2024-08-28

Essays on Textual Information in Finance: Intangible Value, Abnormal Temperature Premium, and Corporate Uncertainty

Hosseini Golafshani, Seyed Amir Farhang

UNIVERSITY OF CALGARY

Essays on Textual Information in Finance:
Intangible Value, Abnormal Temperature Premium, and Corporate Uncertainty

by

Seyed Amir Farhang Hosseini Golafshani

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN MANAGEMENT

CALGARY, ALBERTA

AUGUST, 2024

Abstract

This thesis includes three studies that utilize textual information in their methodology.

The first chapter explores a textual measure of intangible intensity—intangibles talk—based on discussions in firms' 10-K filings. This measure reflects managers' views on the successful outcomes of intangible investments, which often differ from initial expectations. We find that our measure is correlated with, but orthogonal to, prior measures of intangibles. By constructing a long-short portfolio based on high and low values of our measure, we demonstrate that it outperforms portfolios formed on book-to-market ratio and its intangible–augmented versions, generating an average annual alpha of 3.37% from 1995 to 2020. Positive alphas are concentrated among stocks with higher arbitrage risk, suggesting intangible–intensive such stocks are often mispriced by investors.

The second chapter investigates the premium for exposure to abnormal temperatures among firms with a presence across U.S. states. Firm-level abnormal temperature exposure is estimated by extracting state names from firms' Form 10-K filings and averaging the reported abnormal temperatures across those states. Higher exposure to abnormal temperatures predicts lower earnings in five industries, with a stronger effect among firms with low geographic dispersion. A long-short portfolio based on abnormal temperature exposure shows that stocks with the highest exposure outperform those with the lowest, particularly post-Paris Agreement. The abnormal temperature premium averages 62 bps per month in this period and is concentrated in industries sensitive to abnormal temperatures and firms with low geographic dispersion. This premium responds positively to monthly shocks in a news-based index of climate concerns, indicating that investor climate concerns drive the premium.

The third chapter develops a textual measure of uncertainty within context–SECUX–using SEC filings and identifies the six most common sources of uncertainty: monetary policy, tax, financial market, exchange markets, financial intermediation, and oil & gas markets. I find that the predicted delaying effect of uncertainty on firm-level investment discussed in the literature only holds uncertainty related to monetary policy and oil & gas categories. No conclusive evidence supports delaying effects from tax, financial market, and exchange market categories, suggesting that investment response to uncertainty depends on the source of uncertainty.

Preface

All parts are original work that are not previously published. Chapter 1 of this thesis is a working paper with Professors Alexander David and Anup Srivastava entitled "Is Intangibles Talk Informative about Future Returns? Evidence from 10-K Filings." The remaining parts are my independent work.

Acknowledgements

As my years at the University of Calgary come to a close, I remember many individuals who supported me in overcoming the obstacles to complete my doctoral degree. Below I thank those who supported me the most.

Above all, I wish to express my sincerest thanks to Professor Alexander David, for his invaluable help during these years. He has consistently offered great economic intuition, thoughtful comments on my thinking and logical arguments, and constructive critiques of my research design and methodology. He also provided very generous financial support which is .

I am also grateful to the members of my committee. I extend my sincere thanks to Professor Miguel Palacios for his continuous support and encouragement, and to Professor Yrjo Koskinen for his insightful comments and guidance. I owe my deepest gratitude to Professor Anup Srivastava for his unwavering support from our first meeting and his commitment to improving my academic writing.

I want to thank my fellow doctoral students for their feedback, cooperation and of course friendship. Also, I would like to express my gratitude to the administrative staff at the Haskayne School of Business, in particular, Lesley DiMarzo, Professor Peter Sherer ,Mina Ojaghi, and Professor Tunde Ogunfowora.

Last but not least, I would like to thank my parents for encouraging and supporting me throughout my life.

To my mom and dad ...

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years, the finance literature is increasingly focusing on the analysis of textual information to capture the narratives that describe economic events, market conditions, and overall investor beliefs about the future. Leveraging the recent advancements from the field of natural language processing (NLP), researchers have been able to extract indices that add new dimensions to traditional financial and economic indicators. This thesis is an attempt to contribute to this literature by addressing three different topics in finance through leveraging established and intuitive textual analysis methods applied to the US Securities and Exchange Commission (SEC) filings. Annual and quarterly filings mandated by the SEC provide a most comprehensive and detailed summary of the firm's business and financial conditions on an ongoing basis. These filings commonly include a thorough discussion of risk factors that may impact any aspect of the firm's business along with forward-looking statements that reflect the management's assessment of value of intangible assets. Therefore the indicators developed in this study primarily benefit from the textual information closely reflecting firms' internal knowledge of their business and expectations regarding future economic conditions.

In the first chapter, we focus on the discussion of intangible assets in firm's 10-K filings. The study emphasizes the growing importance of intangible assets in the modern economy and their significant impact on firm value. Traditional accounting approaches, when applied to intangible assets face limitations that are associated with accounting criteria that prevent immediate capitalization of intangible expenses. Therefore the information regarding intangible assets is not readily available in financial statements, unlike their tangible counterparts. Various methods are developed to address this issue, mostly through capitalization of portion of intangibles expenses and accumulating it over time. However the proposed methods suffer from two limitations. First, no consensus exists on what percentage of intangible expenditures should be capitalized. Second, any capitalized intangible stock measure based on past expenses does not take into account the lottery-type payoff that often comes from serendipitous investments. These features of intangible assets lead to a potential mispricing of intangible intensive stocks that are discussed intensively in the literature. We contribute to this literature by constructing intangibles talk, a text-based measure that relies on intangibles-related keywords appearing in firms' 10-K filings. Our measure represents the ex-post value of successful intangibles investments that is not yet reflected in financial numbers. We argue that if intangibles talk contains some information about the value of the firm beyond numerical data in financial statements, it is expected to be informative about future returns.

We follow the long-short portfolio sorting methodology to measure the informativeness of intangibles talk against other indicators of value such as book–to–market and its intangible-augmented versions. Our results show that the portfolio constructed based on intangibles talk generates a significant positive annual alpha from 1995 to 2020 after controlling for various specifications of Fama and French factors along with momentum factor. We further test the mispricing hypothesis of explanation for the return informativeness of intangibles talk by relying on the literature that examines the effects of idiosyncratic volatility on mispricing of stocks. In line with the findings in this literature, we show that alphas of portfolios sorted on intangibles talk increase in idiosyncratic volatility and are only significant among high idiosyncratic volatility stocks. This confirms our hypothesis on mispricing of intangible intensive stocks as measured by intangibles talk, which points to the informativeness of our measure in comparison to accounting measures of intangible intensity in the literature.

In the second chapter, I examine the link between the economic effects of abnormal temperature and asset returns motivated by the recent observations on earnings sensitivity to extreme temperatures. My work primarily contributes to the literature investigating climate change risks and in particular physical climate risk which has received less attention in comparison to transition risk. In this study, I use the textual information in 10-K filings to determine the geographical presence of firms across the US states with the goal of measuring firm-level exposure to abnormal temperatures. I establish that exposure to abnormal temperature is a source of risk by first showing that earnings are negatively impacted by exposure to abnormal temperature using my measure. Next, I investigate the implications of differential exposure to abnormal temperature across firms for their stock returns. I argue that global warming creates a source of uncertainty due to its destabilizing effects on weather conditions which includes abnormal temperatures. In the presence of such uncertainty investors cannot predict the magnitude of future abnormal temperature based on historical temperature data and therefore prefer stocks of firms with lower exposure to abnormal temperatures. Accordingly, I hypothesize that firms with greater exposure to abnormal temperature should exhibit higher expected returns, as these stocks command a risk premium to compensate investors for the heightened exposure of their profitability to adverse impacts from abnormal temperatures.

Once again, I follow a long-short portfolio strategy based on the firm-level expected exposure to abnormal temperatures. The returns of this portfolio reveal that the abnormal temperature premium has grown rapidly in recent years, following the growing climate concerns. The portfolio generates a positive alpha between January 2015 and December 2021 after controlling for various specifications of Fama and French model along with momentum factor. This is expected since rise in climate concerns is expected to amplify concerns about the occurrence of larger abnormal temperature associated with global warming. This urges investors to demand a higher premium for exposure to abnormal temperatures. I also find that the premium only exists among the stocks belonging to industries that were demonstrated to have negative earnings sensitivity to abnormal temperature according to my earlier results. To further examine the direct effect of climate concerns on the abnormal temperature premium I use a proxy for climate concerns borrowed from the literature. In line with the results on the rapid growth of the abnormal temperature premium in recent years, I find that this premium positively responds to unexpected shocks in climate concerns. Therefore I conclude that the outperformance of stocks with higher exposure to abnormal temperature is explained by elevated climate concerns in recent years.

In the final chapter of this thesis, I examine the role of uncertainty in investment decisions using a simple approach that gauges uncertainty in the SEC filing along with the most common topics (sources) associated with it. My methodology aggregates the neighboring terms to uncertain words across the SEC filings and identifies the neighboring terms with the highest frequency across all filings. I find the six most common sources of uncertainty are monetary policy, tax, financial market, exchange markets, financial intermediation, and oil & gas markets. Next, b aggregating the neighboring terms under each category, I construct a quarterly measure of aggregate uncertainty corresponding to each of the six sources of uncertainty identified earlier. I then examine how the shocks to uncertainty series representing each source affect the firm-level investment in the economy. The literature on investment-uncertainty relation argues that investment activities are delayed in response to uncertainty in the presence of irreversibility of investment. My results, however, suggest that only uncertainty shocks associated with monetary policy and oil & gas markets have a delaying effect on investments, while rising uncertainty under other categories does not lead to delaying of investment activities at the firm-level. Therefore, my analysis contributes to the uncertainty–investment literature by showing that the source of uncertainty plays a significant role in the predicted delaying impact of uncertainty on investment which is predicted under the real options framework.

Chapter 2

Is Intangibles Talk Informative about Future Returns? Evidence from 10-K Filings

2.1. Introduction

The United States economy has experienced a dramatic shift toward intangible assets in recent decades. Investment in knowledge capital and organizational capacity among US firms has risen steadily (see Corrado, Hulten and Sichel, 2005; Eisfeldt and Papanikolaou, 2013; Enache and Srivastava, 2018; Ewens, Peters and Wang, 2019), allowing knowledge-intensive firms to launch new products and services, or gain a competitive edge in existing marketplaces, through innovation, lower costs, and better customer relations. Because the current US Generally Accepted Accounting Principles (GAAP) requires immediate expensing of internal outlays on intangibles, internally generated intangible capital is largely missing from the balance sheet, and no reliable measures of firms' total intangible capital exist. As a result, knowledge firms have more mispriced securities than do firms with physical assets (see Lev and Sougiannis, 1996; Daniel and Titman, 2006; Eisfeldt and Papanikolaou, 2013; Edmans, Li and Zhang, 2014). Meanwhile, the importance of intangibles in the US economy keeps increasing, as each new cohort of public firms spends more on intangibles than its predecessor cohort (Corrado, Hulten and Sichel, 2005; Srivastava, 2014, 2023).

Numerous studies attempt to address investors' problem via estimating the value of internally generated capital by first capitalizing and then amortizing outlays reported as research and development (R&D) and selling, general, and administrative (SG&A) expenses.[1] Edmans, Li and Zhang (2014) signifies the challenges of accurately defining and

[1]Those studies using the perpetual inventory model are Hulten and Hao (2008), Peters and Taylor (2017), Lev and Srivastava (2022), Eisfeldt,

gauging intangible value using these methods. We extend this literature by proposing a new, text-based measure that relies on intangibles-related keywords appearing in firms' 10-K filings following the glossary created by Filipovic and Wager (2019). Our method could reflect ex-post, firm-specific successful outcomes of intangibles investments, which cannot be fully captured in the capitalization of ex-ante investments using uniform capitalization and amortization parameters. We demonstrate the informativeness of our measure through alphas generated by long-short portfolios formed on our measure while controlling for momentum and Fama and French factors.

US GAAP requires that expenditures on internally generated intangibles be immediately expensed. The same GAAP rules permit capitalization of expenditures on property, plant, and equipment (PP&E) and acquired intangibles. As a result, information on in-house-developed intangible assets such as innovation, knowledge, and brand capital are not readily available in financial statements compared with data related to tangible capital (Belo et al., 2022). To address this accounting limitation, numerous studies in the finance and accounting literature have focused on improving measures of, and proposing new methods to estimate, intangible capital (see Peters and Taylor, 2017; Enache and Srivastava, 2018; Park, 2019; Eisfeldt, Kim and Papanikolaou, 2020; Lev and Srivastava, 2022). Most of those studies rely on a perpetual inventory model (PIM), that is, capitalizing and amortizing past R&D and SG&A expenditures, using uniform or industry-specific parameters. These methods are based on the expectation that a fixed percentage of investments creates value for all firms and depreciates predictably over time. For example, 30% of SG&A creates future value and has a life of three years.

While the new methods, based on PIM, yield improvements over investment models that ignore in-house intangibles, they suffer from two limitations. First, no consensus exists on what percentage of intangible expenditures should be capitalized. Capitalization percentages used by those studies range from 30% to 100% (Iqbal et al., 2023). Second, any capitalized intangible stock measure based on past expenses does not take into account the lottery-type payoff that often comes from serendipitous investments. For example, the discovery of a search formula by Google founders led to a trillion-dollar valuation company, and no amount of capitalization of past expenses would yield a number close to the value of that discovery. Furthermore, these methods do not consider the fact that most intangible investments do not create value. Thus, these methods could suffer from both type-1 and type-2 errors. Stated differently, perpetual inventory models, even if correctly specified, would not capture the lottery-type outcome from initial investments. We rely on the idea that managers are more likely to describe the successful discoveries and self-developed intangible assets in their communications with investors when those assets are expected to create benefits for the company, while ignoring investments that did not yield anything.

We contribute to the literature by identifying new measure of intangibles, which arguably tracks developed intangibles, that is, ex-post outcomes, but are not yet reflected in financial numbers presented ex-ante in income statement numbers and ex-post in balance sheet. We extract the informational content relating to intangible capital embedded in

Kim and Papanikolaou (2020), Iqbal et al. (2023), and Falato et al. (2022)

the text of a firm's 10-K filings. The reason for differences in information presented in text and numbers as follows. Financial Accounting Standards Board (FASB) Concepts Statement No. 5 prescribes strict criteria for recognition transactions in financial numbers, such as measurability, relevance, and reliability. Items that fail these criteria but are value-relevant nevertheless are often disclosed in footnotes and in the management discussion and analysis (MD&A) section of the 10-K. Managers describe their assessments of items that will impact future operations and whose discussion will enhance investors' understanding of firm operations. Any forward-looking information supplied in the MD&A section is expressly covered by the safe harbor rule, a legal provision that shields managers from liability if future projections go wrong. Merkley (2014) finds that narrative R&D disclosure in 10-K filings is positively associated with earnings forecast accuracy and predicts lower analyst forecast dispersion, suggesting that narrative disclosures reduce information asymmetry. Textual portion, thus, is particularly useful for conveying soft information (Seamons and Rouse, 1997). Hence, the textual portion of 10-K filings could provide guidance to investors on the value of internally generated intangible capital, particularly the value not extractable from financial numbers because of constraints imposed by accounting rules.

We conduct textual analysis to gauge the intensity of intangibles discussion in 10-K filings and consider it a proxy for the value of intangible capital. Our measure is the relative frequency of words on intangibles topics to total words in 10-K filings. Scaling the frequency of intangibles words by total words allows for comparison across firms with varying document sizes. We focus on three distinct categories of intangible assets: innovation assets and information technology, brand and customer relations, and human resources following (Lev, 2000; Corrado, Hulten and Sichel, 2005). We construct our measure of intangibles talk using a glossary of intangible terms created by Filipovic and Wager (2019). An additional benefit of using a text-based measure, relative to capitalizing past expenses, is the possibility of mapping words to and classifying them in, the three categories, while separately analyzing their individual informativeness with respect to future returns.

In testing the validity of our measure, we find that intangibles talk positively correlates with conventional accounting measures of intangible investments such as R&D and SG&A expenditures (both scaled by total expense). Also, intangibles talk correlates positively with indicators of intangible value called intangible capital advanced by Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020). Furthermore, our measure has a positive correlation with market-to-book ratio, which prior literature considers as a proxy for intangible value not recognized on balance sheet. Considering that intangibles talk is constructed based entirely on textual material, as opposed to market or book data, these correlations point to a strong link between a firm's intangibles discussions in 10-K filings and its underlying intangible capital. We view this link as preliminary evidence that intangibles talk tracks variations in intangible capital and intensity over time.

We analyze variations in our measure across firms in Fama and French twelve industries. As expected, industries such as health care and business equipment score the highest in intangibles talk, and finance and energy industries score the lowest. We also decompose intangibles talk into three categories and investigate the highest-ranking industries under each of them. Results are consistent with intuition. For instance, while the healthcare industry scores the highest on intangibles talk focused on innovation assets and human resources, it scores among the lowest in the brand and customer relations category.

Our main tests of intangibles talk are informativeness with respect to future stock returns based on the idea that future returns, if any, would reflect underpriced intangibles information or risk. Classic studies, as well as recent studies, investigate the mispricing of intangibles information. Lev and Sougiannis (1996) find a systematic mispricing of R&D-intensive stocks and show that incorporating that information leads to an annual return of 4.57%. Chan, Lakonishok and Sougiannis (2001) find that stocks with high R&D relative to the market value of equity deliver an average of 6.12% annual returns. They show similar results for stocks with high advertising expenses. Other studies link excess returns to patent citations (Deng, Lev and Narin, 1999), software developments (Aboody and Lev, 1998), and employee satisfaction (Edmans, Li and Zhang, 2014). Our argument on the link between intangibles talk and future returns, is similar to Edmans (2011), which points to the insufficient salience of intangibles information that could lead to it being overlooked by investors. Recent studies (Arnott et al., 2021; Choi, So and Wang, 2021; Lev and Srivastava, 2022) augment the book-to-market measure of value investing by intangible estimates and show superior returns than book-to-market based high minus low (HML) returns based on Fama and French (2015).

