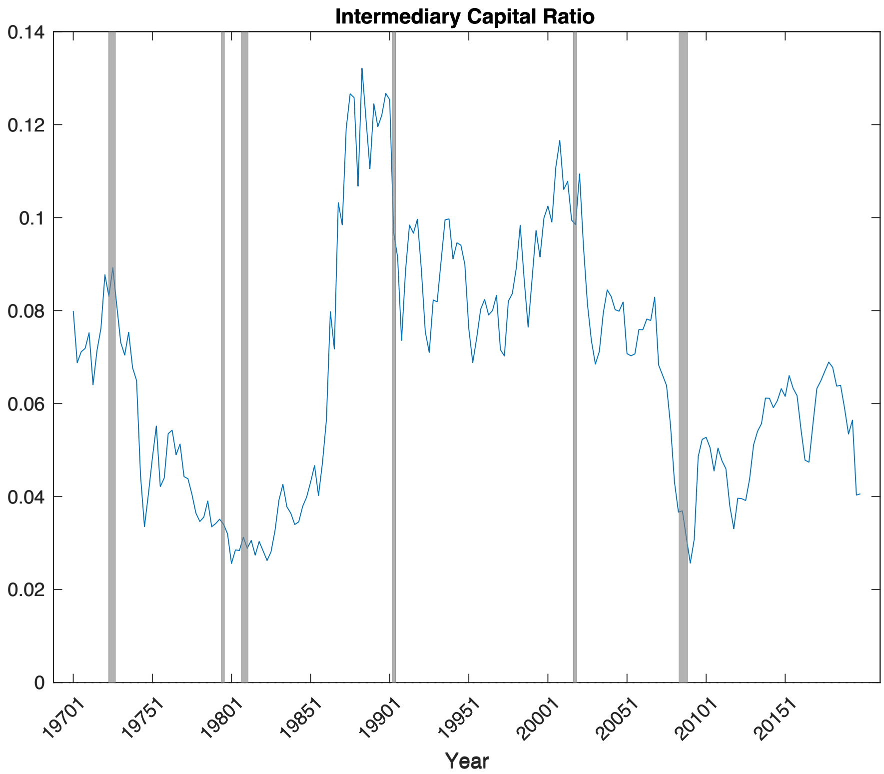


Figure 3.4: Intermediary Capital Risk Factor

The figure plots the value-weighted return of the NY Fed's primary dealers (as identified by their ultimate parent companies) at a quarterly frequency from 1970:Q1 to 2020:Q4. The shaded regions in the chart represent the NBER recession periods.

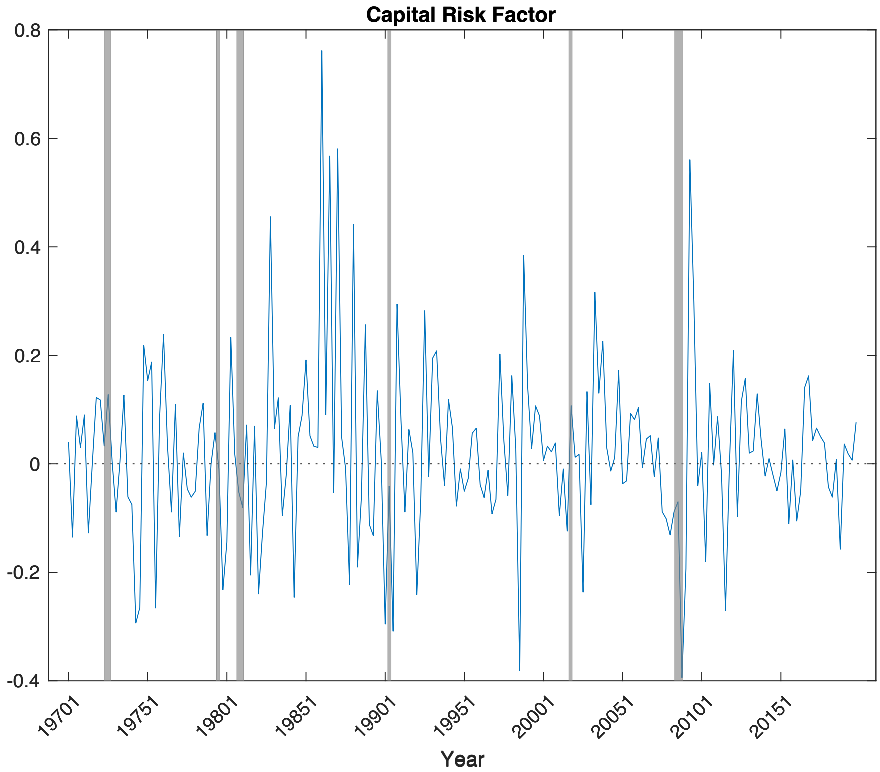


Figure 3.5: Market Risk Factor vs. IXG

The figure shows the excess value-weighted return of the NY Fed's primary dealers and the IXG excess return at a quarterly frequency from 2007:Q2 to 2020:Q4.

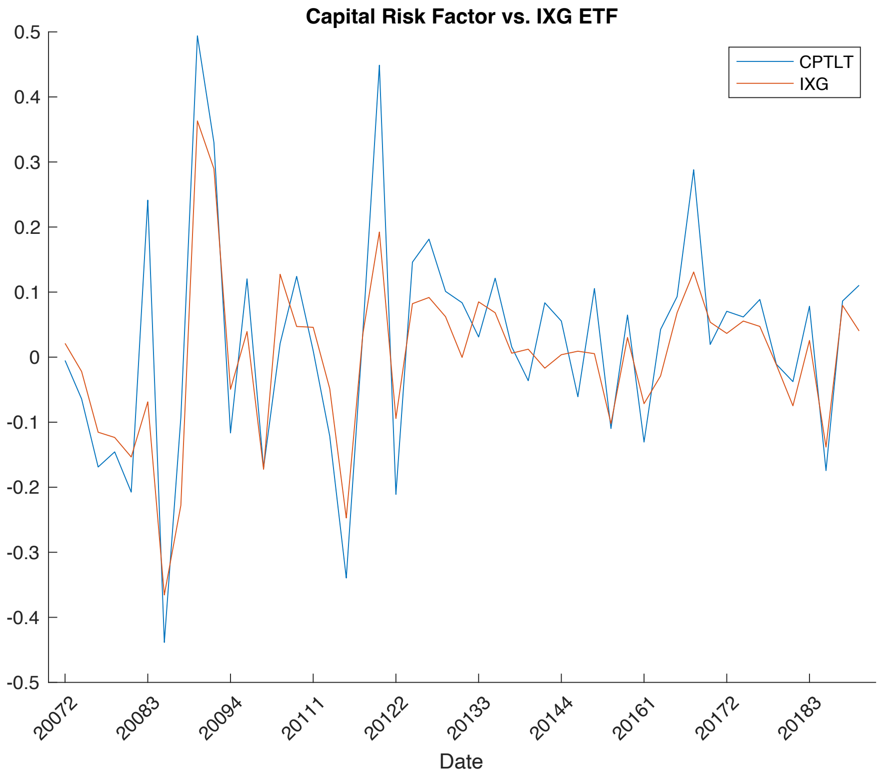


Figure 3.6: Two-Year Realized Volatility

The figure shows the monthly average realized volatility of 74 selected financial intermediaries from 12 months before to 12 months after they become NY Fed's primary dealers. The dashed lines represent the 95% confidence bounds. An event date equal to 0 indicates the month in which the selected financial intermediaries become NY Fed's primary dealers.

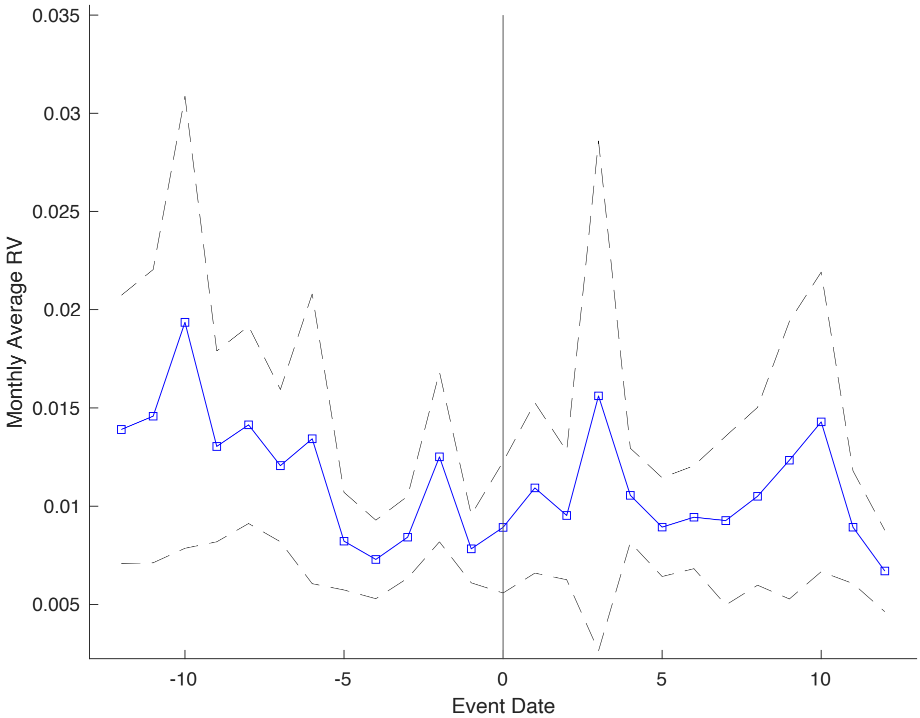


Figure 3.7: Two-Year Cumulative Return

The figure shows the monthly average cumulative return of 74 selected financial intermediaries from 12 months before to 12 months after they become NY Fed's primary dealers. The dashed lines represent the 95% confidence bounds. An event date equal to 0 indicates the month in which the selected financial intermediaries become NY Fed's primary dealers. At month $t-13$, the initial investment in each company is 100$. Then, we compute the cumulative return over time and average it across all of the selected firms.

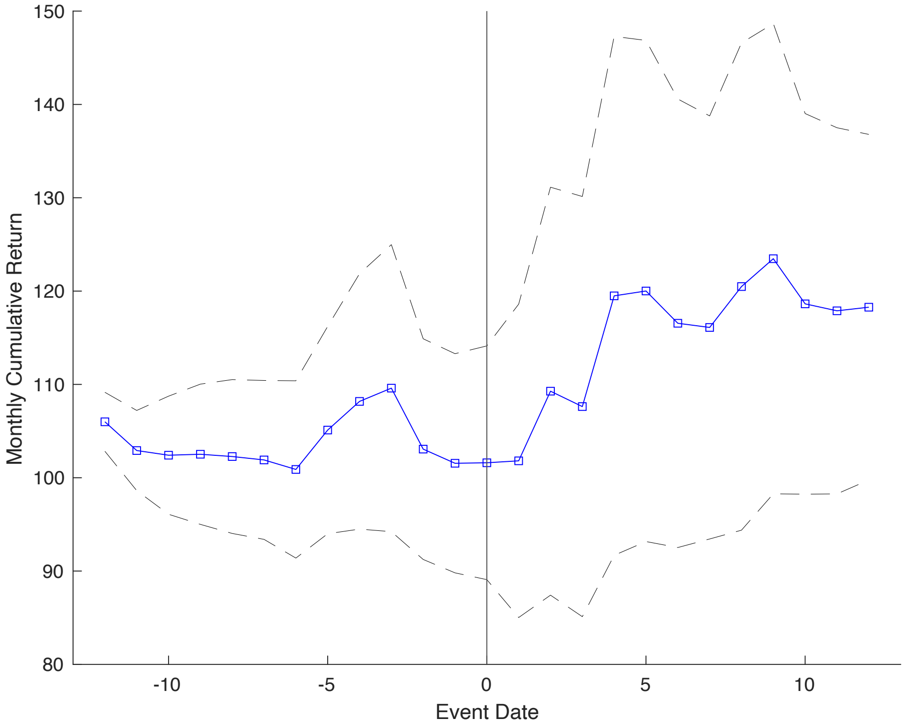
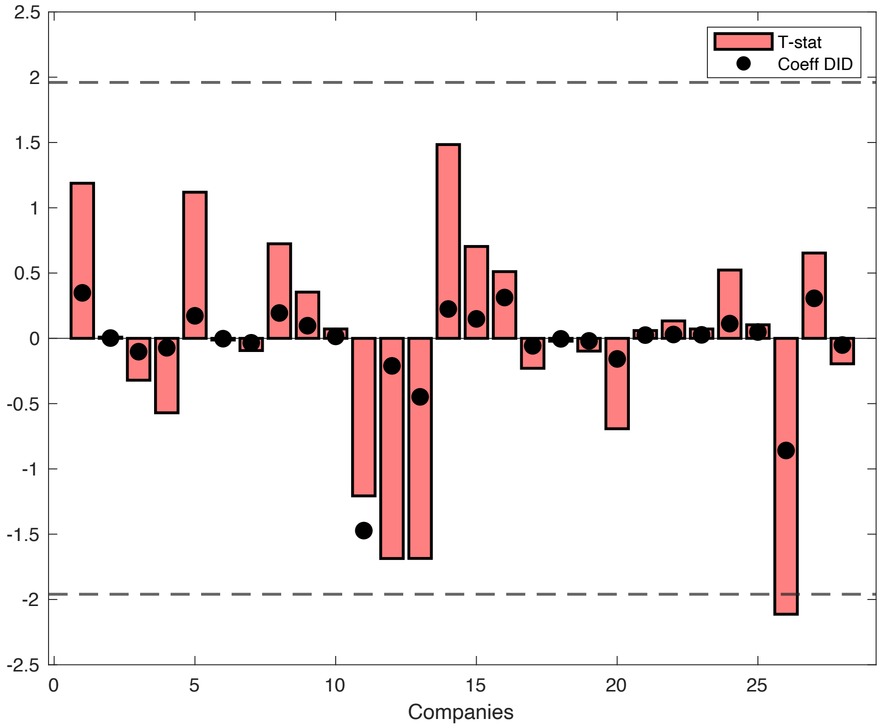


Figure 3.8: Difference-in-Difference Results

The figure displays the results based on estimating Equation (3.6) company by company. The regression is performed using the daily returns of the domestic financial intermediaries for which we have the relevant stock information from one year before until one year after the date in which they become NY Fed's primary dealers for the first time (29 companies in total, treatment group). The control group consists of an equally-weighted portfolio that contains all of the companies classified as broker-dealers (SIC codes 6211 or 6221). In the calculation of this daily equally-weighted portfolio, we exclude those ultimate parent companies whose subsidiaries, on each given day, were operating as NY Fed's primary dealers. The bars represent the t-statistics associated with the interaction term coefficient estimates, while the dots denote the magnitudes of the coefficient estimates (multiplied by 100). The horizontal dashed lines denote statistical significance at the 95% confidence level.



Appendix A

The Effect of Unscheduled News on Systematic Risk

A.1 Patton and Verardo (2012) Main Results

In this Appendix, I briefly illustrate the main results of Patton and Verardo (2012) from January 1, 2006 to December 31, 2008, to highlight the major differences between the behaviour of beta around scheduled (earnings) and unscheduled (SEOs) announcements.

The authors theoretically and then empirically show that, in the presence of earnings announcements of company x and in absence of any announcements about company y, investors revise their expectations about y using the information released by x. Thus, investors interpret good (bad) news from the announcing firm, as good (bad) news about non-announcing firms.

To replicate their empirical analysis, I gather from I/B/E/S the quarterly earnings announcements for the S&P 500 constituent companies. Then, from the TAQ dataset, I calculate the National Best Bid and Offer (NBBO) (Holden and Jacobsen, 2014) of the stocks and the S&P 500 - proxied by the ETF SPY - with a twenty-five-minute sampling frequency from 9:45 a.m. to 4:00 p.m. The event window account for twenty-one-day around the event date. Then I regress the realised beta on a series of dummies as specified in Equation 1.4. This methodology produces the following results:

Table A.1: Change in Beta - Earnings Announcements

Event Day	***Beta***	***Event Day***	***Beta***	***Event Day***	***Beta***
-10	0.003	-3	0.026^{**}	4	-0.022
	(0.28)		*(2.02)*		*(-1.58)*
-9	0.015	-2	-0.003	5	-0.010
	(1.21)		*(-0.17)*		*(-0.81)*
-8	0.010	-1	0.003	6	-0.011
	(0.76)		*(0.24)*		*(-0.82)*
-7	-0.002	0	0.092^{***}	7	-0.006
	(-0.08)		*(3.21)*		*(-0.39)*
-6	0.027^{**}	1	-0.043^{***}	8	0.017
	(2.12)		*(-2.77)*		*(1.24)*
-5	0.006	2	-0.027^{*}	9	0.017
	(0.48)		*(-1.89)*		*(1.31)*
-4	0.001	3	-0.004	10	0.002
	(0.08)		*(-0.27)*		*(0.20)*
No.Obs	143,553				
Adj.R^2	0.5722				

The table shows the results for 143,553 observations for a total of 5,765 announcements.[1] In the above table the only statistically significant coefficient at the 1% level (***) is on the event day and the day after. Furthermore, three days before and two days after the announcement date the change in beta is statistically significant at the 5% level (**), and at the 10% level (*), respectively. The evolution of the beta around earnings announcements shows a clear pattern. The beta starts to increase a few days before the announcements and spikes on the announcement date. The day after the announcement date it becomes negative, implying a lack of further information to be learned. Afterwords it reverts to its long-run average.

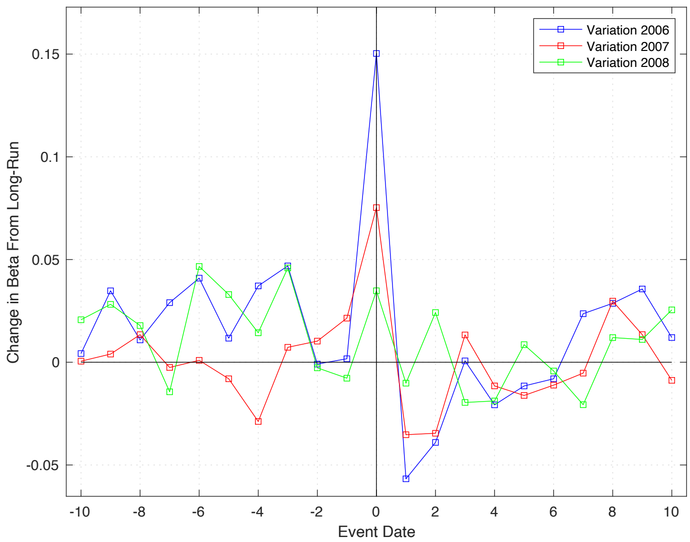
[1] The model estimated by Patton and Verardo (2012) considers a sample that is almost nine times larger. This difference, along with the difference in the period considered lead to a slightly different variation in beta. According to their results the systematic risk starts to increase four days before the announcement date. However, the overall pattern is precisely the same.

A.2 Investors Learning over Recession Period

In this Appendix, I show that over a recession period investors pay less attention to firm-level news, probably in favour of macro-level news. This shift in investors attention penalises the effect documented in this paper. To test this idea, I analysis the change in investors' learning in the presence of earnings announcements for the constituent companies of the S&P 500. To do so, I use the same sampling criteria adopted by Patton and Verardo (2012), which are briefly reviewed in Appendix A.1.

I estimate Equation 1.4 for 2006, 2007 and 2008 separately. If investors' learning is hampered by the recession, the increase in beta around the event date should be smaller and less statistically significant in 2008 compared to 2006. Instead, if the recession does not hamper investors' learning, or more generally they do not shift their attention to macro-level news, then the variation in beta should be the same across the three years. The results are plotted in Figure A.1.

Figure A.1: Change in Beta over Recession Period



Untabulated results show that the systematic risk increases on the announcement day by 0.150 (t-stat of 3.02), 0.075 (t-stat of 2.27) and 0.035 (t-stat of 0.84) in 2006, 2007 and 2008, respectively. Therefore, I conclude that during the recession period investigated here, investors pay less attention to firm-level news.