Prior studies on disclosure versus recognition argue that textual content on intangibles in documents such as 10-K filings are more likely to be ignored by investors than are numbers reported in income statement and balance sheet (Aboody, 1996; Davis-Friday et al., 1999; Ahmed, Kilic and Lobo, 2006). Based on this idea, we hypothesize that intangibles talk could convey at least some informative signals that are ignored by investors. Eisfeldt, Kim and Papanikolaou (2020) show that intangible augmented value factors outperform the traditional value factor. They contribute to the literature by pushing the limits of the available accounting data to capture intangible value. We follow this idea to investigate our hypothesis. We first examine whether long-short portfolios formed on intangibles talk can generate positive, risk-adjusted returns. We then compare these results with returns generated by other benchmark value signals, namely, book-to-market ratio and its intangible augmented versions from Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020).

We begin our analysis by sorting portfolios based on intangibles talk, INT^{10K}. We follow the long-short sorting methodology presented in Fama and French (2015) and construct HML^{FF}, which captures the traditional value strategy based on book-to-market ratio. Similarly, we construct HML^{PT} and HML^{EKP} using the intangible augmented book-to-market ratio based on Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020), respectively.

We find that INT^{10K} outperforms the above three value strategies for the sample period, covering July 1995 to June 2020, while excluding the years that represent the dot-com bubble burst (2000–2001).[2] We achieve an average

[2] We argue that the weak historical performance of INT^{10K} near the burst of the dot-com bubble results from the massive overvaluation

annualized returns return of 5.05%. The returns are particularly high at around 7.6% in the period from 2008 to 2020. The average returns are consistently positive for each year since 2007, with the exception of 2012 and 2016 (see Fig. 2.1). INT^{10K} performs most strongly during the period that records the worst performance of HML^{FF} as shown in the post-financial crisis era (2008–2020) (see Eisfeldt, Kim and Papanikolaou, 2020). Our graph of cumulative returns is consistent with the growing importance of intangible value (see Fig. 2.2). Our portfolio's return displays steady growth over time and reaches its highest levels by the end of our study period in 2020. INT^{10K}'s outperformance relative to HML^{FF}, HML^{PT}, and HML^{EKP}, especially in recent years, suggests that incorporating textual information holds promise as a separate source of intangible value for investors besides the data presented in accounting numbers. Furthermore, the fact that returns can be generated using textual data indicates that text-based intangibles information is not fully incorporated by the investors.

We conduct a more detailed examination on the relative informativeness of intangibles talk against other value indicators. We follow the strategy in Eisfeldt, Kim and Papanikolaou (2020); that is, we go long on INT^{10K} and short other intangibles-enhanced value portfolios. This strategy shows that our measure captures orthogonal, and perhaps superior, information compared with augmented value strategies. We find that our long-short portfolio generates significant positive returns over HML^{FF} except for the period after the dot-com bubble (2000–2007). We find similar results when the short leg of the portfolio is changed to HML^{PT} and HML^{EKP}, providing strong evidence that our measure has additional, and arguably more unpriced, information than intangible augmented versions of value indicators documented in prior studies.

We next test whether the returns generated by intangibles talk represent premiums for risk or some other unpriced factor. We generate positive alphas while controlling for momentum (Carhart, 1997) and the three as well as the five Fama and French factors. The alpha averages 3.37% and 5.81% from 1995 to 2020 in the three- and the five-factor setting, respectively. The alpha remains positive and significant when we replace HML^{FF} with portfolios HML^{PT} and HML^{EKP} as the value factor in the Fama and French regression models. This suggests that the INT^{10K} return cannot be explained by traditional risk measures. We also create five value-weighted portfolios ranked based on intangibles talk and show that the alpha of the highest-ranking portfolio minus the lowest-ranking portfolio is 7.03% and significant.

We further examine the possibility that intangibles talk is an unpriced systematic risk factor by conducting a Fama-MacBeth two-stage regression analysis from Fama and MacBeth (1973). In the first stage, we estimate the factor loadings of a set of test portfolios on the returns of the portfolio sorted based on intangibles talk. In the second stage, we test whether these factor loadings explain the variations in the returns of test portfolios. Our results suggest that intangibles talk is not a consistently systematically priced risk factor in the cross-section of stock returns. This

of technology stocks in previous years. Technology stocks typically score high in intangibles talk and thus are likely picked up by our sorting methodology, which relies on intangibles talk values.

result is inconsistent with risk explanation.

We also test mispricing as an explanation for return informativeness of intangibles talk. We rely on the idea that higher idiosyncratic volatility (IVOL) amplifies returns associated with mispriced stocks. This idea is based on the reasoning that higher IVOL leads to greater arbitrage risk, which, in turn, limits the ability of rational investors to correct mispricing (Pontiff, 1996; Stambaugh, Yu and Yuan, 2015)[3] In the case of intangible-intensive stocks, under higher IVOL market conditions, rational investors would bid less aggressively against overlooking of intangibles information by less sophisticated investors. Accordingly, we argue that mispricing of intangible-intensive stocks because of overlooked information, if any, would be amplified under market conditions with higher arbitrage risk. To test this, we first measure firm-level IVOL by estimating the volatility of the residuals from the three factors of Fama and French for daily returns (Ang et al., 2006). We then sort stocks by IVOL and estimate long-short returns based on our measure in each IVOL-sorted portfolio. We find that alphas of portfolios sorted on intangibles talk increase in IVOL, are positive and significant for high IVOL stocks, and are insignificant at low IVOL stocks. We therefore conclude that intangible-intensive stocks experience greater mispricing under high arbitrage risk, as evidenced by the strongest INT 10K abnormal returns among high IVOL stocks.

Our study contributes to the emerging stream of literature, aiming to create alternative measures of intangible intensity, and using them to earn investment returns. Our study shows that managers could convey their value of intangible assets, through discussions and disclosures, particularly the value that is not represented in balance sheet and is not interpretable through capitalizing past investments. Our results, however, demonstrate that such information is often mispriced by investors, especially in stocks characterized by high arbitrage risk.

In addition to contributing to intangibles-based value literature, our study is closely related to the literature on using textual analysis to predict returns (see Tetlock, 2007; Garcia, 2013; Jiang et al., 2019). We rely on the bag–of–words approach in our textual analysis, which is becoming a standard method used in the literature since Loughran and McDonald (2011). Most studies in the textual analysis literature measure tone sentiment or uncertainty. Our study in this regard deviates from such studies and is closer to those that use specialized glossaries to gauge the intensity of discussion surrounding particular topics.

This paper proceeds as follows. In Section 3.2, we describe the methodology used to construct intangibles talk and discuss and report its values across industries and in relation to other firm characteristics. Section 2.3 presents the results from our portfolio analysis. Section 2.4 examines whether the common risk factors explain the portfolio returns. Section 2.5 tests the idea that intangibles talk and other measures of intangible intensity are unpriced systematic risk factors. Section 2.6 compares the performance of the portfolio sorted based on intangibles talk with technology stocks. In Section 2.7, we present and test the mispricing hypothesis. Section 2.8 concludes.

[3]They show that among overpriced (underpriced) stocks, the ones with the highest IVOL are the most overpriced (underpriced). Another recent study, Birru and Young (2020), utilizes IVOL to show stronger return predictions from investor irrationality.

2.2. Data and methodology

2.2.1 SEC filings

The corpus or textual data for our analysis come from 10-K filings submitted to the SEC by 12,184 public firms from January 1994 to December 2021. This adds up to a total of approximately 107,000 10-K filings in our analysis. The filings are collected using the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. [4] In the parsing process, numbers, tables, and figures are removed for the raw text to be ready for textual analysis. To reduce the noise in the text, we remove all the stop words.[5]

To construct intangibles talk, we use the bag-of-words approach. A filing's vector of words is analyzed against a glossary of intangibles terms (see Table 2.1) based on the recently developed intangibles words list in Filipovic and Wager (2019). See the Appendix for examples. The list is derived from several studies on intangible assets such as Hall (2009) and Lev (2005, 2012). A more detailed description of how our intangibles word list was developed can be found in Filipovic and Wager (2019).

We focus on three broad categories of intangible assets: innovation assets and information technology, brand and customer relations, and human resources. Lev (2005) claims that innovation assets have become synonymous with information technology in recent years with the rise of Internet platforms and software solutions. To account for this aspect, we combine innovation assets and information technology in a category that Wyatt (2008) calls technology resources. Table 2.2 shows that words under the innovation assets and information technology category include terms like databases, websites, and platforms.

Table 2.2 contains the word list belonging to the other two categories of intangibles. The brand and customer relations category, includes reputation, brands, and relations. The last category, human resources, emphasizes skills, abilities, and competencies. Table 2.3 shows that most of the variation in intangibles words frequency across filings is related to the innovation assets and information technology category. This category contributes 47% to the total variance, and the words belonging to it account for 54% of the words in the glossary.

We define intangibles talk for 10-K filing i with n intangibles words as the sum of the frequency across all intangibles words divided by the total number of words in the filing:

$$Intangibles\ talk_{\mathbf{i}} = \frac{\sum_{\mathbf{j=1}}^{\mathbf{n}} Frequency\ of\ intangible\ word_{\mathbf{j}}}{Total\ words_{\mathbf{i}}} \tag{2.1}$$

The ratio, measured on a firm-year basis, reflects the relative intensity of discussion surrounding different intangibles categories in filings. As Filipovic and Wager (2019) assert, the intangibles word list is neither optimized nor reverse-

[4]We use the parsed documents publicly available on the Loughran-McDonald website at: https://sraf.nd.edu/loughranmcdonald-master-dictionary/

[5]Examples of stop words are "a," "the," "are," "and," "could," and "would."

engineered to fulfill return maximization objectives. In a similar fashion, we do not select words from the intangibles word list that maximize our portfolio returns using machine learning techniques.

Fig. 2.3 plots the top 30 most frequent intangibles words across firms in our sample. Some of the most common words, such as ”employee,” ”customers,” and ”data,” can be found under various discussion topics that are not always related to intangible assets. Our glossary contains 128 words, of which the top three in Fig. 2.3 account for almost 20% of the total frequency while their share of the total frequency. Their share of the total frequency would be 2.3% under the equally distributed case. This illustrates the power law distribution of words frequency in natural languages, also known as Zipf’s law.[6]

2.2.2 Intangibles talk across industries

Fig. 2.4 depicts a summary of intangibles word frequency along with intangibles talk across the twelve Fama and French industries. Healthcare and business equipment rank the highest based on the median values of intangibles talk. The lowest-ranking industries are finance and energy. Intangible intensity in the healthcare and business equipment industries likely reflects their principal sources of firms’ competitive advantage, such as patents, data, technical knowhow, and information technology. In contrast, finance and energy rely heavily on tangible assets and financial capital. Later, we delve deeper into each industry’s most frequent intangibles words to identify the main components of intangibles talk in that industry. Fig. 2.4, shows that the absolute frequency of intangibles words is high for some industries and the relative frequency (intangibles talk) can be low in some cases because of varying document sizes across industries.

The portion of intangibles talk’s three categories across the twelve industries reveal the concentration of each class of intangible assets in the economy. Fig. 2.5 shows that the consumer non-durables industry ranks high in the brand and customer relations category, indicating the importance of brands, sales and distribution network, and customer satisfaction in that industry. The healthcare industry ranks relatively low in the brand and customer relations category but ranks high in the human resources and innovation assets categories. This points to the central role that qualified professionals play in the healthcare industry to deliver quality and efficient services. In addition, patents for new drugs and medical devices are important sources of revenue in the pharmaceutical industry.

Another important feature of our measure is that it could indicate the cross-sectional variation of intangible value across firms and industries. Fig. 2.6 plots the top ten most frequent intangibles words by twelve industries. Words such as ”software”, ”data”, and ”technology,” that are associated with information technology appear to be frequent and concentrated in the business equipment industry. Another word, ”employee,” appears as the most frequent word across most industries. Admittedly, the context in which common words such as ”employee” and ”data” are discussed is a

[6] According to Zipf’s law the frequency of words is proportional to the inverse of their ranking: f(r) $\propto \frac{1}{r^{\alpha}}, with \alpha \approx 1$ (see Mandelbrot, 1961).

prerequisite to classify them as indicators of intangibles value. Our measure suffers from this limitation that it does not identify the context in which the words are used. Nonetheless, when a group of words that refer to a particular category of intangible assets is concentrated in one industry, it likely indicates the importance of that category to that industry's firms. For instance, the word "customer" is relatively common across the majority of industries. Closely related words such as "brand," "advertising," "franchise," and "marketing" are less common and are concentrated in industries such as consumer non-durables and wholesale and retail. This indicates the importance of brand and customer relations in consumer goods industries.

Overall, the textual analysis shows that emphasis on different categories of intangible assets varies across industries, as captured by our measure. The variance also aligns with the expected industry characteristics, consistent with intuition.

2.2.3 Validity of intangibles talk measure

In this section, we describe the relationship between intangibles talk and firm characteristics. Table 2.4 reports the time series average of median firm characteristics for firms in the high, middle, and low range of intangibles talk. Prior studies consider R&D to total expenses, SG&A to total expenses, and enhancements in book value because of capitalized intangibles as proxies for intangible intensity.[7] We benchmark our measure against these proxies, by examining whether they vary in predicted directions across quantiles of firms formed based on our measure. All three proxies exhibit an increase by quantiles based on intangibles talk with notably large values amongst the firms in the highest quantile.

R&D to total expenses shows the highest variation with our measure in the innovation assets and information technology category, with a jump from 0.01 to 0.13 between the second and third quantiles. The increase from the second to third quantile is sizable for all three proxies of intangible intensity: R&D, SG&A, and intangible capital. This shows consistency between our intangibles talk measure and proxies considered in the literature.

We also examine variations in firm characteristics, such as size, leverage, and profitability. No clear relationship is discernible between intangibles talk and sales-to-asset ratio, which represents asset turnover or efficiency in utilization of assets. However, debt to total assets falls with intangibles talk, showing that firms in the high quantile are the least leveraged. The opposite holds for the profitability-to-total assets ratio, with its highest value in the third quantile of intangibles talk. This aligns with the research on the relation between R&D and profitability (Lev and Sougiannis, 1996). However, it is inconsistent with Curtis, McVay and Toynbee (2020) who document a decline in the relation

[7] R&D and SG&A are reported as expenses and occur over the normal course of business operations (i.e., flow variables). We divide them by total expenses instead of total assets to capture the variations in the flow of intangible investment over each year. This allows us to compare SG&A and R&D across firms with large and small asset bases.
Intangible capital which is constructed based on the methods in Peters and Taylor (2017) (intangible capital^{PT}) and Eisfeldt, Kim and Papanikolaou (2020) (intangible capital^{EKP}), is a stock of knowledge and organizational capital that accumulates over time through investments reported in SG&A and R&D. We use both versions of intangible capital in our analysis. Intangible capital^{PT} and Intangible capital^{EKP} are available at the GitHub page affiliated with Eisfeldt, Kim and Papanikolaou (2020): `https://github.com/edwardtkim/intangiblevalue`

between R&D and profitability since the 1980s.

We also examine the traditional book-to-market ratio, along with its intangible-augmented versions put forth by Peters and Taylor (2017) (book-to-market$^{\text{PT}}$) and Eisfeldt, Kim and Papanikolaou (2020) (book-to-market$^{\text{EKP}}$).[8] We observe that book-to-market ratio drops along the quantiles, with the lowest book-to-market ratios concentrated in the top quantiles of intangibles talk. This shows consistency between our measure of intangible value and the one implied by the book-to-market ratio (which is the opposite of market-to-book ratio). With respect to intangible-augmented book-to-market ratios, their highest values are in the lowest quantile of intangibles talk, with less variation across quantiles relative to the as-reported book-to-market ratio itself. This arguably indicates that the intangibles-augmented version of the book-to-market ratio is a less correlated with our measure of intangible value than the as-reported one, because the book value has already been at least partly corrected for intangibles.

We provide further evidence that intangibles talk is correlated with other indicators of intangible value in our panel regression analysis in Table 2.5. One standard deviation drop in the book-to-market ratio on average is associated with an approximately 0.18 standard deviation increase in intangibles talk. To capture the relation between SG&A and intangibles talk, we subtract R&D from SG&A because R&D is a constituent item of SG&A (Enache and Srivastava, 2018). As expected, our results show that SG&A (without R&D) and R&D, both scaled by total expenses, are positively related to intangibles talk. A standard deviation increases in SG&A and R&D to total expense ratios corresponds to about a 0.3 standard deviation rise in intangibles talk. A similar result holds for intangible capital based on both Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020).

We view the results from Table 2.5 as evidence that intangibles talk and conventional indicators of intangible value are strongly correlated. This strengthens our assumption that intangibles talk is a proxy for intangible value and intensity across firms. We consider whether our measure conveys additional information on future returns than do current proxies.

2.3. Returns analysis

We investigate the informativeness of intangibles talk by analyzing portfolio returns over time while controlling for the common risk factors. We use factor-mimicking portfolios for our inquiry and follow the long-short sorting method, described in Fama and French (2015), to construct a value-weighted portfolio. We use intangibles talk as an investment signal to identify the long and short portfolios every year in a sample of NYSE, Amex, and Nasdaq stocks, with data available from the Center for Research in Security Prices (CRSP). We analyze the period starting from 1995 to June

[8] The formula is book-to-market$^{\text{PT}}$ = (book value of equity + intangible capital$^{\text{PT}}$ + goodwill)/market value of equity. Similarly, book-to-market$^{\text{EKP}}$ uses intangible capital$^{\text{EKP}}$ in the formula.