A.3 List of ETFs Tracking Sectors

The table shows the ETF tracking sectors. The first column reports the definition of the sector, the second column reports the equivalent of the Global Sectors Classification Standard (GICS) sector code, and the last column reports the trading symbol used in the TAQ dataset. Except for the telecommunication sectors and real estate that are from Vanguard, all the other ETFs refer to SPDR.[2]

Table A.2: ETFs Tracking Sectors

Sector	**Code**	**SPDR Fund Trading Symbol**
Consumer Discretionary	25	XLY
Consumer Staples	30	XLP
Energy	10	XLE
Financials	40	XLF
Health Care	35	XLV
Industrials	20	XLI
Information Technology	45	XLK
Materials	15	XLB
Telecommunication Services	50	VOX
Utilities	55	XLU
Real Estate	60	VNQ

[2]For more information about the SPDR ETFs `https://us.spdrs.com/en/etf/the-consumer-discretionary-select-sector-spdr-fund-XLY`.

A.4 Buyback Analysis

In this Appendix, I show the results when the same framework presented in the papers is applied to firm-level news writing about buyback. In short, around buyback the systematic risk of the announcing company increases, which suggests that when companies announce a distribution of wealth, the systematic risk increases (i.e., earnings). Meanwhile, when companies announcement an increase in capital, the systematic risk decreases.

Table A.3: Changes in Beta - Buyback (Market-Sample)

The table shows the results of Equation 1.4 from January 1, 2010 to December 31, 2013 for the twenty-one days around firm-level news writing about SEO, for a total of 420 announcements. The panel regression accounts for the realised beta of the company as a dependent variable calculated with sixteen intraday returns benchmarked against the S&P 500, which is regressed against the firm-year fix effect, as well as a dummy for each day in the estimation window (equals one if it is the announcement day and zero otherwise). *Beta* represents the deviation of the systematic risk from long-run average in a given day (i.e., *Event Day*) over a twenty-one day estimation window, thus Event Day 0 is the announcement day. *No. Obs* represents the number of observations used to estimate the panel regression of daily realised beta, and *Adj.* R^2 denotes the adjusted R-squared. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm-year and days. .

Event Day	***Beta***	***Event Day***	***Beta***	***Event Day***	***Beta***
-10	-0.020	-3	0.153^{**}	4	-0.062
	(-0.34)		*(2.53)*		*(-1.08)*
-9	0.058	-2	0.104^{*}	5	-0.037
	(0.94)		*(1.69)*		*(-0.62)*
-8	-0.008	-1	0.072	6	0.089
	(-0.11)		*(1.19)*		*(1.40)*
-7	-0.081	0	0.010	7	-0.025
	(-1.38)		*(0.11)*		*(-0.38)*
-6	0.085	1	-0.084	8	-0.004
	(1.33)		*(-1.4)*		*(-0.06)*
-5	-0.007	2	-0.087	9	0.032
	(-0.10)		*(-1.27)*		*(0.64)*
-4	-0.034	3	0.021	10	0.017
	(-0.57)		*(0.39)*		*(0.29)*
No.Obs	10,056				
Adj. R^2	0.4781				

Table A.4: Changes in Beta - Buyback (Sector-Sample)

The table shows the results of Equation 1.4 from January 1, 2010 to December 31, 2013 for the twenty-one-day around a firm-level news writing about buyback, for a total of 420 announcements. The panel regression accounts for the realised beta of the company as a dependent variable calculated with sixteen intraday returns benchmarked against the sector ETF (see Appendix A.3) of each company according to its two-digits GICS sector code, which is regressed against the firm-year fix effect, as well as a dummy for each day in the estimation window (equals one if it is the announcement day and zero otherwise). *Beta* represents the deviation of the systematic risk of the company from its long-run average in a given day (i.e. *Event Day*) over a twenty-one-day estimation window, thus Event Day 0 is announcement day. *No. Obs* represents the number of observations used to estimate the regression, and $Adj.R^2$ denotes the Adjusted R-squared. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm-year and days.

Event Day	***Beta***	***Event Day***	***Beta***	***Event Day***	***Beta***
-10	-0.020	-3	0.105^{**}	4	-0.038
	(-0.44)		*(1.99)*		*(-0.73)*
-9	0.042	-2	0.031	5	-0.099^{*}
	(0.78)		*(0.51)*		*(-1.76)*
-8	-0.051	-1	0.011	6	0.073
	(-0.87)		*(0.19)*		*(1.37)*
-7	-0.049	0	0.002	7	-0.025
	(-0.88)		*(0.04)*		*(-0.44)*
-6	0.061	1	-0.064	8	-0.015
	(1.15)		*(-1.17)*		*(-0.25)*
-5	0.068	2	-0.052	9	-0.022
	(1.23)		*(-0.87)*		*(-0.48)*
-4	-0.042	3	0.005	10	0.026
	(-0.78)		*(0.09)*		*(0.49)*
No.Obs	10,056				
$\boldsymbol{Adj.R^2}$	0.4614				

A.5 Company Characteristics Analysis

To investigate the effect of the company's characteristics in the systematic risk variation, by applying the same framework presented in this paper. I divide the market and sector-samples into tertiles based on the company's turnover, realised volatility and size; all calculated according to the three-day average two weeks before the start of the event window. I sort the companies according to size since investors use "bellwether" companies to price small stocks (Hameed et al., 2015). I also divide companies according to their realised variance. Savor and Wilson (2016) show that it is difficult for investors to learn about non-announcing firms from firm-level news with high volatile stocks. Turnover is used as a liquidity measure, to proxy for stock visibility (Patton and Verardo, 2012).

Panel A of Table A.5 reports the sample statistics of the market-sample, while **Panel B** reports the same statistics for the sector-sample. I then use the two samples to estimate Equation 1.4 and report the results in Table A.6 and A.7.

Table A.5: Summary Statistics - Company Characteristics

The table shows the sample statistics of the realised beta divided into tertiles and sorted according to: *Turn* that is the stock turnover calculated as the daily trading volume divided by the number of shares outstanding; *Vol* that is the trading volume and *Size* is calculated as the number of outstanding shares multiplied by the closing price. The tertiles are created according to the three-day average two weeks before the starts of the event window. **Panel A** shows the sample statistics for the market-sample, while **Panel B** shows the same statistics for the sector-sample. The table considers the observations from January 1, 2010 to December 31, 2013. The statistics refer to the distribution of beta where: *Mean* is the mean; *Std* is the standard deviation; *Skew* is the Skewness; *Q3* is the value of the third quartile; *Med* is the median; *Q1* is the value of the first quartile; *Turn* is the stock turnover calculated as the daily trading volume divided by the number of shares outstanding and *Size* is the average size of the companies reported in thousands.

Panel A: Sorted Market-Sample

	First Tertile			***Second Tertile***			***Third Tertile***		
	Turn	***Vol***	***Size***	***Turn***	***Vol***	***Size***	***Turn***	***Vol***	***Size***
Mean	0.731	0.799	0.950	1.038	1.078	1.099	1.341	1.244	1.071
Std	1.556	1.105	1.948	1.367	1.372	1.560	1.653	2.045	1.082
Skew	-0.066	-1.1600	0.118	-0.134	0.344	-0.375	-0.464	-0.329	-0.396
Q3	1.312	1.257	1.746	1.546	1.686	1.758	2.097	2.190	1.543
Med	0.690	0.764	0.842	0.970	1.023	0.999	1.257	1.182	1.006
Q1	0.107	0.328	0.099	0.447	0.415	0.379	0.546	0.310	0.531
Turn	7.28	12.971	19.829	11.365	12.961	19.821	32.761	26.417	12.756
Size	6,852.9	15,113.0	370.3	10,649.8	6,102.0	1,567.3	5,635.3	1,847.7	19,994.5

Panel B: Sorted Sectors-Sample

	First Tertile			***Second Tertile***			***Third Tertile***		
	Turn	***Vol***	***Size***	***Turn***	***Vol***	***Size***	***Turn***	***Vol***	***Size***
Mean	0.591	0.697	0.773	0.896	0.875	0.909	1.107	1.036	0.920
Std	1.325	0.970	1.687	1.197	1.184	1.295	1.420	1.728	0.938
Skew	-0.618	-1.596	-0.153	-0.835	-0.664	-1.149	-0.724	-0.545	-0.906
Q3	1.086	1.105	1.435	1.341	1.387	1.472	1.695	1.784	1.290
Med	0.555	0.680	0.656	0.841	0.844	0.828	1.064	0.993	0.882
Q1	0.098	0.277	0.048	0.417	0.358	0.314	0.473	0.271	0.487

Table A.6: Market-Sample Sorted on Turnover, Volatility and Size

The table shows the results of Equation 1.4 from January 1, 2010 to December 31, 2013 for the twenty-one days around firm-level news writing about SEO, for a total of 520 announcements. The sample is divided into tertiles and sorted according to: *Turn* that is the stock turnover calculated as the daily trading volume divided by the number of shares outstanding; *Vol* is the trading volume and *Size* that is calculate as the number of outstanding shares multiplied by the closing price. The tertiles are created according to the three-day average two weeks before the starts of the event window. The panel regression accounts for the realised beta of the company as a dependent variable calculated with sixteen intraday returns benchmarked against the S&P 500, which is regressed against the company-year fix effect, as well as a dummy for each day in the estimation window (equals one if it is the announcement day and zero otherwise). Under the nine columns the the deviation of the systematic risk of the company from its long-run average on a given day (i.e., *Event Day*) is reported. Thus, Event Day 0 is announcement day. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm-year and days.

	First Tertile			***Second Tertile***			***Third Tertile***		
Event Day	***Turn***	***Vol***	***Size***	***Turn***	***Vol***	***Size***	***Turn***	***Vol***	***Size***
-10	0.137	0.040	-0.001	0.102	0.058	0.098	-0.149	-0.062	-0.060
	(1.22)	*(0.68)*	*(0.00)*	*(1.10)*	*(0.67)*	*(0.79)*	*(-0.95)*	*(-0.44)*	*(-0.61)*
-9	0.324***	0.169***	0.510***	0.093	0.001	0.036	0.252	0.121	0.122
	(3.03)	*(2.69)*	*(3.19)*	*(0.90)*	*(0.01)*	*(0.32)*	*(1.63)*	*(0.76)*	*(1.36)*
-8	0.068	0.137	0.012	0.045	0.088	0.157	0.182	-0.134	0.120*
	(0.55)	*(1.46)*	*(0.09)*	*(0.48)*	*(0.89)*	*(1.23)*	*(1.40)*	*(-0.89)*	*(1.71)*
-7	0.167	0.067	0.285*	0.087	-0.001	0.064	0.063	-0.197*	-0.007
	(1.29)	*(0.90)*	*(1.86)*	*(0.90)*	*(0.00)*	*(0.53)*	*(0.50)*	*(-1.66)*	*(-0.07)*
-6	-0.005	0.126*	0.173	0.032	0.207*	0.056	0.256**	0.037	0.070
	(-0.04)	*(1.86)*	*(1.27)*	*(0.32)*	*(1.96)*	*(0.48)*	*(2.34)*	*(0.26)*	*(0.99)*
-5	-0.022	0.129	-0.032	0.110	-0.013	0.122	0.125	-0.032	0.107
	(-0.16)	*(1.59)*	*(-0.18)*	*(1.12)*	*(-0.13)*	*(1.19)*	*(0.93)*	*(-0.22)*	*(1.18)*
-4	0.125	0.100	0.095	-0.051	0.110	0.125	0.154	-0.196	0.052
	(0.91)	*(1.52)*	*(0.48)*	*(-0.33)*	*(1.20)*	*(0.96)*	*(1.10)*	*(-1.42)*	*(0.59)*
-3	-0.065	0.175***	-0.299	-0.047	0.079	0.243*	0.041	0.321**	-0.043
	(-0.51)	*(2.82)*	*(-1.72)*	*(-0.36)*	*(0.76)*	*(1.84)*	*(0.29)*	*(2.23)*	*(-0.45)*
-2	0.097	0.123*	0.072	-0.026	0.111	0.198	0.206	0.196	-0.063
	(0.94)	*(1.70)*	*(0.43)*	*(-0.23)*	*(0.99)*	*(1.59)*	*(1.35)*	*(1.58)*	*(-0.72)*
-1	0.314**	0.137**	0.209	-0.103	0.196*	-0.053	-0.073	-0.001	-0.052
	(2.32)	*(2.17)*	*(1.14)*	*(-0.75)*	*(1.89)*	*(-0.37)*	*(-0.51)*	*(-0.01)*	*(-0.65)*
0	-0.293	0.136	-0.166	-0.392**	0.041	-0.638***	-0.301	-0.049	-0.141
	(-1.27)	*(0.83)*	*(-0.61)*	*(-2.07)*	*(0.28)*	*(-3.02)*	*(-1.48)*	*(-0.26)*	*(-1.21)*
1	0.034	-0.064	-0.076	-0.072	0.085	-0.102	-0.155	-0.158	-0.031
	(0.27)	*(-0.85)*	*(-0.57)*	*(-0.66)*	*(0.70)*	*(-0.87)*	*(-1.38)*	*(-1.39)*	*(-0.37)*
2	0.132	0.028	-0.009	-0.067	0.14	-0.075	-0.101	-0.318**	0.051
	(1.09)	*(0.43)*	*(-0.06)*	*(-0.64)*	*(1.29)*	*(-0.55)*	*(-0.72)*	*(-2.03)*	*(0.66)*
3	0.058	0.116*	-0.019	-0.105	0.134	0.053	0.120	-0.073	0.090
	(0.58)	*(1.69)*	*(-0.13)*	*(-1.00)*	*(1.34)*	*(0.52)*	*(0.99)*	*(-0.60)*	*(1.07)*
4	0.038	0.163**	0.053	0.025	-0.092	-0.104	-0.080	-0.130	0.027
	(0.36)	*(2.30)*	*(0.41)*	*(0.29)*	*(-0.81)*	*(-0.86)*	*(-0.63)*	*(-1.16)*	*(0.40)*
5	0.130	0.117**	0.140	0.047	0.139*	-0.123	-0.116	-0.257*	0.043
	(1.25)	*(2.08)*	*(1.08)*	*(0.51)*	*(1.66)*	*(-1.03)*	*(-0.77)*	*(-1.78)*	*(0.47)*
6	0.239**	0.159*	0.149	-0.036	0.041	0.052	0.007	0.173	0.040
	(2.02)	*(1.71)*	*(1.06)*	*(-0.35)*	*(0.45)*	*(0.46)*	*(0.07)*	*(1.27)*	*(0.50)*
7	0.304***	0.072	0.124	-0.023	0.003	0.099	0.062	-0.042	0.113
	(2.66)	*(0.87)*	*(0.91)*	*(-0.24)*	*(0.03)*	*(0.92)*	*(0.52)*	*(-0.29)*	*(1.49)*
8	-0.046	0.040	0.016	0.091	0.192**	-0.094	-0.041	-0.134	0.029
	(-0.44)	*(0.64)*	*(0.12)*	*(1.04)*	*(2.12)*	*(-0.96)*	*(-0.38)*	*(-0.94)*	*(0.46)*
9	0.31***	0.094*	0.190	-0.026	0.073	0.077	0.066	0.043	0.075
	(2.53)	*(1.64)*	*(1.31)*	*(-0.28)*	*(0.79)*	*(0.81)*	*(0.55)*	*(0.39)*	*(0.91)*
10	0.242**	0.142 *	0.160	-0.037	0.141	-0.044	0.025	-0.122	0.068
	(2.36)	*(2.06)*	*(1.05)*	*(-0.35)*	*(1.61)*	*(-0.38)*	*(0.19)*	*(-0.88)*	*(0.91)*

Table A.7: Sector-Sample Sorted on Turnover, Volatility and Size

The table shows the results of Equation 1.4 from January 1, 2010 to December 31, 2013 for the twenty-one days around firm-level news writing about SEO, for a total of 520 announcements. The sample is divided into tertiles and sorted according to: *Turn* that is the stock turnover calculated as the daily trading volume divided by the number of shares outstanding; *Vol* is the trading volume and *Size* that is calculate as the number of outstanding shares multiplied by the closing price. The tertiles are created according to the three-day average two weeks before the starts of the event window. The panel regression accounts for the realised beta of the company as a dependent variable calculated with sixteen intraday returns benchmarked against the sector ETF (see Appendix A.3) of each company according to its two-digit GICS sector code, which is regressed against the company-year fix effect, as well as a dummy for each day in the estimation window (equals one if it is the announcement day and zero otherwise). Under the nine columns the the deviation of the systematic risk of the company from its long-run average on a given day (i.e., *Event Day*) is reported. Thus, Event Day 0 is announcement day. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm-year and days.

	First Tertile			Second Tertile			Third Tertile		
Event Day	Turn	Vol	Size	Turn	Vol	Size	Turn	Vol	Size
-10	0.033	0.001	-0.098	0.134	0.030	0.065	-0.139	-0.068	0.021
	(0.32)	(0.03)	(-0.69)	(1.53)	(0.36)	(0.74)	(-1.11)	(-0.42)	(0.23)
-9	0.103	0.041	0.201	0.046	-0.010	-0.065	0.019	0.093	0.012
	(1.22)	(0.67)	(1.59)	(0.55)	(-0.11)	(-0.79)	(0.16)	(0.66)	(0.15)
-8	0.024	0.040	-0.011	0.035	0.206*	-0.018	0.016	-0.205	0.096
	(0.21)	(0.58)	(-0.08)	(0.46)	(1.74)	(-0.16)	(0.14)	(-1.59)	(1.56)
-7	0.104	0.078	0.125	0.016	0.142*	0.080	0.111	0.007	0.044
	(0.96)	(0.81)	(1.06)	(0.19)	(1.84)	(0.69)	(0.93)	(0.05)	(0.51)
-6	-0.023	-0.001	0.076	-0.010	0.123	0.044	0.207*	0.035	0.062
	(-0.21)	(0.00)	(0.55)	(-0.10)	(1.35)	(0.40)	(1.80)	(0.28)	(1.04)
-5	-0.038	0.101	-0.041	0.039	0.044	0.149*	0.186*	0.020	0.080
	(-0.28)	(1.56)	(-0.25)	(0.46)	(0.41)	(1.68)	(1.66)	(0.13)	(1.09)
-4	0.154	0.035	0.172	0.038	-0.022	0.018	-0.028	0.171	0.013
	(1.18)	(0.35)	(1.02)	(0.38)	(-0.18)	(0.16)	(-0.21)	(1.17)	(0.20)
-3	-0.033	0.082	-0.185	-0.010	0.122	0.159*	-0.097	-0.385**	-0.131
	(-0.28)	(1.17)	(-1.22)	(-0.09)	(1.51)	(1.65)	(-0.83)	(-2.27)	(-1.52)
-2	0.032	0.061	0.154	-0.139	-0.040	-0.058	0.200	0.041	-0.010
	(0.36)	(0.75)	(1.07)	(-1.06)	(-0.40)	(-0.44)	(1.41)	(0.24)	(-0.12)
-1	0.055	0.142*	0.047	-0.061	-0.064	0.024	-0.026	-0.099	-0.101
	(0.45)	(1.66)	(0.30)	(-0.56)	(-0.56)	(0.22)	(-0.22)	(-0.67)	(-1.32)
0	-0.094	-0.215*	0.015	-0.350*	-0.155	-0.481***	-0.214	-0.268	-0.147
	(-0.52)	(-1.87)	(0.06)	(-1.96)	(-0.92)	(-2.86)	(-1.23)	(-1.16)	(-1.41)
1	-0.017	0.086	-0.062	-0.011	0.066	-0.055	-0.082	-0.296**	-0.007
	(-0.17)	(1.21)	(-0.55)	(-0.11)	(0.88)	(-0.56)	(-0.83)	(-2.26)	(-0.11)
2	0.176*	0.113*	0.179	0.091	0.209**	-0.055	-0.028	-0.090	0.119
	(1.94)	(1.69)	(1.61)	(0.99)	(2.44)	(-0.60)	(-0.25)	(-0.70)	(1.50)
3	0.103	0.008	-0.008	-0.107	0.125	0.006	-0.014	-0.166	-0.008
	(1.16)	(0.11)	(-0.07)	(-1.08)	(1.42)	(0.07)	(-0.13)	(-1.28)	(-0.10)
4	-0.014	-0.040	0.014	0.082	0.052	-0.063	-0.107	-0.074	0.010
	(-0.15)	(-0.43)	(0.12)	(1.05)	(0.63)	(-0.56)	(-0.91)	(-0.62)	(0.20)
5	0.096	0.079	0.043	0.077	0.056	0.038	-0.050	-0.047	0.038
	(1.17)	(1.39)	(0.41)	(1.03)	(0.72)	(0.48)	(-0.50)	(-0.41)	(0.56)
6	0.095	0.099*	-0.026	-0.007	0.028	0.054	-0.030	-0.097	0.027
	(0.89)	(1.65)	(-0.21)	(-0.08)	(0.30)	(0.61)	(-0.31)	(-0.81)	(0.42)
7	0.238**	0.147***	0.140	0.010	-0.010	0.079	0.001	0.089	0.029
	(2.42)	(2.73)	(1.14)	(0.14)	(-0.12)	(1.01)	(0.02)	(0.66)	(0.47)
8	-0.037	0.089	-0.044	0.085	-0.004	0.000	-0.135	-0.209*	-0.067
	(-0.38)	(1.46)	(-0.38)	(1.16)	(-0.04)	(0.01)	(-1.48)	(-1.79)	(-1.10)
9	0.155	0.076	0.136	-0.036	0.074	-0.036	0.036	-0.027	0.049
	(1.43)	(1.16)	(1.04)	(-0.45)	(1.01)	(-0.43)	(0.35)	(-0.19)	(0.67)
10	0.079	0.103	0.095	0.033	0.054	-0.016	-0.002	-0.074	0.033
	(1.09)	(1.71)	(0.84)	(0.37)	(0.67)	(-0.16)	(-0.01)	(-0.62)	(0.51)