2020. [9] In this section, we describe our sorting methodology and long-short portfolio performance. We then conduct numerous analyses using several subperiods, such as before and after the dot-com bubble burst and the global financial crisis. We then benchmark the returns of our portfolio against the value strategies based on book-to-market ratio, and its augmented versions with capitalized intangibles, studied in prior literature.

2.3.1 Intangibles talk and value strategies

We follow Fama and French in constructing long-short portfolios except that we use intangibles talk as the sorting variable instead of book-to-market ratio. We sort firms as of June 30 of each year, based on intangibles talk for the fiscal year that ended in the previous calendar year. This method assumes that at least the December fiscal year-end firms have published their annual report by June 30 of the next year, consistent with Fama and French (2015). The portfolio (INT^{10K}) is constructed based on last reported intangibles talk. We identify stocks above the 70th percentile of intangibles talk, and put them in the long portfolios while shorting those below the 30th percentile.[10] Portfolios are held constant from July 1, following the June 30 portfolio formation date, to June 30 of the next year. We use delisting returns in the month that a stock delists. Monthly returns are calculated for each long and short portfolio, by value weighting returns of their constituent securities using their share in total market cap at the end of December of the previous year. Annual returns are calculated using monthly returns from January to June, based on portfolios formed in June of the previous year, and from July to December, based on portfolios formed in June of this year.[11]

We first compare the performance of INT^{10K} against portfolios formed based on book-to-market ratio and its two intangible-augmented versions (Peters and Taylor, 2017; Eisfeldt, Kim and Papanikolaou, 2020). In this analysis, HML^{FF} represents a value strategy that takes into account only reported value of assets. HML^{PT} and HML^{EKP} incorporate book value enhanced by non-reported intangible assets estimated with capitalization of past intangible investments. Notably, the definition of intangible capital used to calculate the augmented versions differs between Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020) and so do the returns generated by HML^{PT} and HML^{EKP}.

Performance statistics in Table 3.5 report the average monthly returns of the four portfolios for the period between July 1995 and June 2020. None of the returns is statistically different from zero on average. Nevertheless, the returns from INT^{10K} are the highest, and the Sharpe ratios suggest that HML^{EKP} and INT^{10K} provide the best-performing strategies. Fig. 2.2 shows that the returns from INT^{10K} drop around the dot-com bubble, arguably because

[9] Given data requirements from the previous year, our portfolio sorting method generates returns from July 1995 based on intangibles talk values available from 1994.

[10] A more detailed description of the sorting method can be found in Fama and French (2015). Using the NYSE median market cap as the breaking point, for the long leg of our portfolio, we average the returns of two portfolios, namely big and small stocks with high intangibles talk, and repeat the same procedure for the short leg of our portfolio.

[11] We test the validity of our sorting methodology, by constructing a portfolio based on book-to-market ratio as defined in Fama and French (2015). We achieve a 95% correlation with the HML^{FF} returns reported in the Kenneth R. French data library: (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). We use the same sorting methodology to construct other portfolios described in this study based on various firm-level intangible value indicators including our measure.

many intangible-intensive companies suffered large negative returns, while numerous others were delisted. We also report the portfolio performances excluding years 2000 and 2001, the peak bubble years. INT^{10K} now outperforms all the value strategies with statistically significant average annualized returns of 5.05%.

A more detailed examination of the subperiods is reported in the last three columns of Table 3.5. The performance of INT^{10K} is stronger than other value strategies from 1995 to 1999, but no portfolio shows any statistically significant returns. During the pre–financial crisis period from 2000 to 2007, INT^{10K} performs poorly. The returns of HML portfolios are positive with high Sharpe ratios but still are not statistically significant from zero. Examining the post–financial crisis period (after 2008) is important because HML^{FF} gave negative returns year after year during this period (Lev and Srivastava, 2022). INT^{10K} delivers the highest returns between 2008 to 2020 with a statistically significant return of 7.63% and a Sharpe ratio of 0.92, the highest amongst all four strategies. The results support two ideas: the growing importance of intangible value in the economy and our ex–post measure of intangibles talk capturing intangible capital better than the ex-ante capitalization methods of Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020).

Results based on intangibles talk follow the pattern presented in Eisfeldt, Kim and Papanikolaou (2020), showing that value investing based on intangible-augmented book-to-market ratio outperforms HML^{FF}, particularly in the post–financial crisis era. More importantly, the results demonstrate that our measure captures orthogonal, and arguably superior, information on intangibles compared with the ones based on capitalization of past expenses.

2.3.2 Other intangible intensity indicators

We turn to other indicators of intangible value from the literature and compare their portfolio performance against the performance of INT^{10K}. We create four portfolios based on four intangible proxies: SG&A expenses and R&D expenses, and both versions of intangible capital from Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020). We use intangible capital based on capitalized investment (scaled by total assets) to create the last two portfolios, not the book-to-market ratios augmented with intangibles. In constructing intangible capital, Peters and Taylor (2017) already account for R&D and SG&A expenses to a degree, and we include the two variables separately as well. We do this to account for any direct intangible value signal obtainable from these items that could be informative about future returns and is lost in book-to-market augmentation. We sort portfolios as of the end of June every year based on the ratio of R&D and SG&A to total expenses reported at the end of the last fiscal year. Similar to Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020), we sort portfolios based on intangible capital as of the end of the fiscal year in the previous calendar year. We divide these measures of intangible capital by total assets at the beginning of that fiscal year. We name them INT^{PT} and INT^{EKP} to distinguish them from HML^{PT} and HML^{EKP}. This exercise additionally allows us to compare the returns of portfolios sorted on other measures of intangible intensity, which is a more comparable strategy to INT^{10K}.

The returns of these four portfolios, along with INT^{10K} are reported in Table 2.7. In the full sample between 1995 and 2020, none of the portfolios generates statistically significant returns. Similar to intangible augmented HML returns, when we exclude the years around the dot-com bubble, most strategies perform better, with INT^{10K} slightly outperforming the rest. Columns 3 to 5 do not report any significant outperformance by any of the portfolios, and the period from 2008 to 2020 presents significant positive returns for the INT^{10K}, INT^{PT}, INT^{EKP}, and $R\&D$ portfolios. The returns of INT^{10K}, and INT^{PT} are essentially the same, with INT^{10K} having a slightly larger Sharpe ratio. Therefore, the results from Table 2.7 show that intangibles talk is at least as informative about future returns as the other measures of intangible intensity in the literature, if not superior.

2.3.3 Returns across categories of intangibles talk

We break down intangibles talk into its three categories based on the words associated with each category (see Table 2.2). We repeat the sorting strategy based on each category of intangibles talk and report the returns. The returns of the portfolios based on these three categories are presented in Table 2.8. The results are similar to INT^{10K} over the full period and the sub-periods. The highest returns are associated with the first category of intangibles talk related to innovation assets and information technology. However, the returns under all the categories are positive and significant when dot-com bubble years are excluded and become economically important especially in recent years. This indicates that the informativeness of intangibles talk is not limited to any single category and that the effect is present across all the categories.[12]

2.4. Intangibles talk and common risk factors

We have established that INT^{10K} is informative about future returns and outperforms value strategies based on traditional and intangible augmented book-to-market ratio, especially in recent years. We next examine whether the returns associated with intangibles talk are simply premiums for known risk factors. We first test the hypothesis that the return for INT^{10K} is compensation for systematic risk. We regress our portfolio's returns against the systematic factors discussed in the literature, namely the three and five factors in the Fama and French (2015) models plus

[12] We also sort stocks based on intangibles talk within industries, using industry-specific benchmarks. We use the twelve Fama and French industry classifications and form our portfolio using stocks from each industry. The returns are insignificant and mostly positive except in the finance industry, for which the returns are positive and statistically significant. This indicates that the returns generated based on intangibles talk is not primarily a within-industry phenomenon, unlike value (see Asness, Porter and Stevens, 2000).

momentum. We control for all the factors in the regression:

$$R_t = \alpha + \beta_{MKT} \times MKT_t + \beta_{SMB} \times SMB_t + \beta_{HML} \times HML_t + \\ \beta_{RMW} \times RMW_t + \beta_{CMA} \times CMA_t + \beta_{UMD} \times UMD_t + \epsilon_t, \quad (2.2)$$

where R_t is the return of INT^{10K} in month t, and α is the intercept that captures the abnormal returns after controlling for risk factors (alpha hereafter). MKT_t, SMB_t, HML_t, RMW_t, CMA_t, and UMD_t are, respectively, the returns of the market, size, value, profitability, investment, and momentum portfolios taken from Ken French's website. The standard errors are estimated using Newey and West (1987), which allows for serially correlated and heteroskedastic error terms. The alphas are reported in Table 3.6 for the period between July 1995 and June 2020. The first two columns report the alphas over the Fama and French three and five factors plus momentum. In the first column, as the baseline regression, we use HML^{FF} as the value factor. In the third and fourth columns, we use HML^{PT} and HML^{EKP} as the value factor, respectively, to account for intangibles-enhanced book value used in prior studies. The alphas remain positive and statistically significant in the first four columns, with the highest alpha, on an annualized basis, reported at 8.35%, and the lowest at 3.37%. The positive and significant alphas show that the informativeness of intangibles talk with respect to future returns is not fully explained by the common risk factors: market, size, value, profitability, investment, and momentum. This result rules out the possibility that the returns associated with intangibles talk capture the effects of profitability factor (RMW). Profitability is positively associated with intangibles talk (see Table 2.4). In columns 5 to 8 in Table 3.6, we also include INT^{PT}, INT^{EKP}, $R\&D$ and $SG\&A$ portfolios in our regressions. In the last column of the table, we include the acquired intangibles portfolio, $Acq.INT^{portf.}$, which is based on intangible assets from COMPUSTAT and scaled by total assets. This controls for the idea that acquired intangibles that are already presented on balance sheet drive returns reflected in intangibles talk. In all cases, the alpha remains positive and statistically significant, indicating that the informativeness of intangibles talk is not captured by other measures of intangible intensity constructed based on capitalizing past reported R&D and SG&A expenses and acquired intangibles.

The alphas are similar for each of the three categories of intangibles talk as well. Table 2.10 reports the alphas over the Fama and French factors for the three portfolios. The highest alpha belongs to innovation assets and information technology with 3% and 5.3%, respectively, for the Fama and French three and five factors plus momentum. Therefore, while the magnitude of alpha varies across categories, the abnormal returns associated with intangibles talk are not limited to a particular class of intangible assets.

The limited availability of textual data from 10-K filings restricts the exploration of alpha from before 1995. However, the results here are comparable with the 3.48% alpha (over the three Fama and French factors plus momentum) generated by the value-weighted portfolio that picks stocks based on employee satisfaction, reported between 1984 and 2009 by Edmans (2011). Similarly, the alpha associated with R&D and advertising expenses reported in Lev and Sougiannis (1996) is around 4.57% in their sample, which goes back to 1975. So, our results seem to be consistent with prior results, using an alternative measure of intangible intensity.

Table 2.11 better illustrates the relatively cheap (expensive) valuation of stocks with high (low) levels of intangibles talk in the market. We construct five portfolios of stocks sorted on intangibles talk and estimate the alphas in each portfolio from the five-factor model. We repeat the exercise based on other intangible intensity indicators discussed so far. We create five sorted portfolios based on each measure and report their alphas from the five-factor model. These alphas are presented in Table 2.11.

The highest-ranking portfolio based on intangibles talk generates significant positive alpha, while a significant negative alpha is reported for the lowest-ranking portfolio, suggesting that intangible intensive stocks, as measured by intangibles talk, are underpriced. The results for the portfolio of stocks ranked based on R&D to total expenses also confirm the previous findings by reporting a positive and significant alpha for the highest-ranking portfolio. When stocks are ranked based on intangible capital to total assets using the Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020) measures, the alphas are not significant for the highest-ranking portfolios, but the lowest-ranking portfolio in the case of intangible capital$^{\mathrm{PT}}$ generates a significant negative alpha.

The results from this section and Section 2.3 are consistent with the hypothesis that the information on intangibles is not fully considered by investors, with the highest underpricing being associated with our measure of textual disclosures. However, another plausible explanation is that intangibles talk, along with other measures of intangible intensity from the literature that lead to abnormal returns, are proxies for a new systematic risk factor that is different from the commonly known risk factors in the literature.

2.5. Is intangibles talk a systematic risk factor?

We next investigate the extent to which intangible intensity, as measured by intangibles talk and other intangible-augmented value factors, can be considered systematic risk factors that influence the cross-sectional variation in stock returns. We start by constructing aggregate value-weighted intangible factors using three methods: Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020), as well as intangibles talk developed in this paper. We next include these factors in the Fama and MacBeth two-stage regressions as in Fama and MacBeth (1973) to estimate the price that investors pay to bear the risk of these intangible factors.

To ensure that our estimated prices of intangibles' risk are robust, we use alternative sets of test portfolios. By examining various sets of test portfolios, we reduce the likelihood of finding 'false' factors in any one set of test portfolios as pointed out by **?**. We conduct two-stage regressions for each set of portfolios in the Kenneth R. French data library, which has at least ten portfolios.[13] Overall, we have $K = 36$ sets of test portfolios, listed in Table2.12. The test portfolios are created using a diverse set of firm characteristics, such as size, market-to-book, investment, repurchases, and several others. The sets of portfolios, therefore, are unlikely to be based on similar sets of firms.

We use the Fama-Macbeth (FM) two-stage procedure as in Fama and MacBeth (1973) to estimate the price investors would pay to assume intangibles risk. Consider the portfolio set $k \in 1, 2, \cdots, K$. We obtain monthly returns on the factors (MKT, SMB, HML, and UMD) from the Kenneth R. French data library. Each month, risk loadings are estimated using the current and previous thirty-five monthly returns for the first-stage equation:

$$R^p_{t-s} = \alpha_{p,t} + \beta^{MKT}_{p,t} \times MKT_{t-s} + \delta_{p,t} \times INT_{t-s-1} + \beta^{SMB}_{p,t} \times \mathbf{SMB}_{t-s} + \beta^{HML}_{p,t} \times \mathbf{HML}_{t-s} + \beta^{UMD}_{p,t} \times \mathbf{UMD}_{t-s} + \epsilon_{p,t-s}, \quad (2.3)$$

for $p = 1, ..., N_k$, $s = 0, \cdots, 35$. In this equation, $\mathbf{R}^p_t$ is the monthly return of the p^{th} portfolio at t, N_k is the number of portfolios of stocks in portfolio set k, and INT_t is either the INT^{PT}, INT^{EKP}, or INT^{10K} factor. Next, for the second-stage, we estimate for each month the prices of risk at date t for portfolio set k using the equation

$$R^p_t = \kappa^k_t + \pi^k_t \times \beta^{MKT}_{p,t} + \omega^k_t \times \delta_{p,t} + \phi^{SMB,k}_t \times \beta^{SMB}_{p,t} + \phi^{HML,k}_t \times \beta^{HML}_{p,t} + \phi^{UMD,k}_t \times \beta^{UMD}_{p,t} + \epsilon_{t,p}, \quad (2.4)$$

for $p = 1, ..., N_k$. Finally, following Cochrane (2001), we calculate the time-series mean of the estimated prices of risk and the estimated variance (our equations illustrate the case of the price of intangibles risk), for portfolio set k as

$$\hat{\omega}^k = \frac{1}{T}\sum_{t=1}^{T}\hat{\omega^k_t} \quad (2.5)$$

and,

$$\sigma^2(\hat{\omega^k}) = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{\omega^k_t} - \hat{\omega^k})^2, \quad (2.6)$$

respectively. The prices of risk of the other factors are estimated analogously.

In Fig. 2.7, we report the average prices of risk from the second stage regression for each of the factors in the top panels. As can be seen, the prices of risk for each of the three factors are similar; they are almost all positive but small, mostly less than 0.06. The middle panels show the t-statistics of the estimated coefficients. As seen, most of the estimated coefficients are not statistically significant at the five percent level. Overall, neither of the factors seem

[13] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

to be consistently systematically priced in the cross-section of stock returns. The bottom panels show the R^2 of the 2nd stage regression for each of the three factors (along with the four-factor model). The R^2s using each factor are similar and range between 20% and 55% for the different sets of portfolios. We, therefore, conclude that there is no compelling evidence that any of the intangibles-based factors is priced for risk and would improve the specification of the four-factor model.

2.6. Intangibles talk versus. technology stocks

A plausible explanation for abnormal returns based on our measure is that it reflects the strong performance of technology stocks over the past decade. Because technology firms rely more on intangible assets, discussions of intangibles are more likely to occur in 10-K filings from this sector, which would heighten their measured intangibles talk. This is what we find when we plot the distribution of intangibles talk values across the twelve Fama and French industry classifications (see Figs. 2.4 and 2.5). Notably, the business equipment industry ranks the highest among the twelve industries in intangibles talk, especially in the category of innovation assets and information technology. This suggests that the portfolio sorted based on intangibles talk may partially capture the strong performance of technology stocks by taking a long position in stocks in the highest percentiles of intangibles talk in the market. To address this concern, we conduct two exercises to disentangle the effects of intangibles talk from the outperformance of technology stocks.

For our first exercise, we construct an INT^{10K} portfolio independently within each of the twelve Fama and French industries and investigate the magnitude of its abnormal returns against the common risk factors. If INT^{10K}'s strong performance reported in the previous sections comes only from technology stocks, then we should not observe a statistically significant alpha for INT^{10K} within the technology industry. This, however, is not the case based on the results reported in Table 2.13. The alphas of INT^{10K} are significant in only four industries, with its highest values reported in the business and equipment industry at around 6% annually. Therefore even among technology stocks, returns could be significantly enhanced by sorting stocks based on intangibles talk. Other industries with significant alphas are health care, finance, and ”other” industries based on the twelve Fama and French industry classifications. Notably, the ”other” industry classification includes many new businesses that cannot be classified in traditional industries.