A.6 Individual Sectors Analysis

Table A.8: Changes in Beta - Individual Sector

The table shows the results of Equation 1.4 from January 1, 2010 to December 31, 2013 for the twenty-one days around a firm-level news writing about SEO, for a total of 520 announcements. The entire sample is divided according to the company two-digit GICS sector code. The panel regression accounts for the realised beta of the company as a dependent variable calculated with sixteen intraday returns benchmarked against the sector ETF (see Appendix A.3) of each company according to its two-digit GICS sector code, which is regressed against the company-year fix effect, as well as a dummy for each day in the estimation window (equals one if it is the announcement day and zero otherwise). Under the eight columns the the deviation of the systematic risk of the company from its long-run average on a given day (i.e., *Event Day*) is reported. Thus, Event Day 0 is announcement day. The asterisks *, **, *** denote statistical significance at the 90%, 95% and 99% significance level, respectively. The *t-statistics* (in brackets) are computed by using robust standard errors clustered by firm-year and days.

Event Day	*XLY*	*XLP*	*XLE*	*XLV*	*XLI*	*XLK*	*XLB*	*VOX*
-10	0.378***	0.214	0.185*	-0.442**	0.113	0.176	-0.093	0.327
	(2.70)	*(0.87)*	*(1.95)*	*(-2.08)*	*(0.84)*	*(0.76)*	*(-0.42)*	*(1.06)*
-9	0.003	-0.150	0.083	-0.068	0.097	0.410*	0.185	0.448
	(0.02)	*(-0.33)*	*(0.98)*	*(-0.49)*	*(0.69)*	*(1.79)*	*(1.08)*	*(1.30)*
-8	0.115	0.432	0.063	-0.091	0.150	-0.154	0.268	0.214
	(0.67)	*(1.32)*	*(0.61)*	*(-0.47)*	*(0.81)*	*(-0.50)*	*(1.50)*	*(0.84)*
-7	-0.188	-0.146	-0.057	0.186	0.361*	0.456*	0.385*	0.041*
	(-1.16)	*(-0.57)*	*(-0.44)*	*(1.40)*	*(1.82)*	*(1.73)*	*(1.96)*	*(0.13)*
-6	0.082	0.583*	0.002	0.047	0.154	0.331***	0.251*	-0.245
	(0.51)	*(1.86)*	*(0.02)*	*(0.26)*	*(0.92)*	*(1.64)*	*(1.90)*	*(-0.76)*
-5	0.164	0.656*	0.218*	-0.295	0.114	0.163	0.338*	0.187
	(0.63)	*(1.81)*	*(1.94)*	*(-1.57)*	*(0.76)*	*(0.89)*	*(1.76)*	*(0.38)*
-4	0.273*	0.133	0.005	-0.152	0.479**	-0.027	0.622**	-0.124
	(1.77)	*(0.44)*	*(0.05)*	*(-0.71)*	*(2.25)*	*(-0.09)*	*(2.16)*	*(-0.35)*
-3	-0.101	0.956***	-0.168	-0.093	-0.022	0.053	0.170	0.048
	(-0.40)	*(4.02)*	*(-1.06)*	*(-0.44)*	*(-0.16)*	*(0.26)*	*(1.01)*	*(0.19)*
-2	0.048	0.361*	-0.067	0.140	0.010	0.050	0.437**	-0.611
	(0.31)	*(1.94)*	*(-0.67)*	*(0.53)*	*(0.05)*	*(0.25)*	*(2.42)*	*(-1.45)*
-1	0.263	0.767***	0.098	-0.478**	-0.027	0.063	0.005	0.545
	(1.35)	*(3.02)*	*(1.08)*	*(-2.17)*	*(-0.14)*	*(0.27)*	*(0.03)*	*(1.11)*
0	-0.417	0.363	-0.279*	0.110	0.104	-0.517*	0.121	-0.854
	(-1.06)	*(0.68)*	*(-1.87)*	*(0.33)*	*(0.39)*	*(-1.85)*	*(0.50)*	*(-1.48)*
1	0.101	0.384	0.069	-0.338*	0.088	0.054	0.118	-0.303
	(0.58)	*(1.50)*	*(0.85)*	*(-1.88)*	*(0.64)*	*(0.29)*	*(0.80)*	*(-0.72)*
2	0.013	0.428**	0.038	0.108	0.122	0.370*	0.063	0.404
	(0.08)	*(2.26)*	*(0.43)*	*(0.69)*	*(0.64)*	*(1.65)*	*(0.40)*	*(1.01)*
3	-0.008	0.406	0.028	-0.054	-0.058	0.071	0.220*	-0.087
	(-0.03)	*(1.47)*	*(0.37)*	*(-0.28)*	*(-0.35)*	*(0.38)*	*(1.65)*	*(-0.33)*
4	0.140	0.287	0.077	-0.195	-0.040	0.049	0.364**	-0.112
	(1.06)	*(0.93)*	*(0.92)*	*(-0.99)*	*(-0.29)*	*(0.26)*	*(2.36)*	*(-0.26)*
5	0.004	0.501**	0.046	-0.050	0.219	0.073	0.288**	-0.146
	(0.03)	*(1.83)*	*(0.56)*	*(-0.34)*	*(1.52)*	*(0.44)*	*(1.69)*	*(-0.55)*
6	0.277	0.506***	0.008	-0.194	0.241	-0.049	0.096	-0.262
	(1.63)	*(2.96)*	*(0.10)*	*(-1.08)*	*(1.53)*	*(-0.25)*	*(0.52)*	*(-0.94)*
7	0.409*	0.590***	0.211**	-0.183	0.176	0.092	0.321**	-0.616**
	(1.88)	*(3.39)*	*(2.46)*	*(-1.21)*	*(1.31)*	*(0.51)*	*(2.29)*	*(-2.02)*
8	-0.214	0.307	0.047	-0.114	0.264*	0.210	-0.045	-0.345
	(-1.14)	*(1.41)*	*(0.56)*	*(-0.77)*	*(1.81)*	*(1.25)*	*(-0.30)*	*(-1.11)*
9	0.467*	0.276	0.079	-0.039	-0.032	0.188	-0.086	-0.006
	(1.80)	*(1.06)*	*(1.14)*	*(-0.24)*	*(-0.25)*	*(0.82)*	*(-0.51)*	*(-0.02)*
10	0.203	-0.005	0.063	-0.058	0.182	0.155	0.053	-0.331
	(1.55)	*(-0.02)*	*(0.78)*	*(-0.41)*	*(0.94)*	*(0.77)*	*(0.32)*	*(-1.30)*

Appendix B

Dissecting Firm-Level News: a New Measure to Capture the Time-Varying Risk of the Company

B.1 Firm-Level New - Selection and Pre-Proces

The news is obtained and then pre-processed as follows:

1. Identify the firm-level from the Lexis-Nexis archive with a company "relevance score" ranging between 80% and 99%, while from ProQuest the name of the company and/or the ticker have to appear in the headline or the abstract.

2. Exclude news that is not from the business section (Pittsburgh Post Gazette, The Atalanta Journal-Constituent, The Philadelphia Inquirer and St. Louis Post Dispatch), money section (USA Today), financial desk section (New York Times), investing and business section (Tulsa World), section A ,B, C or missing (Wall Street Journal) (Bybee et al., 2020) and the A-Section or financial section (The Washington Post).

3. Exclude everything that is not in the main corpus of the news (ie., headline).

4. Eliminate stop-words[1], words with less than three letters, words written all in upper case (i.e., AAPL, AMZN) and words that start with capital letters that are not at the beginning of the sentence (i.e., Amazon, Franco). Words in capital letters at the beginning of the sentence are, instead, transformed into lower case.

5. Exclude articles with more than 40% digits and with less than 100 characters.

[1] The list can be downloaded from the following two links: `http://ir.dcs.gla.ac.uk/resources/linguistic_utils/stop_words` and `http://snowball.tartarus.org/algorithms/english/stop.txt`.

6. Eliminate all digits. Words divided by a hyphen are separated, and the hyphen eliminated.

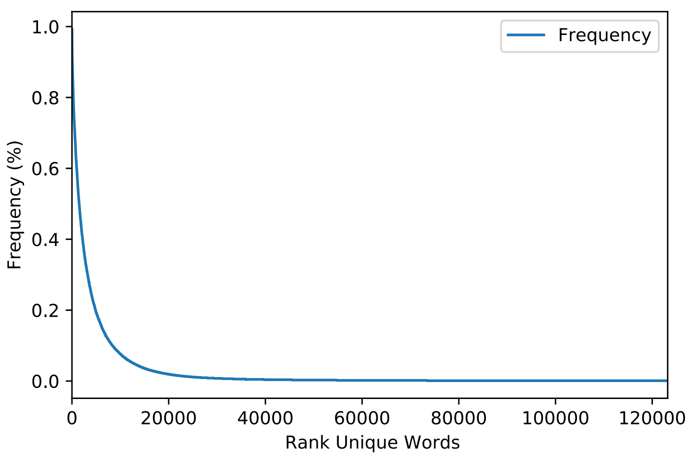
7. The dot at the end of the sentence is used to split the corpus into sentences and then the punctuation is eliminated. If a sentence has less than three words it is eliminated, while if more than three and less than ten is concatenated with the previous sentence.