Our second exercise uses a blind portfolio strategy in which we take long positions on technology stocks and short stocks from other industries. We call this portfolio $Tech^{portf}$. Like other portfolios in this paper, $Tech^{portf}$ is value-weighted and re-balanced at the end of June every year. We first investigate whether such a strategy would yield a significant and positive abnormal return against the common risk factors. We then test whether $Tech^{portf}$ can explain INT^{10K}'s abnormal returns and vice versa.

Table 2.14 reports the results for this exercise. $Tech^{portf}$ has a significant and positive alpha over the Fama and

French five factors plus momentum. This alpha, however, is partly explained by including INT^{10K} in the regression specifications in columns (3) and (4). This result is consistent with the idea that returns associated with our measure partly reflect the runup in technology stocks. Nevertheless, this does not fully explain returns from our measure. The last two columns show that the positive alpha of INT^{10K} remains significant when we include $Tech^{portf}$ in our regression specifications. In sum, the results from Table 2.14 show that including INT^{10K} in factor regressions renders the alpha of $Tech^{portf}$ insignificant, while INT^{10K}'s alpha remains significant even after including $Tech^{portf}$ in factor regressions. Overall a blind strategy of buying technology stocks and shorting other stocks, even though achieving a positive alpha, does not fully explain the alpha linked to intangibles talk.

2.7. Do intangibles returns represent mispricing?

We next examine whether return informativeness of intangibles talk is related to mispricing associated with limits to arbitrage. To answer this question, we rely on IVOL as a measure for limits to arbitrage that could amplify mispricing. To the extent that the INT^{10K} alphas come from limits to arbitrage for mispriced stocks, they should increase in IVOL.

Diversification of idiosyncratic risks is central to the capital asset pricing model, yet its limiting effect on arbitrage is proven in the literature. Stambaugh, Yu and Yuan (2015) claim that adverse price moves are more likely under higher IVOL and, therefore, high IVOL is a source of arbitrage risk. This is because capital-constrained investors are forced to close their positions prematurely, under high IVOL conditions, before subsequent price corrections occur. We follow Stambaugh, Yu and Yuan (2015) as well as subsequent studies that consider IVOL as a proxy for mispricing (see Cao and Han, 2016; Birru and Young, 2020) to examine whether our results differ under different IVOL conditions.

To the extent that returns from intangibles talk are because of mispricing of intangible-related information in 10-K filings, we expect an amplification of such an effect under high IVOL conditions. In addition, in line with the results in Stambaugh, Yu and Yuan (2012) on the IVOL-return relation, we expect the IVOL-return relation to be negative among stocks that score the lowest in intangibles talk (that is, the putative overpriced stocks) and to be positive for the highest in intangibles talk stocks (that is, putative underpriced stocks).

We test our hypothesis by creating twenty-five double-sorted portfolios, five times five each, based on intangibles talk and IVOL. We then estimate the alphas of each portfolio over the five factors of Fama and French to examine the abnormal returns. Table 2.15 reports the alphas for each of the twenty-five portfolios. The difference between abnormal returns of the highest and lowest-ranking portfolios (high minus low alpha) based on intangibles talk (rows) is not significantly different from zero among stocks with low IVOL. Meanwhile, the alpha difference for the bottom two rows with the highest IVOL is significant and positive. To account for size effects, we also report the alphas for the twenty-five portfolios separately for small and big firms. Table 2.16 shows that the results are similar for small and big firms separately. The results for the big firms imply the same, with the alpha difference being 19.47% among stocks with the highest IVOL.

In Fig. 2.8, we plot the strong positive relation between IVOL and INT^{10K} abnormal returns over Fama and French's five factors. The highest alpha belongs to INT^{10K}, constructed using the stocks in the highest IVOL decile in our sample. More importantly, a positive, and increasing, trend appears in the magnitude of alpha across IVOL deciles. Alpha increases to above 50% average annualized monthly returns in the highest decile of IVOL.

Another observation from Table 2.15 and Table 2.16 comes from the sorting of stocks based on IVOL (columns), demonstrating the negative relation between IVOL and subsequent returns. This aligns with Ang et al. (2006) and the literature on the idiosyncratic volatility puzzle (negative IVOL-return relation). However, this relation disappears among stocks with the highest intangibles talk, for both small and big firms. This result aligns with Stambaugh, Yu and Yuan (2015), which shows that the negative relation of IVOL and returns is stronger for overpriced (low intangibles talk) stocks and is positive among underpriced (high intangibles talk) stocks.

Overall, the results suggest that the abnormal returns associated with intangibles talk is significant and growing in size with higher levels of IVOL and, thus, with limits to arbitrage. The positive relation between alpha and IVOL is also in line with Birru and Young (2020), which finds that investor sentiment is a stronger predictor of subsequent returns (mispricing) in the presence of uncertainty (measured through IVOL).

The abnormal returns of INT^{10K} reported in Section 2.4 indicate that the informativeness of intangibles talk is not captured by the common risk factors. Results from this section support the mispricing hypothesis and therefore ties our findings to the literature that examines mispricing in relation to the limits to arbitrage.

2.8. Conclusion

We construct a textual measure of intangibles talk using 10-K filings and test its informativeness for future stock returns. We devise a long-short portfolio based on our measure and report significant and positive returns between 1995 and 2020, after excluding the years 2000 and 2001. We compare our portfolio's returns with value strategies based on traditional book-to-market ratio and its intangible–augmented versions reported in the literature and document a superior performance, especially in recent years.

We test whether the common risk factors from Fama and French explain intangibles talk's informativeness. Our results indicate a positive and significant alpha over the three- and five–factor models from Fama and French in addition to momentum. We achieve the same results when we replace the value factor in the regressions with its intangible augmented versions from the literature. An implication of the abnormal returns of our portfolio is that investors do not fully price intangibles information disclosed in 10-K filings, perhaps due to the challenging nature of detecting, defining, and valuing intangible assets from that information. Our results align with similar findings on mispricing of

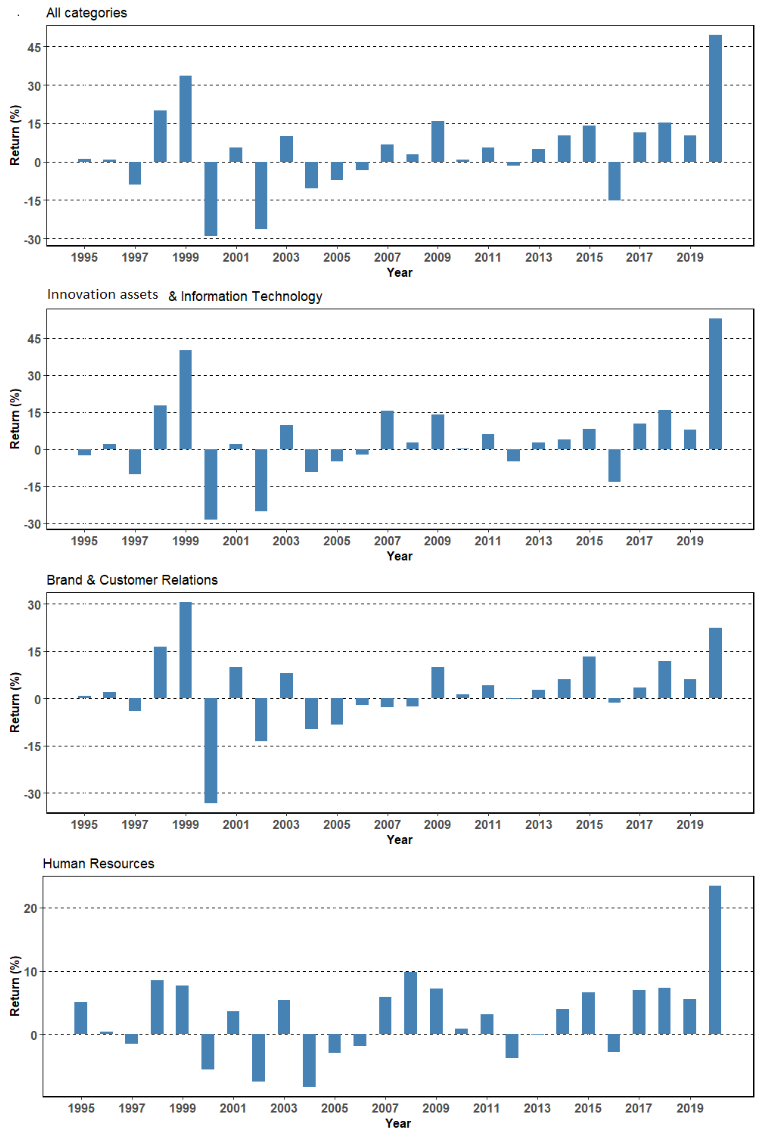
R&D and employee satisfaction in stock valuations.

We test the mispricing hypothesis for intangibles talk's abnormal returns in the presence of limits to arbitrage. We find that the abnormal returns associated with intangibles talk is present only among stocks with high levels of IVOL, which measure arbitrage risks. This supports our mispricing hypothesis and shows that the limits to arbitrage exacerbate the mispricing of intangibles talk.

Our contribution to the literature is twofold. First we show that value-relevant information on successful intangibles investments can be measured from firms' textual disclosures. Second, this information is not fully incorporated in stock prices.

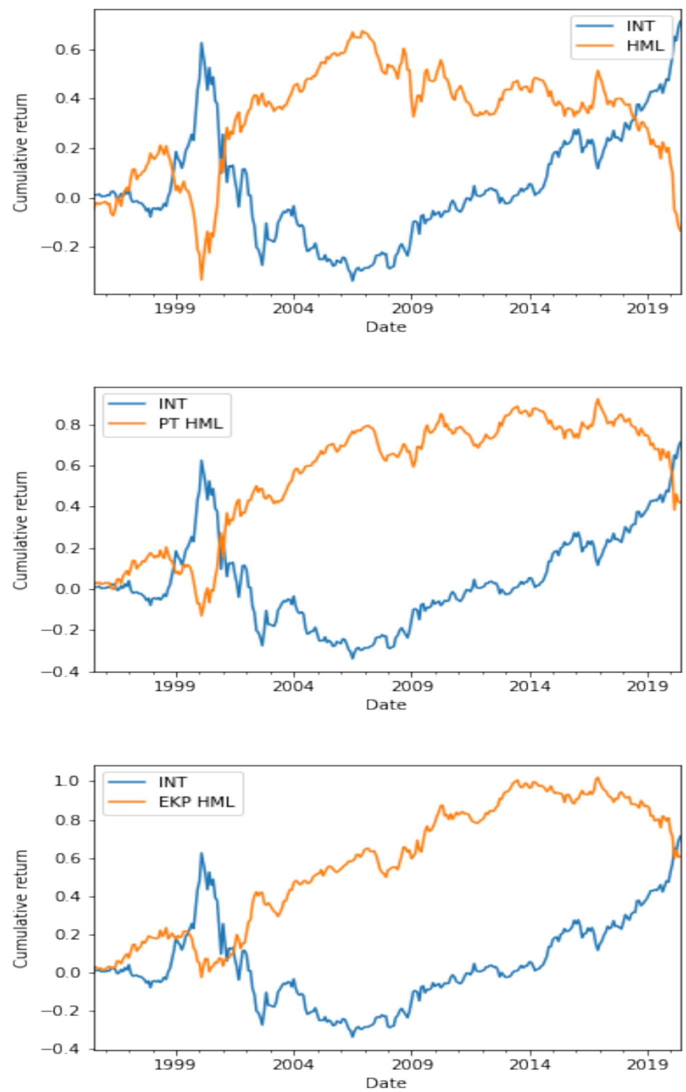
Tables and figures

Figure 2.1: Average Monthly Return Performance



The figure above plots the average monthly returns of INT^{10K} for each year between July 1995 to June 2020. For each category stocks are sorted based on intangibles talk measure in that category. The returns are in percent per year (monthly return multiplied by twelve). Detailed calculations are described in Section 2.3.

Figure 2.2: Cumulative Return Performance



The figure above plots the cumulative returns of one dollar invested in INT^{10K} in comparison to value strategies HML^{FF}, HML^{PT}, and HML^{EKP} for the period between July 1995 to June 2020. Detailed calculations are described in Section 2.3.

Table 2.1: Intangibles words glossary

ABILITIES	CUSTOMER RELATION	INNOVATION	NETWORKS	TALENT
ABILITY	CUSTOMERS	INNOVATIONS	PATENT	TALENTS
ADVERTISING	DATA	INNOVATOR	PATENTED	TEAM
ALGORITHM	DATABASE	INNOVATORS	PATENTS	TEAMS
AUTHORSHIP	DATABASES	INTELLECTUAL	PLATFORM	TEAMWORK
AUTHORSHIPS	DESIGN		PLATFORMS	TECHNOLOGIES
BRAND	DESIGNS	INTELLECTUAL PROPERTY	PRESENCE	TECHNOLOGY
BRANDING	DISCOVERIES	INTERNET	PRODUCTIVITY	TRADE MARK
BRANDS	DISCOVERY	INTERNET ACTIVITIES	PROTECTED DESIGN	TRADE MARKS
CLIENT	EMPLOYEE	INTERNET ACTIVITY	PROTECTED DESIGNS	TRADE NAME
CLIENT RELATIONS	EMPLOYEES	INVENT	REGISTERED DESIGN	TRADE SECRET
CLIENTS	EXPERIENCE	INVENTED	REGISTERED DESIGNS	TRADE SECRETS
COMPETENCE	EXPERT	INVENTING	RELATION	TRADEMARK
COMPETENCES	EXPERTISE	INVENTION	RELATIONS	TRADEMARKS
COMPETENCIES	EXPERTS	INVENTIONS	RELATIONSHIP	TRAINING
COMPETENCY	FORMULA	INVENTS	RELATIONSHIPS	USER
CONNECTIONS	FORMULAE	KNOWHOW	REPUTATION	USERS
CONNECTIVITY	FRANCHISE	KNOWLEDGE	RESEARCH	WEBSITE
CONSUMER	FRANCHISES	LABEL	RESEARCHES	WEBSITES
CONSUMERS	HUMAN	LABELS	SITE VISITS	WORKFORCE
COPYRIGHT	HUMAN CAPITAL	LICENCE	SKILL	
COPYRIGHTS	HUMAN RESOURCES	LICENCES	SKILLS	
CUSTOMER	INNOVATE	LOGO	SOFTWARE	
CUSTOMER BASE	INNOVATE PARTNERS	LOYALTY	SOLUTION	
CUSTOMER BASES	INNOVATED	MARKETING	SOLUTIONS	
CUSTOMER LIST	INNOVATES	NAMES	SYSTEM	
CUSTOMER LISTS	INNOVATING	NETWORK	SYSTEMS	

This table shows 128 words and terms used to calculate intangibles talk measure for each 10-K filings. The words in the table are based on Filipovic and Wager (2019). Calculations of intangibles talk measure is described in Section 3.2.

Table 2.2: Intangibles words by categories of intangibles

Innovation assets & information technology		Brand & customer relations	Human resources
INNOVATE PARTNERS	INNOVATOR	CLIENT RELATIONS	HUMAN CAPITAL
INTELLECTUAL PROPERTIES	INNOVATORS	CUSTOMER BASE	HUMAN RESOURCES
INTELLECTUAL PROPERTY	INTELLECTUAL	CUSTOMER BASES	ABILITIES
INTERNET ACTIVITIES	INTERNET	CUSTOMER LIST	ABILITY
INTERNET ACTIVITY	INVENT	CUSTOMER LISTS	COMPETENCE
PROTECTED DESIGN	INVENTED	CUSTOMER RELATION	COMPETENCES
PROTECTED DESIGNS	INVENTING	ADVERTISING	COMPETENCIES
REGISTERED DESIGN	INVENTION	BRAND	COMPETENCY
REGISTERED DESIGNS	INVENTIONS	BRANDING	EMPLOYEE
SITE VISITS	INVENTS	BRANDS	EMPLOYEES
TRADE MARK	KNOWHOW	CLIENT	EXPERIENCE
TRADE MARKS	KNOWLEDGE	CLIENTS	EXPERT
TRADE NAME	LICENCE	CONNECTIONS	EXPERTISE
TRADE SECRET	LICENCES	CONNECTIVITY	EXPERTS
TRADE SECRETS	NETWORK	CONSUMER	HUMAN
ALGORITHM	NETWORKS	CONSUMERS	PRODUCTIVITY
AUTHORSHIP	PATENT	CUSTOMER	SKILL
AUTHORSHIPS	PATENTED	CUSTOMERS	SKILLS
COPYRIGHT	PATENTS	FRANCHISE	TALENT
COPYRIGHTS	PLATFORM	FRANCHISES	TALENTS
DATA	PLATFORMS	LABEL	TEAM
DATABASE	RESEARCH	LABELS	TEAMS
DATABASES	RESEARCHES	LOGO	TEAMWORK
DESIGN	SOFTWARE	LOYALTY	TRAINING
DESIGNS	SOLUTION	MARKETING	WORKFORCE
DISCOVERIES	SOLUTIONS	NAMES	
DISCOVERY	SYSTEM	PRESENCE	
FORMULA	SYSTEMS	RELATION	
FORMULAE	TECHNOLOGIES	RELATIONS	
INNOVATE	TECHNOLOGY	RELATIONSHIP	
INNOVATED	TRADEMARK	RELATIONSHIPS	
INNOVATES	TRADEMARKS	REPUTATION	
INNOVATING	WEBSITE	USER	
INNOVATION	WEBSITES	USERS	
INNOVATIONS			

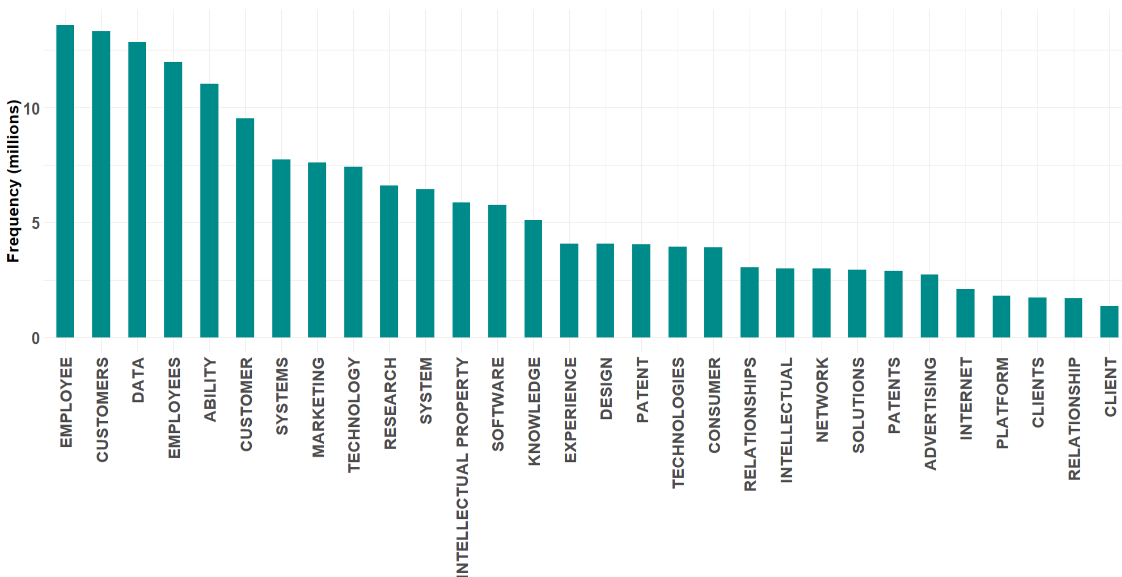
This table shows the list of words to calculate intangibles talk measure for each 10-K filing, by three categories of intangibles. The words in the table are based on Filipovic and Wager (2019). Calculations of intangibles talk measure is described in Section 3.2.

Table 2.3: Variance shares of three intangibles categories in intangibles talk measure

Category	Portion of glossary	Variance share	Top words
Innovation assets & information technology	54%	46.74%	Data, System, Technology, Research, Intellectual Property
Brand & customer relations	27%	28.6%	Customers, Marketing, Consumer, Advertising, Relationship
Human resources	19%	24.66%	Employee, Ability, Experience Expertise, Talent

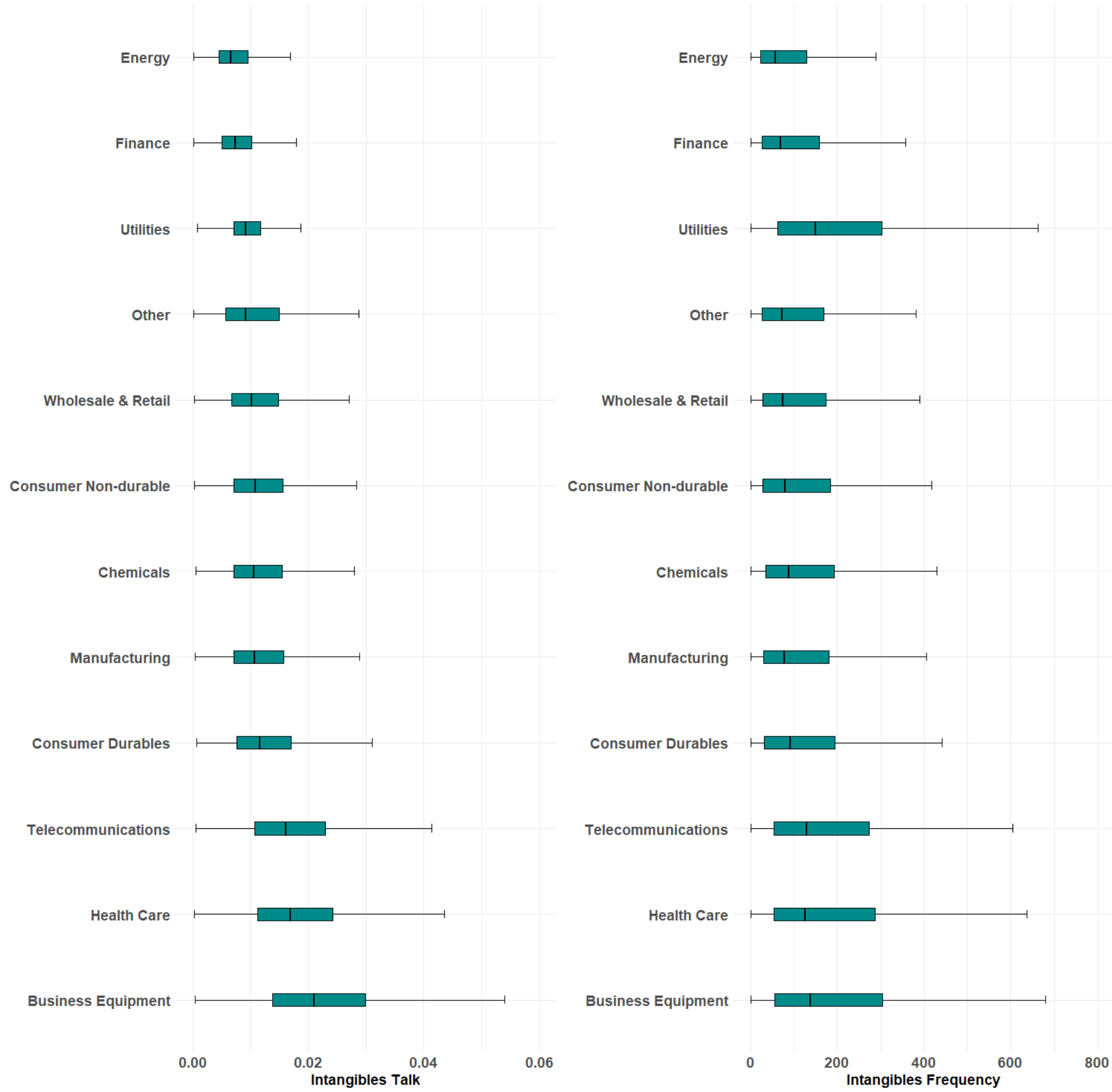
The table reports the share of the total variance coming from each category of intangibles words. The total variance is $[\Sigma_j\, v(j)]$ summed across all the words in our glossary and $v(j)$ is the variance of frequency of word j across all the filings in our sample.

Figure 2.3: Most frequent terms appearing in intangibles talk measure across all firms



This figure shows the frequency of words appearing in intangibles talk measure. The words in the table are based on Filipovic and Wager (2019). Calculations of intangibles talk measure is described in Section 3.2.

Figure 2.4: Intangibles talk measure for each Fama and French twelve industries

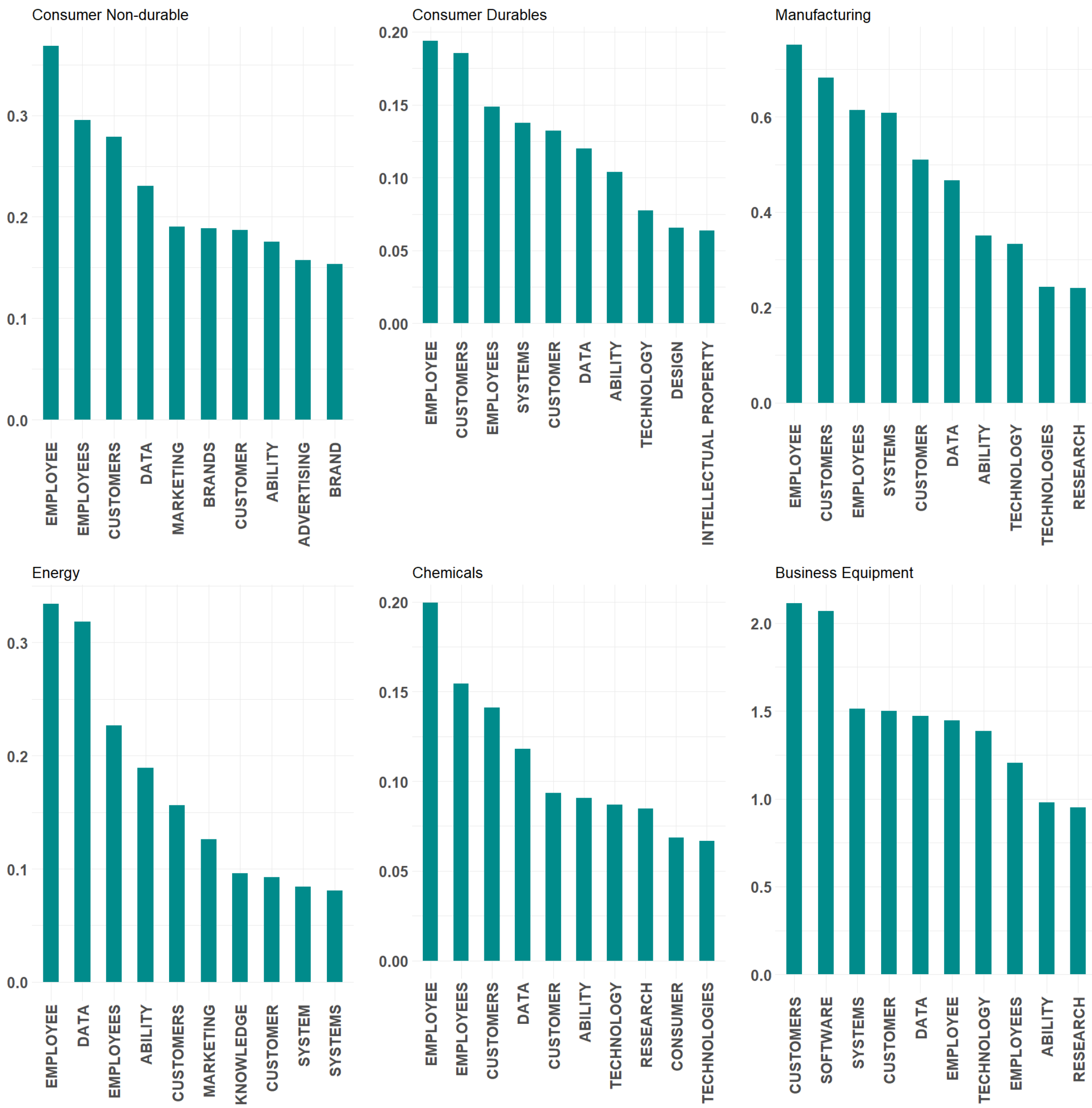


This figure shows the Box and Whisker plot for intangibles talk measure for each Fama and French industry category, with dark central line in the middle representing the median value, and the box representing variation from 25th to 75th percentile. Calculation of intangibles talk measure is described in Section 3.2. Words used to calculate intangibles talk measure come from Filipovic and Wager (2019).

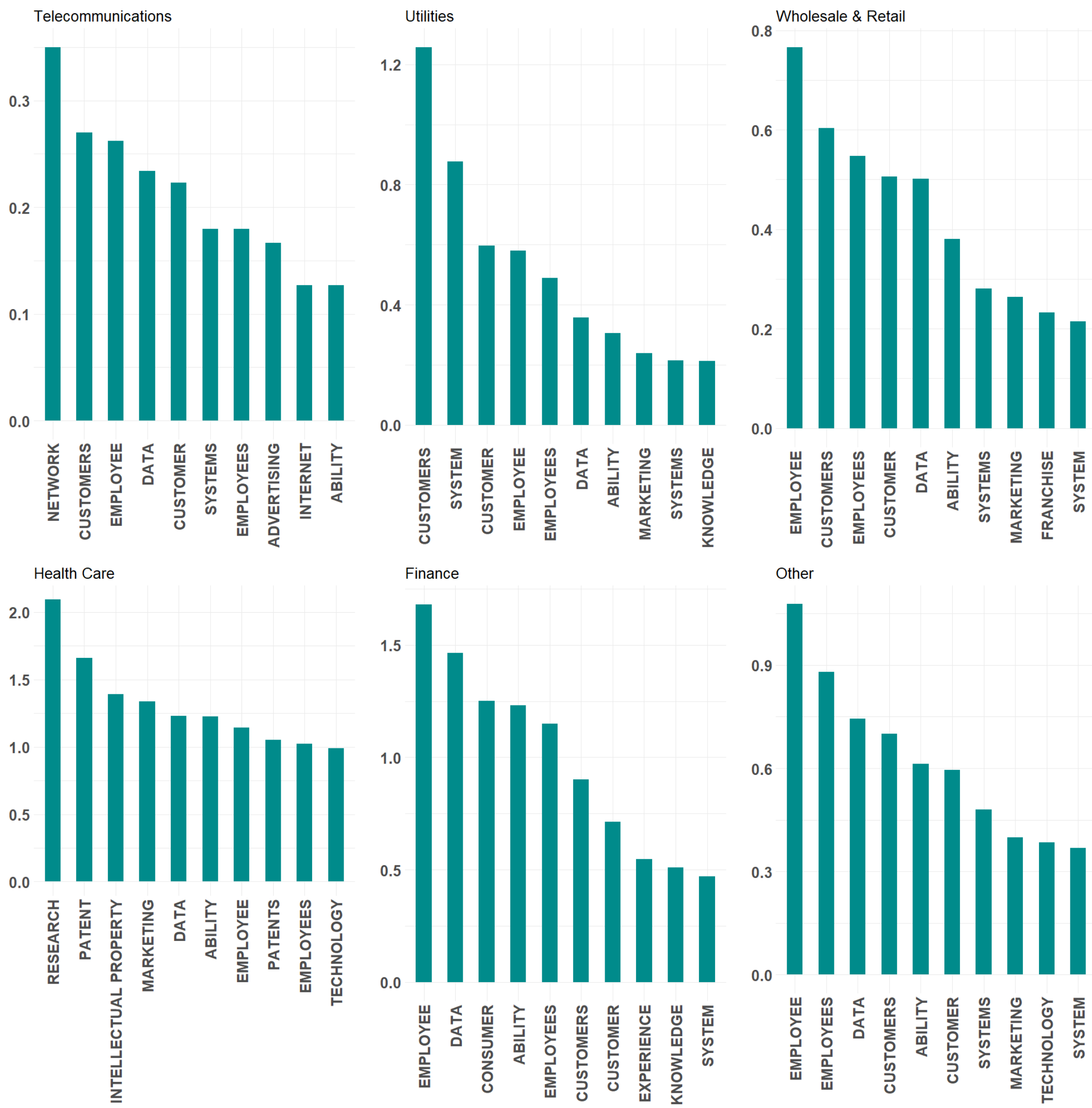
Figure 2.5: Distribution of intangibles talk measure by three intangible categories in Fama and French twelve industries

This figure shows the Box and Whisker plot for intangibles talk measure for each of the three intangible categories, in each Fama and French industry category. Dark central line in the middle representing the median value, and the box representing variation from 25th to 75th percentile. Calculation of intangibles talk measure is described in Section 3.2. Words used to calculate intangibles talk measure come from Filipovic and Wager (2019).

Figure 2.6: Most frequent terms entering intangibles talk measure by Fama and French twelve industries



This figure shows the frequency of words appearing in intangibles talk measure, by Fama and French industry categories. The words in the figure are based on Filipovic and Wager (2019). Calculation of intangibles talk measure is described in Section 3.2.


Telecommunications
0.3
0.2
0.1
0.0
NETWORK
CUSTOMERS
EMPLOYEE
DATA
CUSTOMER
SYSTEMS
EMPLOYEES
ADVERTISING
INTERNET
ABILITY
Utilities
1.2
0.8
0.4
0.0
CUSTOMERS
SYSTEM
CUSTOMER
EMPLOYEE
EMPLOYEES
DATA
ABILITY
MARKETING
SYSTEMS
KNOWLEDGE
Wholesale & Retail
0.8
0.6
0.4
0.2
0.0
EMPLOYEE
CUSTOMERS
EMPLOYEES
CUSTOMER
DATA
ABILITY
SYSTEMS
MARKETING
FRANCHISE
SYSTEM
Health Care
2.0
1.5
1.0
0.5
0.0
RESEARCH
PATENT
INTELLECTUAL PROPERTY
MARKETING
DATA
ABILITY
EMPLOYEE
PATENTS
EMPLOYEES
TECHNOLOGY
Finance
1.5
1.0
0.5
0.0
EMPLOYEE
DATA
CONSUMER
ABILITY
EMPLOYEES
CUSTOMERS
CUSTOMER
EXPERIENCE
KNOWLEDGE
SYSTEM
Other
0.9
0.6
0.3
0.0
EMPLOYEE
EMPLOYEES
DATA
CUSTOMERS
ABILITY
CUSTOMER
SYSTEMS
MARKETING
TECHNOLOGY
SYSTEM

.