8. Words are stemmed by using Part Of Speech (POS) and a series of rules.[2] As opposite to the normal stemmer or trimming approach, this methodology always delivers human interpretable words. For example, the verb *make* in any of its forms (i.e., made, makes, making) is always referred back to make.

9. If a word appears in more than 40% of the news it is excluded. This step prevents the inclusion of common words.

10. I discard words with the least relative frequency across the collection of news for each single company. To do so, I group all news of each company and retain only unique words, and count their frequency. Then, I count the number of times each unique word is observed across all companies and rank them in ascending order. Figure B.1 shows that the relative frequency across companies (i.e., y-axis) of each unique word (i.e., x-axis) follows Zipf's law.[3]

Figure B.1: Frequency of Unique Words



---

[2] Code available from the author upon request.

[3] Zipf's law predicts that the word distribution follows the power law, or else that few words account for a large fraction of the document, whereas a large number of words account for a low fraction of the document (Loughran and McDonald, 2016).

Zipf's law asserts that the word rank frequency decays by a factor of n, so the word ranked in the position n is observed $1/n$ less time than the word ranked in the first place. Therefore, out of the 123,893 unique words, I retain only 16,241, which is where the blue line starts to flatten out. Stated differently, a word must be observed in at least twenty-five different documents (i.e., companies). A parsimonious assumption is that these 16,241 words are strictly related to information related to financial matters, which does not mean that they are financially-related words (i.e., a financial vocabulary would count 1,242 words (Calomiris and Mamaysky, 2019)). The number of words delivered by the proposed approach is similar to those determined by Bybee et al. (2020), who identified 18,432 words related to economic matters.

11. I calculate the most common bi-grams through a maximum likelihood estimation and retain 3,247 (i.e., 20% of the unique words). The identified bi-grams are then used to replaced the corresponding uni-grams (i.e., second quarter becomes second_quarter). I calculate the maximum likelihood estimation as follows:

$$P(w_n|w_{n-1}) = \frac{C(w_{n-1}w_n)}{C(w_{n-1})} \tag{B.1}$$

Where $C(w_{n-1}w_n)$ is the count of the time the word w_{n-1} is followed by the word w_n divided by the count of the word w_{n-1} (Jurafsky and Martin, 2008).

B.2 Market Similarity Sorted Tertiles

Table B.1: Market Similarity Sorted Tertiles - Summary Statistics and Systematic Risk

The table shows sample statistics from January 1, 1990 to December 31, 2017 of three equally-weighted tertile portfolios sorted in ascending order according to their MS in each week, where *Q1* is composed by companies with the lowest MS and *Q3* is composed by companies with the highest MS, *Mkt* are the statistics of the S&P 500 and *No-News* is a portfolio composed by companies without news items in each given week. **Panel A** shows the annualised weekly statistics for: *Ret* is the average returns of the tertiles; *Std* is the standard deviation of returns; *SR* is the Sharpe Ratio; *No. Comp* is the average number of companies in each portfolio; *Size* is the market capitalisation of the companies, which is calculated as the latest quarterly number of shares outstanding from Compustat, multiplied by the current stock's price, reported in thousands; *Turn* is the turnover calculated as the average weekly volume divided by the last available number of shares outstanding, reported in thousands. **Panel B** shows the systematic risk (β) of the tertiles, and the t-statistics between brackets calculated with standard errors corrected for heteroskedasticity and with the Newey-West procedure with one lag. *No. Obs* is the number of weekly observations used to estimate the model, and R^2 is the the R-squared of the model. The standard errors of the regressions are corrected for heteroskedasticity and with the Newey-West procedure with one lag. The t-statistics are reported between brackets and the *, ** and *** symbols denote statistical significance at the 10%, 5% and 1% levels.

Panel A: Sample Statistic

	Q1	***Q2***	***Q3***	***No-News***	***Mkt***
Ret	0.088	0.094	0.113	0.074	0.079
Std	0.175	0.200	0.234	0.173	0.162
SR	0.345	0.332	0.367	0.271	0.320
No.Comp	35	34	36	395	500
Size	49.370	47.142	47.350	13.971	-
Turn	86.694	133.233	276.094	63.718	-

Panel B: Sorted Portfolios Systematic Risk

	Q1	***Q2***	***Q3***	***No-News***	***Mkt***
$\boldsymbol{\beta}$	0.980^{***}	1.088^{***}	1.178^{***}	1.013^{***}	-
	(54.01)	*(44.19)*	*(36.24)*	*(65.43)*	-
No.Obs	1460	1460	1460	1460	-
$\boldsymbol{R^2}$	0.819	0.776	0.662	0.902	-

Appendix C

Are the Primary Dealers of the New York Fed Really Special?

C.1 He et al. (2017) Replication Exercise

In this appendix, we report the results of our replication exercise of the He et al. (2017) series. In the interest of brevity, we only include figures for the intermediary capital ratio and traded capital risk factor at a quarterly frequency (the complete set of results is available from the authors upon request). The computations of the intermediary capital ratio are based on Equation (3.1) and are displayed in Figures C.1-C.3. Similarly, the computations of the traded intermediary capital risk factor are based on Equation (3.4) and are displayed in Figures C.4-C.6. We report, in order, results for all of the ultimate parent companies (Figures C.1 and C.4), for all but the Japanese ultimate parent companies (Figures C.2 and C.5), and for all but the Japanese and a few more ultimate parent companies (Figures C.3 and C.6).

Figure C.1: Intermediary Capital Ratio (All Companies)

The figure plots two capital ratio series from 1970:Q1 to 2018:Q3. HKM is the series of He et al. (2017) from Asaf Manela's website, while Calculated denotes the series based on our own calculations and includes all domestic and foreign ultimate parent companies.

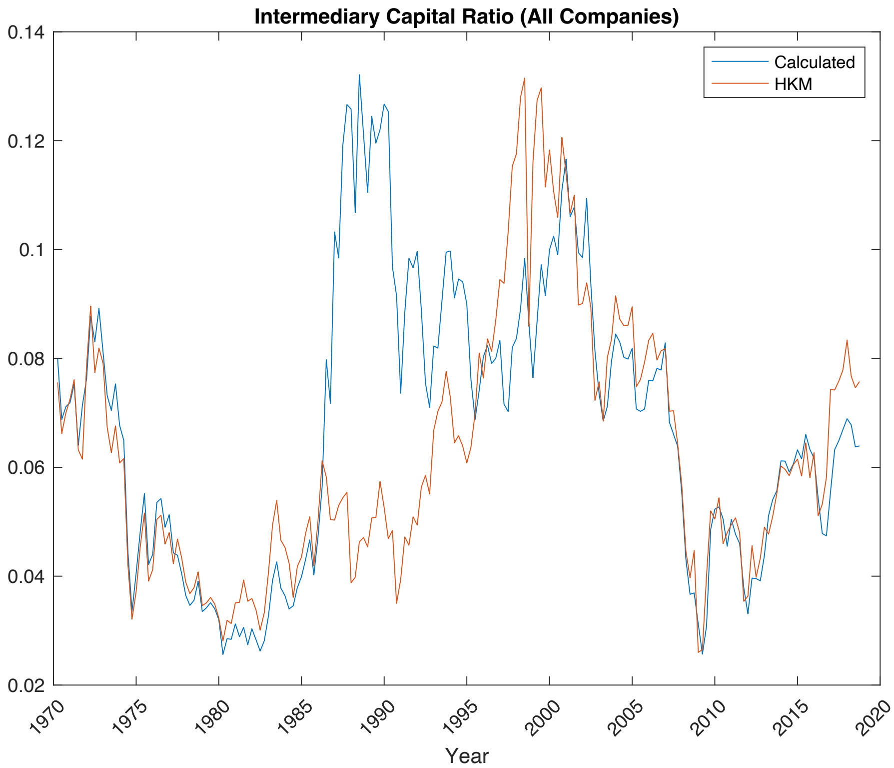


Figure C.2: Intermediary Capital Ratio (All Companies Minus Japanese)

The figure plots two capital ratio series from 1970:Q1 to 2018:Q3. HKM is the series of He et al. (2017) from Asaf Manela's website, while Calculated denotes the series based on our own calculations by excluding all of the Japanese ultimate parent companies.

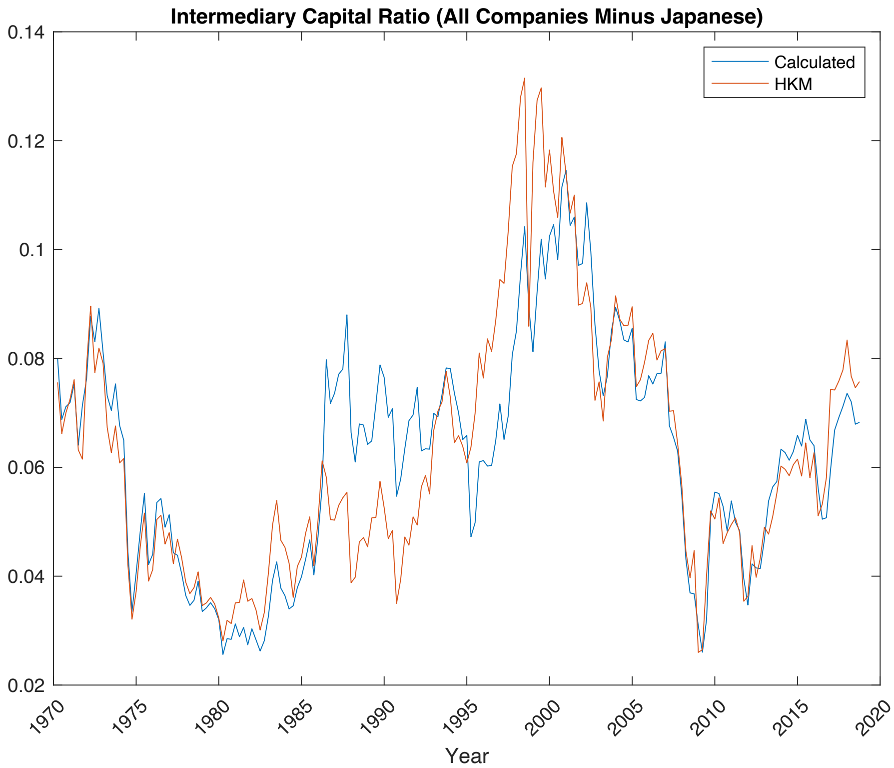


Figure C.3: Intermediary Capital Ratio (All Companies Minus Selected Companies)

The figure plots two capital ratio series from 1970:Q1 to 2018:Q3. HKM is the series of He et al. (2017) from Asaf Manela's website, while Calculated denotes the series based on our own calculations by excluding all of the Japanese and some other ultimate parent companies.