Table 2.4: Summary statistics for firms sorted on intangibles talk measure

	Intangibles Talk (All Categories)			Innovation assets & Information Technology		
	Low 30	Mid 40	High 30	Low 30	Mid 40	High 30
R&D to total expenses	0	0.02	0.12	0	0.01	0.13
SG&A to total expenses	0.24	0.23	0.38	0.27	0.21	0.4
Intangible capital$^{\text{PT}}$ to total assets	0.06	0.43	0.73	0.05	0.42	0.76
Intangible capital$^{\text{PT}}$ to sales	0.34	0.44	0.88	0.34	0.39	0.99
Sales to total assets	0.34	0.96	0.8	0.32	1.04	0.75
Sales to stockholder's equity	1.01	1.98	1.37	1.02	2.2	1.25
Price to sales	0.05	0.03	0.07	0.05	0.03	0.08
Debt to EBITDA	2.22	1.3	0.15	2.35	1.37	0.19
Debt to total assets	0.26	0.2	0.07	0.26	0.2	0.07
Profitability to total assets	0.09	0.28	0.35	0.08	0.29	0.33
Investment to physical capital	0.09	0.09	0.12	0.08	0.09	0.12
Market Cap (millions)	401.16	566.35	382.35	377.97	655.49	364.4
Book to market^{FF}	0.73	0.51	0.38	0.73	0.51	0.37
Book to market^{EKP}	1.2	1.21	1.12	1.23	1.22	1.07
Book to market^{PT}	0.85	0.73	0.63	0.87	0.73	0.6
	Brand & Customer Relations			Human Resources		
	Low 30	Mid 40	High 30	Low 30	Mid 40	High 30
R&D to total expenses	0.08	0.04	0.06	0.02	0.05	0.09
SG&A to total expenses	0.22	0.3	0.32	0.26	0.28	0.32
Intangible capital$^{\text{PT}}$ to total assets	0.14	0.23	0.6	0.12	0.37	0.58
Intangible capital$^{\text{PT}}$ to sales	0.41	0.53	0.62	0.43	0.53	0.64
Sales to total assets	0.45	0.68	0.95	0.53	0.73	0.85
Sales to stockholder's equity	1.1	1.3	1.75	1.21	1.43	1.57
Price to sales	0.05	0.06	0.05	0.05	0.05	0.05
Debt to EBITDA	1.46	1.05	0.58	1.64	1.11	0.22
Debt to total assets	0.21	0.18	0.11	0.23	0.18	0.09
Profitability to total assets	0.12	0.21	0.36	0.16	0.24	0.32
Investment to physical capital	0.09	0.09	0.11	0.09	0.1	0.11
Market Cap (millions)	447.57	387.42	480.18	425.67	401.57	473.28
Book to market^{FF}	0.58	0.61	0.46	0.66	0.56	0.44
Book to market^{EKP}	1.04	1.24	1.3	1.21	1.2	1.14
Book to market^{PT}	0.75	0.8	0.71	0.83	0.76	0.67

This table summarizes the characteristics of firms sorted by intangibles talk measure above 70th percentile, between 30th to 70th percentile, and the bottom 30th percentile. The values are the time-series average of the median firm characteristics within each bucket. The sample period is from January 1994 to December 2019. Calculation of intangibles talk measure is described in Section3.2. Words used to calculate intangibles talk measure come from Filipovic and Wager (2019).

Table 2.5: Association between intangibles talk measure and other proxies for intangible intensity

	Dependent variable:					
	Intangibles Talk$_t$					
	(1)	(2)	(3)	(4)	(5)	(6)
Book-to-Market$_t$	−0.176					0.006
	(−56.96)					(1.03)
$\frac{(SG\&A_t - R\&D_t)}{Total\ Expenses_t}$		0.299				0.255
		(88.05)				(40.63)
$\frac{R\&D_t}{Total\ Expenses_t}$			0.256			0.520
			(55.33)			(59.45)
$\frac{Intangible\ Capital_t^{PT}}{Total\ Assets_{t-1}}$				0.372		−0.512
				(112.43)		(−28.51)
$\frac{Intangible\ Capital_t^{EKP}}{Total\ Assets_{t-1}}$					0.370	0.525
					(108.74)	(29.98)
Observations	99,112	82,713	43,986	84,283	81,927	31,923
Adjusted R^2	0.032	0.086	0.065	0.130	0.126	0.218

In this table, we report the pooling regression with firm-level intangibles talk as the dependent variable:

$$\text{Intangibles Talk}_{i,t} = \boldsymbol{\alpha_i} + \boldsymbol{\beta} X_{i,t} + \epsilon_{i,t} \tag{2.7}$$

where $X_{i,t}$ are firm-level book-to-market, (SG&A - R&D)/Total Expense, R&D/Total Expense, intangible capital$^{\text{PT}}$. The panel covers the period between January 1994 to December 2021. To separately capture the effect of only SG&A we use (SG&A - R&D) since R&D is already included in SG&A. All the variables are annual, with their negative values dropped, winsorized at 1% from above, and normalized by dividing by standard deviation. The error terms are clustered at the firm level and time fixed effects are accounted for. Calculation of intangibles talk measure is described in Section3.2. Words used to calculate intangibles talk measure come from Filipovic and Wager (2019).

Table 2.6: Risk and returns from intangible talk measure versus intangibles-enhanced value strategies

		Full Sample (1995-2020)	Full Sample (Exc. Dot-com Bubble)	Dot-com Bubble (2000-2001)	1995-1999	2000-2007	2008-2020
		(1)	(2)	(3)	(4)	(5)	(6)
INT^{10K}	Ret	3.79	**5.05**	-10.69	10.78	-6.13	**7.63**
		(1.36)	(2.3)	(-0.85)	(1.52)	(-1.22)	(3.15)
	σ	12.13	9.31	29.24	9.86	17	8.26
	Sharpe	0.31	0.54	-0.37	1.09	-0.36	0.92
HML^{FF}	Ret	0.22	-2.18	**27.82**	-3.41	9.64	-4.51
		(0.07)	(-0.91)	(2.49)	(-0.44)	(1.66)	(-1.38)
	σ	11.13	9.44	21.75	10.12	12.33	10.4
	Sharpe	0.02	-0.23	1.28	-0.34	0.78	-0.43
HML^{PT}	Ret	2.01	0.06	**24.51**	-1.83	9.21	-1.21
		(0.81)	(0.03)	(2.65)	(-0.37)	(1.8)	(-0.42)
	σ	10.24	8.53	21.12	8.31	11.94	9.54
	Sharpe	0.2	0.01	1.16	-0.22	0.77	-0.13
HML^{EKP}	Ret	2.55	2.32	5.26	0.91	5.72	1.12
		(1.27)	(1.26)	(1.46)	(0.24)	(1.55)	(0.4)
	σ	8.56	8.22	12.01	7.39	9	8.67
	Sharpe	0.3	0.28	0.44	0.12	0.64	0.13

In this table, we summarize the risk and return associated with intangibles talk measure and other measures of intangible value documented in the literature. INT^{10K} is the portfolio sorted based on the intangible talk. Calculation of intangibles talk measure is described in Section3.2. Words used to calculate intangibles talk measure come from Filipovic and Wager (2019). HML^{FF} is sorted based on traditional book-to-market value. HML^{PT} and HML^{EKP} are sorted based on intangible augmented book-to-market calculated using Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020) methods. Calculation of portfolio returns is described in Section2.3. The values in parentheses are Newey-West T-statistics that test the difference in the means from zero. The full sample is from July 1995 to June 2020. We exclude the years 2000 and 2001 in the sample, that is, the bursting of dot-com bubble. The returns are in percent per year (monthly return multiplied by twelve).

Table 2.7: Outperformance of intangible talk measure over other intangibles-based investment strategies

		Full Sample (1995-2020)	Full Sample (Exc. Dot-com Bubble)	Dot-com Bubble (2000-2001)	1995-1999	2000-2007	2008-2020
		(1)	(2)	(3)	(4)	(5)	(6)
INT^{10K}	Ret	3.79	**5.05**	-10.69	10.78	-6.13	**7.63**
		(1.36)	(2.3)	(-0.85)	(1.52)	(-1.22)	(3.15)
	σ	12.13	9.31	29.24	9.86	17	8.26
	Sharpe	0.31	0.54	-0.37	1.09	-0.36	0.92
$\text{SG\&A}^{portf.}$	Ret	0.53	1.44	-9.91	9.39	**-7.65**	2.58
		(0.25)	(0.75)	(-1.16)	(1.89)	(-2.52)	(1.17)
	σ	10.29	9.1	19.41	9.4	12.01	9.11
	Sharpe	0.05	0.16	-0.51	1	-0.64	0.28
$\text{R\&D}^{portf.}$	Ret	3.29	**4.92**	-15.39	13.58	-6.18	**5.65**
		(0.94)	(2.16)	(-0.92)	(1.61)	(-0.87)	(2.44)
	σ	14.43	11.32	33.79	12.72	20.21	9.56
	Sharpe	0.23	0.43	-0.46	1.07	-0.31	0.59
INT^{PT}	Ret	3.61	**4.52**	-6.44	6.36	-3.87	**7.62**
		(1.45)	(2.07)	(-0.59)	(0.77)	(-1.1)	(2.74)
	σ	11.09	9.35	22.94	11.83	12.81	9.45
	Sharpe	0.33	0.48	-0.28	0.54	-0.3	0.81
INT^{EKP}	Ret	3.44	**4.26**	-5.6	5.52	-3.07	**7.02**
		(1.62)	(2.12)	(-0.59)	(0.75)	(-1.04)	(2.67)
	σ	9.83	8.58	18.87	10.69	10.72	8.83
	Sharpe	0.35	0.5	-0.3	0.52	-0.29	0.79

In this table, we summarize the risk and return associated with textual intangible value and other measures of value and intangible value in the literature. SG&A and R&D portfolios are sorted based on (SG&A - R&D)/Total Expense and R&D/Total Expense. INT^{PT} and INT^{EKP} are sorted based only on the capitalized intangibles based, respectively in Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020) methods, both scaled by total assets. Calculation of portfolio returns is described in Section2.3. The values in parentheses are Newey-West T-statistics that test the difference in the means to be zero. The returns are monthly in percent per year (monthly return multiplied by twelve).

Table 2.8: Risk and returns from intangible talk measure for different subperiods and different intangible categories

Category		Full Sample (1995-2020) (1)	Full Sample (Exc. Dot-com Bubble) (2)	Dot-com Bubble (2000-2001) (3)	1995-1999 (4)	2000-2007 (5)	2008-2020 (6)
Innovation assets & Information Technology	Ret	3.52	**4.9**	-12.43	10.35	-4.77	**6.36**
		(1.16)	(2.19)	(-0.88)	(1.19)	(-0.81)	(2.5)
	σ	12.64	9.62	30.78	10.93	17.73	8.49
	Sharpe	0.28	0.51	-0.4	0.95	-0.27	0.75
Brand & Customer Relations	Ret	2.47	**3.69**	-11.57	10.84	-6.11	**4.95**
		(1.15)	(2.14)	(-1.15)	(1.58)	(-1.63)	(3.15)
	σ	9.89	7.55	23.91	7.88	13.77	6.85
	Sharpe	0.25	0.49	-0.48	1.38	-0.44	0.72
Human Resources	Ret	**2.47**	**2.71**	-0.28	**4.46**	-1.38	**4.21**
		(2.06)	(2.35)	(-0.06)	(2.4)	(-0.69)	(2.74)
	σ	6.29	5.71	11.15	4.92	8.08	5.3
	Sharpe	0.39	0.47	-0.03	0.91	-0.17	0.8

In this table, we summarize the risk and return associated with intangible talk measure, by three intangible categories. Calculation of intangibles talk measure is described in . Words used to calculate intangibles talk measure come from Filipovic and Wager (2019). Calculation of portfolio returns is described in Section2.3. The values in parentheses are Newey-West T-statistics that test the difference in the means to be zero. The full sample is from July 1995 to June 2020. We exclude the years 2000 and 2001 in the sample, excluding the dot-com bubble. The returns are monthly in percent per year (monthly return multiplied by twelve).

Table 2.9: Risk-adjusted future returns (alphas) from intangible talk measure.

	Dependent variable: INT^{10K}								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\alpha(\%)$	**3.369****	**5.814*****	**6.374*****	**8.349*****	**2.540*****	**3.014*****	**2.924****	**4.056*****	**6.178*****
	(2.352)	(4.030)	(4.319)	(4.103)	(2.831)	(3.245)	(2.428)	(3.740)	(4.026)
Mkt-RF	0.109**	0.033	0.056	−0.030	0.061***	0.066***	0.026	0.105***	0.059*
	(2.387)	(0.912)	(1.391)	(−0.520)	(3.738)	(3.998)	(0.829)	(3.358)	(1.691)
SMB	0.153*	0.010	0.091	−0.027	−0.083	−0.086	0.075	0.095	0.019
	(1.937)	(0.123)	(0.981)	(−0.353)	(−1.491)	(−1.511)	(1.167)	(1.285)	(0.264)
HML^{FF}	−0.824***	−0.650***			−0.045	−0.123***	−0.433***	−0.551***	−0.577***
	(−9.996)	(−9.004)			(−1.027)	(−2.785)	(−5.705)	(−10.788)	(−7.509)
HML^{PT}			−0.716***						
			(−8.656)						
HML^{EKP}				−0.219**					
				(−2.166)					
RMW		−0.427***	−0.386***	−0.656***	−0.335***	−0.426***	−0.080	−0.251***	−0.509***
		(−4.653)	(−3.643)	(−6.538)	(−5.362)	(−6.848)	(−0.982)	(−2.609)	(−5.691)
CMA		−0.050	−0.007	−0.485***	−0.101	−0.063	−0.008	0.023	−0.123
		(−0.474)	(−0.069)	(−2.889)	(−1.460)	(−0.908)	(−0.087)	(0.246)	(−1.177)
UMD	−0.098	−0.077	−0.041	−0.012	−0.051*	−0.053**	−0.066*	−0.081**	−0.054
	(−1.511)	(−1.469)	(−1.067)	(−0.189)	(−1.845)	(−2.043)	(−1.665)	(−2.114)	(−1.117)
INT^{PT}					0.780***				
					(19.181)				
INT^{EKP}						0.775***			
						(16.812)			
R&D$^{portf.}$							0.399***		
							(6.639)		
SG&A$^{portf.}$								0.380***	
								(6.362)	
Acq.INT$^{portf.}$									0.358***
									(2.649)
Observations	300	300	300	300	288	288	300	300	300
Adjusted R^2	0.622	0.681	0.669	0.533	0.853	0.843	0.743	0.743	0.707

In this table, we report portfolio alphas and betas by regressing the returns of INT^{10K} against factor models. Calculation of intangibles talk measure is described in Section3.2. Words used to calculate intangibles talk measure come from Filipovic and Wager (2019). Calculation of portfolio returns (INT^{10K}) is described in Section2.3. Columns (1) and (2) use the Fama and French (2015) three and five factors and momentum factor from Carhart (1997). HML portfolios is based on traditional book-to-market. Columns (3) and (4) replace HML with its intangible augmented versions from Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020). Columns (5) and (6) include INT^{PT} and INT^{EKP} sorted based only on the capitalized intangibles based, respectively, in Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020) methods, both scaled by total assets. Columns (7) and (8) include portfolios sorted based on R&D/Total Expense and (SG&A - R&D)/Total Expense, and column (9) includes the portfolio sorted based on acquired intangibles which is intangible assets from COMPUSTAT scaled by total assets. We include Newey-West T-statistics. The sample is monthly from July 1995 to June 2020. All coefficients are reported in percentage per year (monthly returns multiplied by twelve).

Table 2.10: Risk-adjusted future returns (alphas) from intangible talk measure, by intangible categories

	Dependent variable: Categorical INT^{10K}					
Category:	Innovation assets &information technology	Band & customer relation	Human resources	Innovation assets & information technology	Band & customer relation	Human resources
$\alpha(\%)$	3.000**	2.279*	2.286**	5.298***	4.674***	2.619**
	(2.091)	(1.757)	(2.458)	(4.045)	(3.019)	(2.581)
Mkt-RF	0.109**	0.087**	0.039	0.039	0.007	0.034
	(2.498)	(1.973)	(1.576)	(1.177)	(0.168)	(1.387)
SMB	0.182***	0.042	0.064*	0.039	−0.068	0.012
	(2.921)	(0.814)	(1.661)	(0.625)	(−1.269)	(0.232)
HML^{FF}	−0.913***	−0.531***	−0.334***	−0.755***	−0.337***	−0.335***
	(−13.108)	(−4.835)	(−10.228)	(−13.933)	(−4.075)	(−7.020)
RMW				−0.419***	−0.346***	−0.135*
				(−4.992)	(−4.541)	(−1.766)
CMA				−0.016	−0.173	0.126
				(−0.166)	(−1.369)	(1.207)
UMD	−0.084	−0.090	−0.030	−0.064	−0.070	−0.027
	(−1.284)	(−1.234)	(−0.751)	(−1.223)	(−1.135)	(−0.741)
Adjusted R^2	0.704	0.389	0.379	0.759	0.453	0.417
obs	300	300	300	300	300	300

In this table, we report portfolio alphas and betas by regressing the returns from each category of INT^{10K} against factor models. Calculation of intangibles talk measure is described in Section3.2. Words used to calculate intangibles talk measure come from Filipovic and Wager (2019). Calculation of portfolio returns (INT^{10K}) is described in Section2.3. Columns (1) through (3) use the Fama and French (2015) three factors and momentum factor from Carhart (1997). Columns (4) through (6) use the Fama and French (2015) five factors and momentum factor from Carhart (1997). We include Newey-West T-statistics. The sample is monthly from July 1995 to June 2020. All coefficients are reported in percentage per year (monthly returns multiplied by twelve).