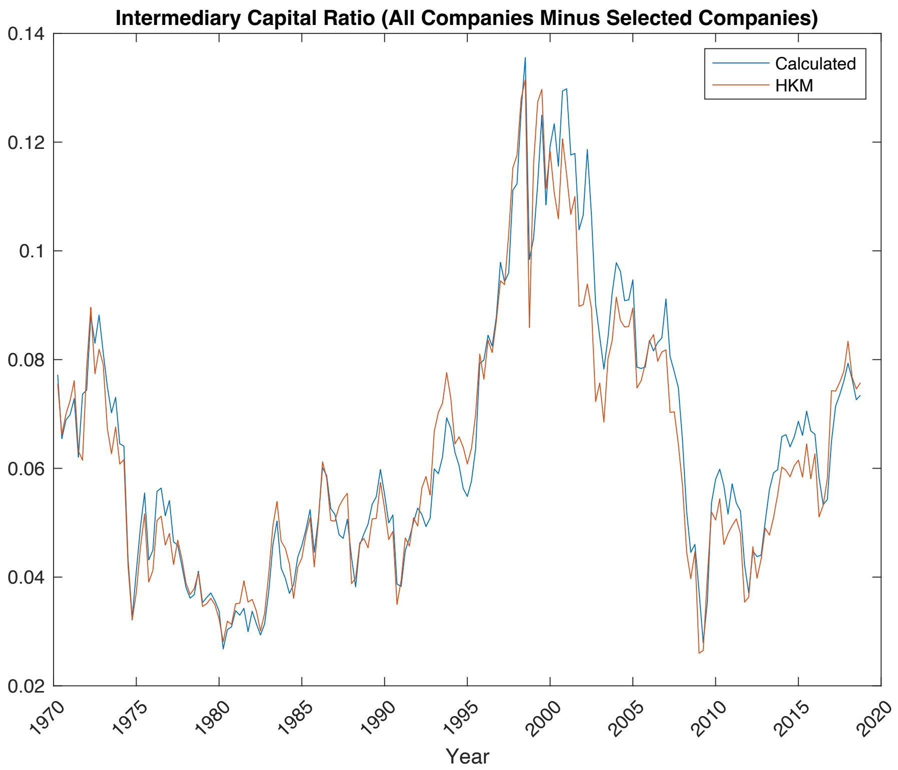


Figure C.4: Capital Risk Factor (All Companies)

The figure plots two traded capital risk factor series from 1970:Q1 to 2018:Q3. HKM is the series of He et al. (2017) from Asaf Manela's website, while Calculated denotes the series based on our own calculations and includes all domestic and foreign ultimate parent companies.

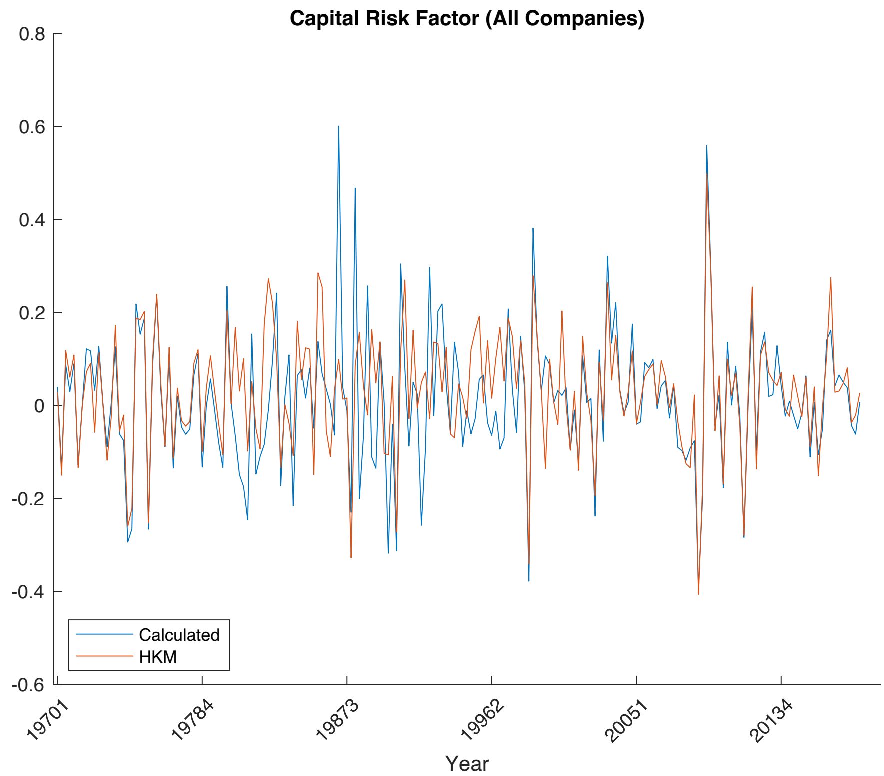


Figure C.5: Capital Risk Factor (All Companies Minus Japanese)

The figure plots two traded capital risk factor series from 1970:Q1 to 2018:Q3. HKM is the series of He et al. (2017) from Asaf Manela's website, while Calculated denotes the series based on our own calculations by excluding all of the Japanese ultimate parent companies.

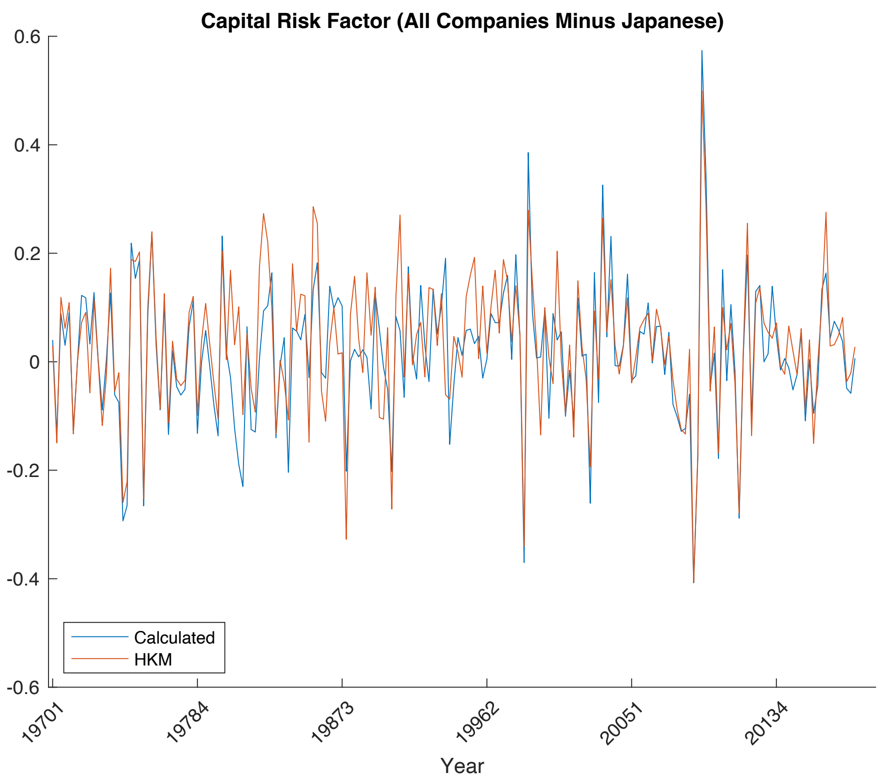
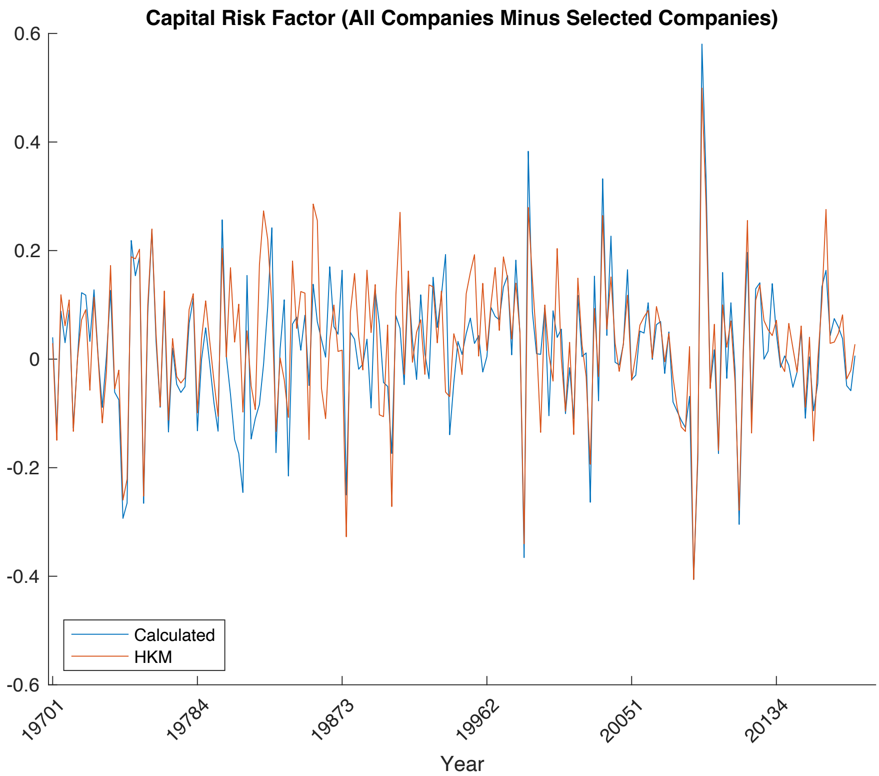


Figure C.6: Capital Risk Factor (All Companies Minus Selected Companies)

The figure plots two traded capital risk factor series from 1970:Q1 to 2018:Q3. HKM is the series of He et al. (2017) from Asaf Manela's website, while Calculated denotes the series based on our own calculations by excluding all of the Japanese and some other ultimate parent companies.