Table 2.11: Risk-adjusted returns of portfolios sorted on intangible intensity indicators

Intangibles Talk →	Low	2	3	4	High	H-L
$\alpha(\%)$	**-3.09**	-0.2	0.09	0.79	**3.94**	**7.03**
	(-3.12)	(-0.19)	(0.11)	(1.08)	(3.65)	(4.18)
R&D to Total Expenses →	**Low**	**2**	**3**	**4**	**High**	**H-L**
$\alpha(\%)$	0.76	-2.51	0.37	-0.15	**4.81**	**4.04**
	(0.5)	(-1.48)	(0.3)	(-0.13)	(3.58)	(2.4)
SG&A to Total Expenses →	**Low**	**2**	**3**	**4**	**High**	**H-L**
$\alpha(\%)$	-1.71	-1.05	0.37	2.13	1.031	**2.54**
	(-1.02)	(-0.84)	(0.3)	(1.74)	(1.02)	(2.38)
Intangible Capital$^{\text{PT}}$ to Total Assets →	**Low**	**2**	**3**	**4**	**High**	**H-L**
$\alpha(\%)$	6.01	6.62	**9.12**	**7.97**	**8.85**	2.84
	(1.35)	(1.85)	(2.51)	(2.48)	(2.53)	(0.94)
Intangible Capital$^{\text{EKP}}$ to Total Assets →	**Low**	**2**	**3**	**4**	**High**	**H-L**
$\alpha(\%)$	6	**6.89**	**7.71**	**7.82**	**9.63**	3.63
	(1.36)	(1.97)	(2.06)	(2.41)	(2.91)	(1.35)

This table reports risk-adjusted returns (alphas) from five-factor Fama and French model and momentum factor from Carhart (1997). In each regression, the dependent variable is the excess returns (over risk-free rate) of value-weighted portfolios sorted on one intangible intensity at a time: intangibles talk, R&D and SG&A to total expenses, Intangible Capital$^{\text{PT}}$ to total assets, and Intangible Capital$^{\text{EKP}}$ to total assets. The regression model is:

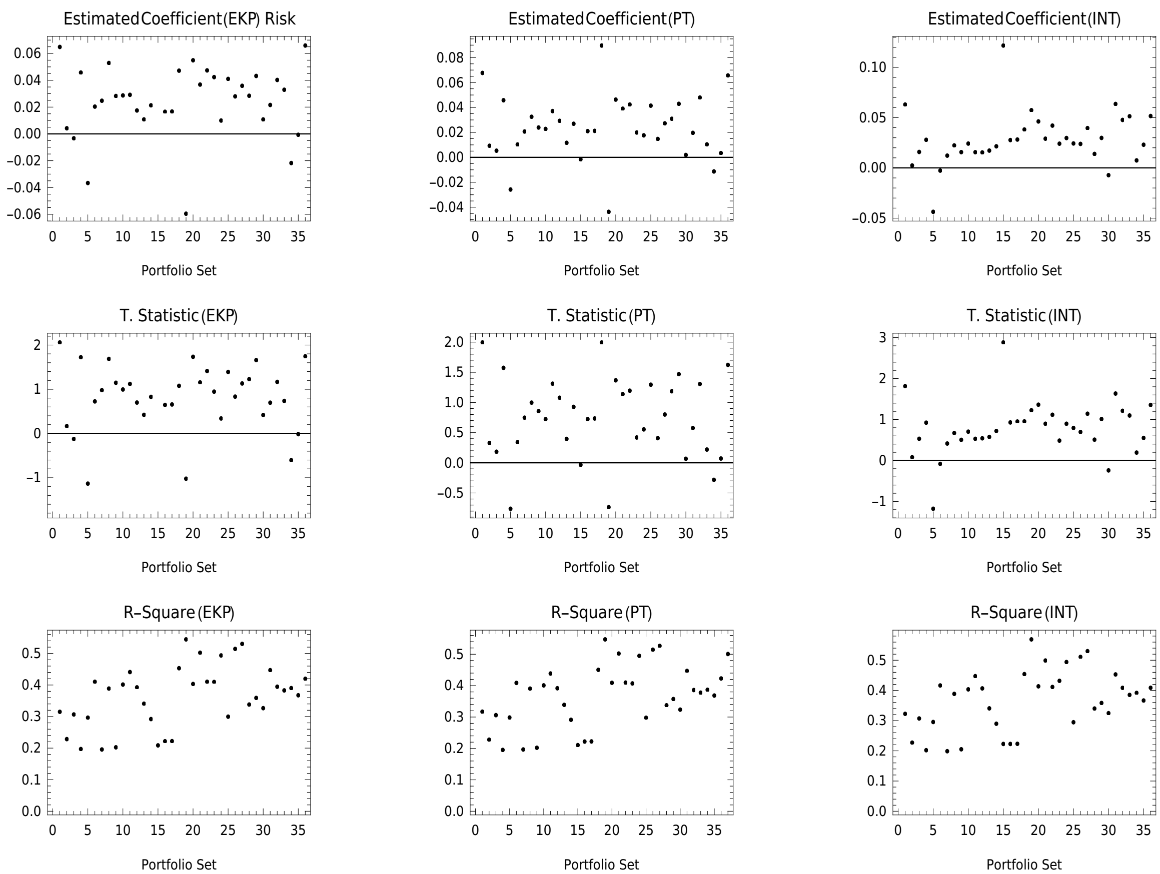
$$\mathrm{R}_{i,t} = \boldsymbol{\alpha_i} + \boldsymbol{\beta_{MKT}} MKT_t + \boldsymbol{\beta_{SMB}} SMB_t + \boldsymbol{\beta_{HML}} HML_t + \boldsymbol{\beta_{RMW}} RMW_t + \boldsymbol{\beta_{CMA}} \mathbf{CMA_t} + \boldsymbol{\beta_{UMD}} \mathbf{UMD_t} + \epsilon_{\mathbf{i,t}}$$

Where $R_{i,t}$ is the excess return (over risk-free rate) of a value-weighted portfolio in month t that is long in stocks belonging to the i^{th} (i from 1 to 5) quantile based on intangible intensity measures. Portfolio returns calculation is described in Section 2.3. Only the alphas from the regressions are reported for expositional purposes. We include Newey-West T-statistics in parentheses. The last column displays alphas from going long on highest quantile (5^{th}) and going short on the lowest quantile (1^{st}) of intangible-intensity measures. The sample is monthly from July 1995 to June 2020.

Table 2.12: Kenneth French's Sets of Portfolios (July 1995 - June 2020)

Number	Name
1	25 Portfolios Based on Size and Accruals
2	25 Portfolios Based on Book-to-Market and Investment
3	25 Portfolios based on Book-to-Market and Operating Profitability
4	25 Portfolios Based on Operating Profitability and Investment
5	100 Portfolios Based on Size and Book-to-Market
6	25 Portfolios Based on Size and Book-to-Market
7	100 Portfolios Based on Size and Investment
8	25 Portfolios Based on Size and Investment
9	100 Portfolios Based on Size and Operating Profitability
10	25 Portfolios Based on Size and Operating Profitability
11	10 Industry Portfolios
12	12 Industry Portfolios
	17 Industry Portfolios
14	30 Industry Portfolios
15	38 Industry Portfolios
16	48 Industry Portfolios
17	49 Industry Portfolios
18	10 Portfolios Based on Long-Term Reversal
19	10 Portfolios Based on Momentum
20	25 Portfolios based on Size and Long-Term Reversal
21	25 Portfolios Based on Size and Momentum
22	25 Portfolios Based on Size and Short-Term Reversal
23	10 Portfolios Based on Short Term Reversal
24	25 Portfolios Based on Size and Market Beta
25	25 Portfolios Based on Size and Net Share Issuance
26	25 Portfolios Based on Size and Residual Variance
27	25 Portfolios Based on Size and Variance
28	32 Portfolios Based on Book-to-Market, Investment, and Size
29	32 Portfolios Based on Book-to-Market, Operating Profitability and Size
30	32 Portfolios Based on Operating Profitability, Investment, and Size
31	10 Portfolios Based on Book-to-Market
32	10 Portfolios Based on Cash Flow Divided by Price
33	10 Portfolios Based on Dividend Yield
34	10 Portfolios Based on Earnings Divided by Price
35	10 Portfolios Based on Investment
36	10 Portfolios Based on Operating Profitability

Figure 2.7: Price of Intangibles Risk in Different Sets of Portfolios Created by Kenneth French



The price of intangibles risk is estimated from Fama-Macbeth two-stage regressions for each set of portfolios listed in Table2.12 and shown on the x-axis of each panel. Each month, risk loadings are estimated from the first-stage regression for portfolio p

$$\mathbf{R^p_{t-s}} = \alpha_{\mathbf{p,t}} + \beta^{\mathbf{MKT}}_{\mathbf{p,t}} \times \mathbf{MKT_{t-s}} + \delta_{\mathbf{p,t}} \times \mathbf{INT_{t-s-1}} + \beta^{\mathbf{SMB}}_{\mathbf{p,t}} \times \mathit{SMB}_{\mathbf{t-s}}$$
$$+\beta^{HML}_{p,t} \times \mathit{HML}_{t-s} + \beta^{UMD}_{p,t} \times \mathit{UMD}_{t-s} + \epsilon_{p,t-s}$$

for $p = 1, ..., N_k$, $s = 0, \cdots, 35$. In this equation, R^p_t is the monthly return of the p^{th} portfolio at t, N_k is the number of portfolios of stocks in portfolio set k, and INT_t is either the INT^{PT}, INT^{EKP}, or INT^{10K} factor. For the second-stage, each month we estimate the prices of risk at date t for portfolio set k using the equation

$$R^p_t = \kappa^k_t + \pi^k_t \times \beta^{MKT}_{p,t} + \omega^k_t \times \delta_{p,t} + \phi^{SMB,k}_t \times \beta^{SMB}_{p,t} + \phi^{HML,k}_t \times \beta^{HML}_{p,t} + \phi^{UMD,k}_t \times \beta^{UMD}_{p,t} + \epsilon_{t,p},$$

for $p = 1, ..., N_k$. We calculate the time-series mean of the estimated prices of risk and the estimated variance (our equations illustrate the case of the price of intangibles risk), for portfolio set k as

$$\hat{\omega}^k = \frac{1}{T}\sum_{t=1}^{T}\hat{\omega^k_t}; \qquad \sigma^2(\hat{\omega^k})\frac{1}{T^2}\sum_{t=1}^{T}(\hat{\omega^k_t} - \hat{\omega^k})^2,$$

respectively. The top panels report the estimated price of risk for each intangible, the middle panels report the t-statistics of the coefficients, and the bottom panel reports the R^2 of the 2nd stage regression, for the INT^{PT}, INT^{EKP}, and INT^{10K} factors, respectively.

Table 2.13: Risk-adjusted future returns (alphas) from intangible talk measure, by twelve Fama and French industry classification

	Dependent variable: *INT^{10K} in 12 Fama–French industries*					
	NoDur	Dur	Manuf	Enrgy	Chems	BusEq
$\alpha(\%)$	−2.736	0.688	2.721	−2.183	−2.127	**6.079****
	(−1.221)	(0.226)	(1.473)	(−0.750)	(−0.716)	(2.530)
Mkt-RF	0.238***	−0.165**	−0.092*	0.046	−0.132	−0.100*
	(4.983)	(−2.464)	(−1.887)	(0.648)	(−1.384)	(−1.745)
SMB	0.101	−0.103	0.016	0.279**	−0.180**	−0.191**
	(1.143)	(−1.127)	(0.330)	(2.414)	(−2.041)	(−1.997)
HML	0.059	−0.176**	−0.101	−0.034	−0.174	−0.033
	(0.723)	(−2.113)	(−1.182)	(−0.248)	(−1.334)	(−0.396)
RMW	0.135	−0.268**	−0.264***	−0.118	−0.184	−0.327***
	(1.479)	(−2.460)	(−3.746)	(−1.274)	(−1.572)	(−3.065)
CMA	−0.097	−0.227	−0.115	−0.056	0.263	−0.324*
	(−0.731)	(−1.420)	(−0.843)	(−0.309)	(1.311)	(−1.841)
UMD	−0.001**	−0.00001	0.0001	−0.001	0.0004	0.0005
	(−2.145)	(−0.018)	(0.322)	(−1.449)	(0.727)	(0.975)
Observations	300	300	300	300	300	300
Adjusted R^2	0.134	0.059	0.134	0.082	0.049	0.151

	Telcm	Utils	Shops	Hlth	Finance	Other
$\alpha(\%)$	−2.679	0.070	0.809	**5.451***	**3.952*****	**5.346****
	(−0.680)	(0.054)	(0.455)	(1.872)	(3.548)	(2.489)
Mkt-RF	−0.133	−0.120	0.156***	0.003	0.010	−0.097*
	(−1.148)	(−1.439)	(2.766)	(0.057)	(0.275)	(−1.861)
SMB	−0.403**	−0.122	0.150**	0.055	0.064	−0.284***
	(−2.519)	(−1.305)	(2.311)	(0.518)	(1.193)	(−3.596)
HML	−0.427***	−0.047	−0.204***	−0.270***	−0.447***	−0.195*
	(−3.874)	(−0.520)	(−3.872)	(−2.687)	(−4.647)	(−1.839)
RMW	0.135	−0.268**	−0.264***	−0.118	−0.184	−0.327***
	(1.479)	(−2.460)	(−3.746)	(−1.274)	(−1.572)	(−3.065)
CMA	−0.097	−0.227	−0.115	−0.056	0.263	−0.324*
	(−0.731)	(−1.420)	(−0.843)	(−0.309)	(1.311)	(−1.841)
UMD	−0.001	0.0003	−0.001*	0.001*	−0.0002	0.002***
	(−0.393)	(1.236)	(−1.960)	(1.814)	(−0.858)	(3.059)
Observations	300	300	300	300	300	300
Adjusted R^2	0.144	0.060	0.192	0.317	0.319	0.262

In this table, we report portfolio alphas and betas by regressing the returns of INT^{10K} sorted using the stocks within each of the twelve industries based on Fama and French industry classifications. Calculation of intangibles talk measure is described in Section3.2. Words used to calculate intangibles talk measure come from Filipovic and Wager (2019). Calculation of portfolio returns (INT^{10K}) is described in Section2.3. In all the regressions, the specification includes Fama and French (2015) five factors and momentum factor from Carhart (1997). We include Newey-West T-statistics. The sample is monthly from July 1995 to June 2020. All coefficients are reported in percentage per year (monthly returns multiplied by twelve).

Table 2.14: Risk-adjusted future returns (alphas) for intangibles talk measure vs. technology stocks

	Dependent variable:					
	$Tech^{portf.}$				INT^{10K}	
	(1)	(2)	(3)	(4)	(5)	(6)
$\alpha(\%)$	2.113	**6.689*****	-1.234	1.474	**2.290****	**2.552****
	(0.994)	(3.380)	(-0.819)	(0.965)	(2.134)	(2.285)
Mkt_RF	0.272***	0.119***	0.165***	0.090***	-0.031	-0.025
	(6.420)	(2.787)	(5.417)	(2.790)	(-1.359)	(-1.057)
SMB	0.246***	0.043	0.094**	0.035	0.027	-0.012
	(4.342)	(0.761)	(2.313)	(0.817)	(0.918)	(-0.373)
HML	-1.055^{***}	-0.679^{***}	-0.236^{***}	-0.095	-0.285^{***}	-0.320^{***}
	(-18.706)	(-9.351)	(-3.838)	(-1.421)	(-6.789)	(-6.980)
RMW		-0.644^{***}		-0.261^{***}		-0.113^{**}
		(-8.360)		(-4.128)		(-2.381)
CMA		-0.362^{***}		-0.318^{***}		0.127**
		(-3.511)		(-4.095)		(2.176)
UMD	-0.171^{***}	-0.133^{***}	-0.074^{***}	-0.064^{**}	-0.010	-0.012
	(-4.551)	(-3.905)	(-2.748)	(-2.474)	(-0.516)	(-0.607)
INT^{10K}			0.993***	0.897***		
			(17.403)	(15.063)		
$Tech^{portf.}$					0.511***	0.487***
					(17.403)	(15.063)
Observations	300	300	300	300	300	300
Adjusted R^2	0.622	0.698	0.813	0.830	0.813	0.820

In this table, the portfolio $Tech^{portf.}$ is the value-weighted portfolio that goes long in stocks belonging to the Business Equipment sector based on the Fama and French twelve industry classifications and goes short stocks from other industries at the end of June each year. Columns(1) and (2) report the abnormal returns of $Tech^{portf.}$ against the Fama and French (2015) three and five factors and momentum factor from Carhart (1997). Columns (3) and (4) includes INT^{10K} in the specifications. Columns (5) and (6) report the abnormal returns of INT^{10K} against the Fama and French (2015) three and five factors and momentum factor from Carhart (1997) including $Tech^{portf.}$ in the specifications as well. We include Newey-West T-statistics. The sample is monthly from July 1995 to June 2020. All coefficients are reported in percentage per year (monthly returns multiplied by twelve).

Table 2.15: Risk-adjusted returns (Alphas) from portfolios double sorted on idiosyncratic Volatility (IVOL) and intangibles talk

Intangibles → Talk	Low	2	3	4	High	$[H-L]$	
Low IVOL	-1.83	1.38	0.61	0.67	1.32	3.16	(1.64)
2	-2.42	0.64	-0.54	1.98	1.86	4.29	(1.64)
3	-2.51	-3.16	1.19	-2.69	5.64	**8.15**	(3.23)
4	-8.78	-2.8	-2.26	0.06	6.07	**14.84**	(3.76)
High IVOL	-9.88	-9.43	1.06	2.2	4.34	**14.22**	(3.33)
$[H-L]$	**-8.05**	**-10.81**	0.45	1.52	3.01		
	(-2.22)	(-2.73)	(0.13)	(0.59)	(1.35)		

This table reports risk-adjusted returns (alphas) from five-factor Fama and French model. In each regression, the dependent variable is the excess returns (over risk-free rate) of value-weighted portfolios double sorted (5 by 5) on intangible talk measure and idiosyncratic volatility (IVOL).The regression model is:

$$\mathrm{R}_{i,t} = \boldsymbol{\alpha_i} + \boldsymbol{\beta_{MKT}} MKT_t + \boldsymbol{\beta_{SMB}} SMB_t + \boldsymbol{\beta_{HML}} HML_t + \boldsymbol{\beta_{RMW}} RMW_t + \boldsymbol{\beta_{CMA}} \mathbf{CMA_t} + \boldsymbol{\beta_{UMD}} \mathbf{UMD_t} + \epsilon_{\mathbf{i,t}}$$

Where $R_{i,t}$ is the excess return (over risk-free rate) of a value-weighted portfolio in month t that is long in stocks belonging to one of the 25 groups. Columns show results from lowest to highest quantile of intangible talk measure. Rows show results from lowest to highest quantile of IVOL. Portfolio returns calculation is described in Sections 2.3. We measure firm-level IVOL by estimating the volatility of the residuals from the three factors of Fama and French for daily returns. Only the alphas from the regressions are reported for expositional purposes. We include Newey-West T-statistics in parentheses. The last column displays alphas from going long on highest quantile (5^{th}) and going short on the lowest quantile (1^{st}) of intangibles talk measure for IVOL quantiles. The sample is monthly from July 1995 to June 2020.