Bibliography

Adrian, T., Etula, E., and Muir, T. (2014). Financial intermediaries and the cross-section of asset returns. *Journal of Finance*, 69(6):2557–2596.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Wu, G. (2006). *Realized beta: Persistence and predictability.* Emerald Group Publishing Limited.

Antweiler, W. and Frank, M. (2005). The market impact of corporate news stories. Technical report, Working paper, University of British Columbia.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2):277–297.

Asness, C. and Frazzini, A. (2013). The devil in hml's details. *The Journal of Portfolio Management*, 39(4):49–68.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4):1645–1680.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Ball, R. and Kothari, S. P. (1991). Security returns around earnings announcements. *Accounting Review*, pages 718–738.

Bannouh, K., Martens, M., Oomen, R. C., and van Dijk, D. J. (2012). Realized mixed-frequency factor models for vast dimensional covariance estimation. *ERIM Report Series Reference No. ERS-2012-017-F&A.*

Barillas, F., Kan, R., Robotti, C., and Shanken, J. (2020). Model comparison with sharpe ratios. *Journal of Financial and Quantitative Analysis*, 55(6):1840–1874.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

Bollerslev, T., Law, T. H., and Tauchen, G. (2008). Risk, jumps, and diversification. *Journal of Econometrics*, 144(1):234–256.

Boudoukh, J., Feldman, R., Kogan, S., and Richardson, M. (2019). Information, trading, and volatility: Evidence from firm-specific news. *The Review of Financial Studies*, 32(3):992–1033.

Bradley, D. and Yuan, X. (2013). Information spillovers around seasoned equity offerings. *Journal of Corporate Finance*, 21:106–118.

Brunnermeier, M. K. and Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, 104(2):379–421.

Bybee, L., Kelly, B. T., Manela, A., and Xiu, D. (2020). The structure of economic news. Technical report, National Bureau of Economic Research.

Calomiris, C. W. and Mamaysky, H. (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics*.

Carlson, M., Fisher, A., and Giammarino, R. (2010). Seo risk dynamics. *The Review of Financial Studies*, 23(11):4026–4077.

Chan, W. S. (2003). Stock price reaction to news and no-news: drift and reversal after headlines. *Journal of Financial Economics*, 70(2):223–260.

Coles, J. L., Loewenstein, U., and Suay, J. (1995). On equilibrium pricing under parameter uncertainty. *Journal of Financial and Quantitative analysis*, 30(3):347–364.

Dang, T. L., Moshirian, F., and Zhang, B. (2015). Commonality in news around the world. *Journal of Financial Economics*, 116(1):82–110.

Denis, D. J. and Kadlec, G. B. (1994). Corporate events, trading activity, and the estimation of systematic risk: Evidence from equity offerings and share repurchases. *The Journal of Finance*, 49(5):1787–1811.

Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economics Statistics*, 13.

Eckbo, B. E., Masulis, R. W., and Norli, Ø. (2000). Seasoned public offerings: Resolution of the "new issues puzzle". *Journal of Financial Economics*, 56(2):251–291.

Engelberg, J., McLean, R. D., and Pontiff, J. (2018a). Anomalies and news. *The Journal of Finance*, 73(5):1971–2001.

Engelberg, J., Ozoguz, A., and Wang, S. (2018b). Know thy neighbor: Industry clusters, information spillovers, and market efficiency. *Journal of Financial and Quantitative Analysis*, 53(5):1937–1961.

Engelberg, J. E. and Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, 66(1):67–97.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020). Hedging climate change news. *The Review of Financial Studies*, 33(3):1184–1216.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366a):291–298.

European, U. (2017). Handbook on rapid estimates.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 71(3):607–636.

Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, 64(5):2023–2052.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis*.

Forbes, K. J. and Rigobon, R. (2002). No contagion, only interdependence: measuring stock market comovements. *The Journal of Finance*, 57(5):2223–2261.

Foucault, T. and Fresard, L. (2014). Learning from peers' stock prices and corporate investment. *Journal of Financial Economics*, 111(3):554–577.

Giannone, D. A. (2021). The effect of unscheduled news on systematic risk. *Working paper*.

Gospodinov, N. and Robotti, C. (2021). Common pricing across asset classes: Empirical evidence revisited. *Journal of Financial Economics*, 140(1):292–324.

Hameed, A., Morck, R., Shen, J., and Yeung, B. (2015). Information, analysts, and stock return comovement. *The Review of Financial Studies*, 28(11):3153–3187.

Hanley, K. W. and Hoberg, G. (2018). Dynamic interpretation of emerging systemic risks. *Working Paper*.

He, Z., Kelly, B., and Manela, A. (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics*, 126(1):1–35.

He, Z. and Krishnamurthy, A. (2012). A model of capital and crises. *Review of Economic Studies*, 79(2):735–777.

He, Z. and Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, 103(2):732–770.

Heston, S. L. and Sinha, N. R. (2017). News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal*, 73(3):67–83.

Hillert, A., Jacobs, H., and Müller, S. (2014). Media makes momentum. *The Review of Financial Studies*, 27(12):3467–3501.

Holden, C. W. and Jacobsen, S. (2014). Liquidity measurement problems in fast, competitive markets: expensive and cheap solutions. *The Journal of Finance*, 69(4):1747–1785.

Jagannathan, R. and Wang, Z. (1998). An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *Journal of Finance*, 53(4):1285–1309.

Jiang, G. J., Xu, D., and Yao, T. (2009). The information content of idiosyncratic volatility. *Journal of Financial and Quantitative Analysis*, 44(1):1–28.

Johnson, W. C., Kang, J.-K., Masulis, R. W., and Yi, S. (2018). Seasoned equity offerings and customer–supplier relationships. *Journal of Financial Intermediation*, 33:98–114.

Jurafsky, D. and Martin, J. H. (2008). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.

Kelly, P. J. (2014). Information efficiency and firm-specific return variation. *The Quarterly Journal of Finance*, 4(04):1450018.

Kleibergen, F. and Zhan, Z. (2015). Unexplained factors and their effects on second pass r-squared's. *Journal of Econometrics*, 189(1):101–116.

Lambert, R., Leuz, C., and Verrecchia, R. E. (2007). Accounting information, disclosure, and the cost of capital. *Journal of accounting research*, 45(2):385–420.

Larsen, V. H. and Thorsrud, L. A. (2017). Asset returns, news topics, and media effects. *Working Paper*.

Lewellen, J. and Nagel, S. (2006). The conditional capm does not explain asset-pricing anomalies. *Journal of financial economics*, 82(2):289–314.

Lewellen, J. W., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, 96(2):175–194.

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

Loughran, T. and McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4):1187–1230.

Manela, A. and Moreira, A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*, 123(1):137–162.

Massa, M., Rehman, Z., and Vermaelen, T. (2007). Mimicking repurchases. *Journal of Financial Economics*, 84(3):624–666.

Masulis, R. W. (1983). The impact of capital structure change on firm value: Some estimates. *The journal of finance*, 38(1):107–126.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Mimno, D., Wallach, H., Talley, E., Leenders, M., and McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 conference on empirical methods in natural language processing*, pages 262–272.

Morck, R., Yeung, B., and Yu, W. (2000). The information content of stock markets: why do emerging markets have synchronous stock price movements? *Journal of financial economics*, 58(1-2):215–260.

Myers, S. C. and Majluf, N. S. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, 13(2):187–221.

Nguyen, B. D. and Nielsen, K. M. (2010). The value of independent directors: Evidence from sudden deaths. *Journal of Financial Economics*, 98(3):550–567.

Otchere, I. and Ross, M. (2002). Do share buy back announcements convey firm-specific or industry-wide information?: A test of the undervaluation hypothesis. *International Review of Financial Analysis*, 11(4):511–531.

Patton, A. J. and Verardo, M. (2012). Does beta move with news? firm-specific information flows and learning about profitability. *The Review of Financial Studies*, 25(9):2789–2839.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, 22(1):435–480.

Peyer, U. and Vermaelen, T. (2008). The nature and persistence of buyback anomalies. *The Review of Financial Studies*, 22(4):1693–1745.

Phin, A., Prono, T., Reeves, J. J., and Saxena, K. (2018). Level shifts in beta, spurious abnormal returns and the tarp announcement. *Working Paper*.

Piotroski, J. D. and Roulstone, D. T. (2004). The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices. *The Accounting Review*, 79(4):1119–1151.

Ritter, J. R. (2003). Investment banking and securities issuance. In *Handbook of the Economics of Finance*, volume 1, pages 255–306. Elsevier.

Roll, R. (1988). R-squared. *Journal of Finance*, 43(2):541–566.

Savor, P. and Wilson, M. (2016). Earnings announcements and systematic risk. *The Journal of Finance*, 71(1):83–138.

Teoh, S. H., Yang, Y. G., and Zhang, Y. (2009). R-square and market efficiency. *Available at SSRN 926948*.

Tetlock, P. C. (2011). All the news that's fit to reprint: Do investors react to stale information? *The Review of Financial Studies*, 24(5):1481–1512.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3):1437–1467.

Vermaelen, T. (1981). Common stock repurchases and market signalling: An empirical study. *Journal of Financial Economics*, 9(2):139–183.

West, K. D. (1988). Dividend innovations and stock price volatility. *Econometrica: Journal of the Econometric Society*, pages 37–61.

Whaley, R. E. (2009). Understanding the vix. *The Journal of Portfolio Management*, 35(3):98–105.

Xing, X. and Yan, S. (2019). Accounting information quality and systematic risk. *Review of Quantitative Finance and Accounting*, 52(1):85–103.

Zhang, X. F. (2006). Information uncertainty and stock returns. *The Journal of Finance*, 61(1):105–137.