Table 2.16: Risk-adjusted returns (Alphas) from portfolios double sorted on idiosyncratic Volatility (IVOL) and intangibles talk, by firm size

Small Firms

Intangibles → Talk	Low	2	3	4	High	$[H-L]$	
Low IVOL	-0.07	2.09	0.6	1.1	-0.61	-0.54	(-0.22)
2	1.91	1.51	0.15	-0.5	3.38	1.47	(0.76)
3	-0.61	0.17	-0.73	-0.65	2.97	3.58	(1.89)
4	-4.09	-0.61	-0.02	-1.2	3.36	**7.45**	(2.84)
High IVOL	-7.87	-5.41	-3.82	-1.41	4.02	**11.89**	(3)
$[H-L]$	**-7.8**	**-7.5**	-4.42	-2.51	4.63		
	(-2.13)	(-2.35)	(-1.5)	(-0.87)	(1.33)		

Big Firms

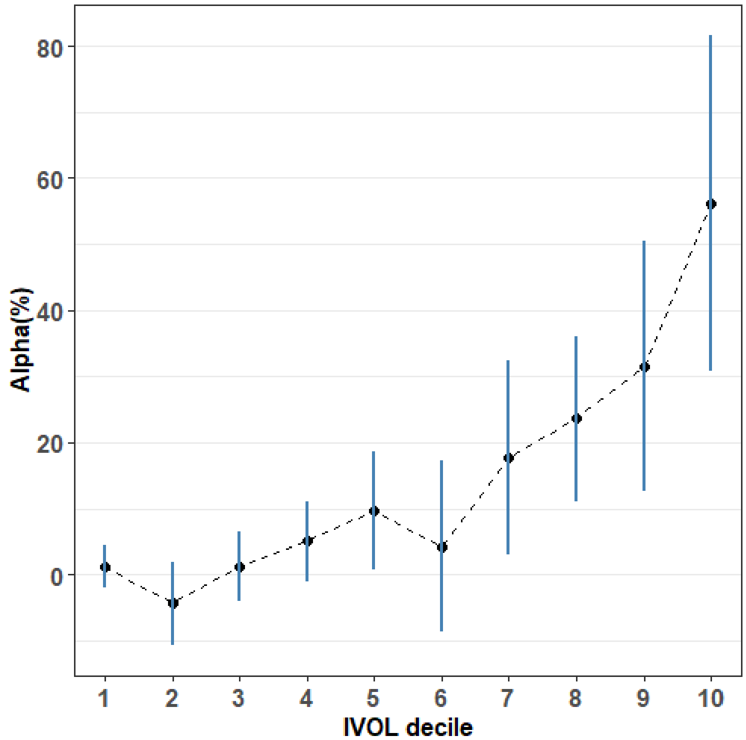
Intangibles → Talk	Low	2	3	4	High	$[H-L]$	
Low IVOL	-1.86	1.32	0.61	0.63	1.33	3.19	(1.64)
2	-2.62	0.81	-0.61	2.11	1.82	4.43	(1.63)
3	-2.54	-3.57	1.51	-3.1	5.79	**8.33**	(3.04)
4	-9.93	-3.94	-3.2	0.34	6.56	**16.50**	(3.75)
High IVOL	-13.07	-9.9	6.69	3.14	6.4	**19.47**	(3.28)
$[H-L]$	**-11.22**	**-11.22**	6.09	2.51	5.07		
	(-2.06)	(-2.23)	(1.21)	(0.55)	(1.76)		

This table reports Table 2.15 results by dividing sample at the median of firm size. This table reports risk-adjusted returns (alphas) from five-factor Fama and French model. In each regression, the dependent variable is the excess returns (over risk-free rate) of value-weighted portfolios double sorted (5 by 5) on intangible talk measure and idiosyncratic volatility (IVOL).The regression model is:

$$\mathrm{R}_{i,t} = \boldsymbol{\alpha_i} + \boldsymbol{\beta_{MKT}} MKT_t + \boldsymbol{\beta_{SMB}} SMB_t + \boldsymbol{\beta_{HML}} HML_t + \boldsymbol{\beta_{RMW}} RMW_t + \boldsymbol{\beta_{CMA}} \mathbf{CMA_t} + \boldsymbol{\beta_{UMD}} \mathbf{UMD_t} + \epsilon_{\mathbf{i,t}}$$

Where $R_{i,t}$ is the excess return (over risk-free rate) of a value-weighted portfolio in month t that is long in stocks belonging to one of the 25 groups. Columns show results from lowest to highest quantile of intangible talk measure. Rows show results from lowest to highest quantile of IVOL. Portfolio returns calculation is described in Sections 2.3. We measure firm-level IVOL by estimating the volatility of the residuals from the three factors of Fama and French for daily returns. Only the alphas from the regressions are reported for expositional purposes. We include Newey-West T-statistics in parentheses. The last column displays alphas from going long on highest quantile and going short on the lowest quantile of intangibles talk measure for IVOL quantiles. The last column displays alphas from going long on highest quantile (5^{th}) and going short on the lowest quantile (1^{st}) of intangibles talk measure for IVOL quantiles. The sample is monthly from July 1995 to June 2020.

Figure 2.8: Risk-adjusted return (Alphas) from portfolios sorted on intangibles talk measure, by decile of idiosyncratic volatility (IVOL)



The table above plots the Fama-French five factor alpha of ten INT^{10K} portfolios by IVOL decile. Every year stocks are sorted into deciles based on their IVOL and INT^{10K} is constructed separately using the stocks in each decile. Calculation of portfolio returns (INT^{10K}) is described in Section2.3. We measure firm-level IVOL by estimating the volatility of the residuals from the three factors of Fama and French for daily returns. Alphas are estimated using annual returns (monthly returns multiplied by twelve), based on Fama and French regressions discussed in Section2.4. The blue line represents the ($\mu \pm 1.96 \times$ NeweyWest standard errors). The sample is monthly, covering the period between July 1995 to June 2020.

Chapter 3

The Abnormal Temperature Premium

3.1. Introduction

The record-breaking warmer temperatures, year after year, fall in the extremes of historical temperature distributions. The magnitude of heat waves across the United States is projected to continue rising with global warming, with some areas experiencing larger abnormal temperatures than others (Smith, Zaitchik and Gohlke, 2013). With the rising temperatures and abundant evidence of global warming looming on the horizon, extreme weather events and their economic impact have gained significant attention. The effects of warmer temperatures on profitability is shown to affect labor productivity (Graff-Zivin and Neidell, 2010), energy consumption (Deschênes and Greenstone, 2011), and consumer demand (Starr-McCluer, 2000). Recent findings in the literature suggest earnings sensitivity to extreme temperatures with bidirectional effects that harm some industries while benefiting others (Addoum, Ng and Ortiz-Bobea, 2023).[1]

Motivated by the observations on earnings sensitivity to extreme temperatures, I examine the link between the economic effects of abnormal temperatures and asset returns. Given the heterogeneity of exposure to abnormal temperatures across the US states, the distribution of firms' presence across those states determines their level of exposure to abnormal temperatures. This allows me to study the link between exposure to abnormal temperatures and asset returns. Thus, the main innovation in this paper is the use of textual information in Form 10-K filings to determine the geographical presence of firms across the US states. The measure of exposure to abnormal temperatures, as defined and used in this paper, are calculated as the average magnitude of abnormal temperatures across the states in which a firm has a presence.[2]

To establish that exposure to abnormal temperatures is a source of risk, I must first demonstrate that earnings are sensitive to abnormal temperatures using my measure and that exposure to larger abnormal temperatures predicts lower earnings across firms. Depending on the quarter, I show that abnormal temperatures can have both positive and

[1] Other studies on the effects of extreme temperatures on economic output include Schlenker and Roberts (2009), Jones and Olken (2010), Hsiang (2010), Fisher et al. (2012), Burke, Hsiang and Miguel (2015),Blanc and Schlenker (2017), and Addoum, Ng and Ortiz-Bobea (2020).

[2] The measure of exposure to abnormal temperatures in this paper is not an indicator of the frequency of abnormal temperatures. Higher values of exposure to abnormal temperatures represent exposure to abnormal temperatures of larger magnitude averaged across the days of a month.

negative effects on earnings. As expected, the negative economic impact of abnormal temperatures is concentrated in the third and fourth quarters, while its occurrence in the first quarter predicts higher earnings. This is because larger abnormal temperatures in warmer seasons often negatively impact earnings, while smaller anomalies in colder seasons can boost profitability. However, exposure to larger abnormal temperatures, when averaged across quarters, predicts lower annual earnings, which suggests that the negative economic effect of abnormal temperatures prevails when pooled across quarters. The magnitude of coefficients in annual regressions suggest that one standard deviation increase in abnormal temperatures corresponds to a reduction in annual earnings per share that is 5% of its sample average.

The negative economic effect of abnormal temperatures on earnings is only present in five industry groups. These are energy, business equipment, utilities, healthcare, and ”other” industries based on the twelve Fama and French industry classifications. The concentration of this effect among a few industries is similar to the findings in Starr-McCluer (2000) and Addoum, Ng and Ortiz-Bobea (2023) who show that extreme temperatures impact earnings in 40% of industries based on the Global Industry Classification Standard (GICS). Further examination of various components in annual earnings, such as revenues and operating expenses, reveals that the negative economic impact of abnormal temperatures can be attributed to two channels, namely consumer demand and energy consumption. In industries such as business equipment and energy, exposure to larger abnormal temperatures predicts higher selling, general, and administrative expenses (SG&A), which is likely linked to an increase in utility bills as a result of higher energy consumption. Exposure to larger abnormal temperatures also predicts lower revenues in the transportation and hospitality industries, which indicates the impact of abnormal temperatures on profitability through consumer demand channels.

I also examine the potential amplifying role of geographic dispersion in the negative economic effects of abnormal temperatures.[3] By splitting the sample into two subsamples based on the median value of geographic dispersion, I observe that abnormal temperatures predict lower earnings more strongly for firms with low geographic dispersion (more localized). This makes intuitive sense since the profitability of more localized firms relies heavily on fewer states, making them more exposed to abnormal local weather patterns in those states.

The results on the economic effect of abnormal temperatures set the stage for investigating the implications of differential exposure to abnormal temperatures across firms for asset returns. I examine the risk premium implications of abnormal temperatures in the presence of investor uncertainty about the magnitude of future abnormal temperatures, which results from climate change and its impact on the distribution of abnormal temperatures. Global warming is assumed to cause abnormal temperatures of larger magnitude over time and across all states. In this setting, investors are unable to accurately predict the magnitude of future abnormal temperatures, but can observe the expected exposure to abnormal temperatures at the firm level, estimated based on historical and current records of abnormal temperatures. Under warmer climate conditions with larger abnormal temperatures, the average profitability across all firms is expected to decline, but more so for firms with a heavier presence in states with higher exposure to abnormal temper-

[3]Geographic dispersion refers to the number of unique state names that are identified in Form 10-K filings based on the definition presented in Garcia and Norli (2012). In their study, firms with the highest number of state names extracted from their Form 10-K filings are geographically dispersed as opposed to local firms with the lowest number of unique state names in their Form 10-K filings.

atures. This makes firms with higher expected exposure to abnormal temperatures more sensitive to climate change. Thus, investors with climate concerns favor the stocks of firms with lower exposure to abnormal temperatures as they are expected to perform better under climate conditions that are conducive to larger abnormal temperatures. Accordingly, I hypothesize that stocks of firms with higher expected exposure to abnormal temperatures command higher risk premiums in the presence of uncertainty about the magnitude of abnormal temperatures, as their profitability is more negatively exposed to abnormal temperatures of larger magnitude.

I turn to the equity market to test my hypothesis. To observe the abnormal temperature premium, I construct a portfolio – henceforth hot-minus-cold (HMC) – that takes long and short positions in the stocks of firms that fall in the top (hot leg) and bottom (cold leg) 20^{th} percentile of the market based on expected exposure to abnormal temperatures. The portfolio return analysis reveals that the abnormal temperature premium, although positive, is insignificant for the period between 1994 and 2021. However, this has changed recently as the portfolio's returns have grown rapidly along with growing climate concerns. The portfolio generates an average return of 62 bps per month with a Sharpe ratio of 0.25 between January 2015 and December 2021. This recent outperformance of stocks with the highest expected abnormal temperatures (hot stocks) is worthy of attention as it relates the abnormal temperature premium to investors' climate concerns. The common risk factors do not explain the outperformance of hot stocks for the period after 2015. The monthly alpha over Fama and French (2015) three-factor model plus the momentum factor from Carhart (1997) is 50 bps, and adding other common risk factors such as profitability, investment, and liquidity does not capture the alpha. The returns for each leg of the portfolio reveal that the long leg of the portfolio generates a monthly alpha of 24.4 bps, while the short leg generates a monthly alpha of -25.7 bps, indicating that investors demand a premium for exposure to abnormal temperatures.

Examination of HMC's performance for industries with and without earnings sensitivity to abnormal temperatures supports my earlier results, suggesting that abnormal temperatures negatively impact earnings in only five industries. Since investors' demand for the abnormal temperature premium is linked to their concerns about its negative economic effects, therefore, the premium is expected to be present only among industry groups with earnings sensitivity to abnormal temperatures. I test this by constructing two separate HMC portfolios using the stocks belonging to industry groups with and without sensitivity to abnormal temperatures. The results suggest that HMC's alpha is only statistically significant when constructed using stocks of the five industries with earnings negatively impacted by abnormal temperatures. The portfolio's alpha is 74 bps per month over Fama and French (2015) three-factor model plus the momentum factor from Carhart (1997), which is larger than the previously reported alpha of 50 bps for the portfolio that included all industries.

Similarly, following the results on the stronger negative effects of abnormal temperatures on profitability among firms with low geographic dispersion, I construct and examine HMC's performance separately for firms with low and high geographic dispersion. Once again, if the abnormal temperature premium is linked with the negative effects of abnormal temperatures on profitability, it is expected to find a larger premium among the stocks of firms with low geographic dispersion. The results show that the alpha of HMC is only significant using the stocks of local (low geographic dispersion) firms with a magnitude of 1.04% per month. The alpha, however, is not significantly different from zero for the stock of firms that are more geographically dispersed (high geographic dispersion). HMC's strong performance within industries sensitive to abnormal temperatures, and especially among firms with low geographic dispersion, closely aligns with my earlier results on the relation between abnormal temperatures and earnings.

Nevertheless, the negative effect of abnormal temperatures on earnings does not explain the recent strong performance of HMC. The analysis of the effect of abnormal temperatures on earnings before and after 2015 suggests that the effect has always been present, which raises the question of why there has been a sudden rise in the abnormal temperature premium in recent years. To answer this question, I closely examine the link between the abnormal temperature premium and investors' climate concerns as the factor that derives the abnormal temperature premium for stocks with earnings sensitivity to abnormal temperatures. I do this by investigating the premium's response to unexpected shocks to climate concerns. An unexpected rise in climate concerns is expected to amplify concerns about larger abnormal temperatures associated with global warming and urge investors to demand a higher premium for exposure to abnormal temperatures. Therefore, we should observe a strengthening in HMC's performance during or after the shocks to climate concerns.

I test this using a proxy for climate concerns based on the recently developed measure of the Media Climate Change Concerns Index ($MCCC$) from Ardia et al. (2022). The $MCCC$ index gauges climate-related concerns based on the major US newspapers and has multiple categories corresponding to various climate change topics. Similar to Pástor, Stambaugh and Taylor (2022), I use the error terms collected from the AR(1) model on the $MCCC$ index as the proxy for shocks to climate concerns. I find that the abnormal temperature premium positively responds to unexpected shocks in climate concerns in the previous month.

Further analysis on each leg of the portfolio suggests that climate concerns shocks impacts only the cold leg, representing firms with the lowest expected exposure to abnormal temperatures. The results show that an increase in climate concerns predict lower returns for the cold leg in the subsequent month, suggesting an increase in demand for cold stocks in the aftermath of the shock, which in turn, leads to their overpricing and lower returns in the subsequent month. Therefore shocks to climate concerns drive the variations in the abnormal temperature premium by impacting the stocks in the cold leg. This confirms my earlier observation on the rapid growth of the premium in recent years which coincides with an increase in climate concerns. This result confirms the link between investors' climate concerns and the abnormal temperature premium and also supports my earlier observation of the rapid growth of the premium in recent years, which coincides with an increase in climate concerns.

My work primarily contributes to the literature investigating climate change risks and their implications for asset prices. The two main categories of climate risks are transition and physical risks (Giglio, Kelly and Stroebel, 2021; Faccini, Matin and Skiadopoulos, 2023). The growing literature surrounding sustainable investment aligned with climate concerns primarily focuses on the transition risks that are associated with investor preferences for a low-carbon economy and green versus brown stocks (Bolton and Kacperczyk, 2021; Pástor, Stambaugh and Taylor, 2021; Ardia et al., 2022; Pástor, Stambaugh and Taylor, 2022; Hsu, Li and Tsou, 2023; Aswani, Raghunandan and Rajgopal, 2023). This study, however, belongs to the strand of literature that addresses the physical climate change risks that are concerned with the direct economic impacts of a warmer climate (Bansal and Ochoa, 2011; Bansal, Kiku and Ochoa, 2016; Co-

Iacito, Hoffmann and Phan, 2019; Alok, Kumar and Wermers, 2020; Cuculiza et al., 2022; Goldsmith-Pinkham et al., 2023).

The methodology in this study differs from that of Addoum, Ng and Ortiz-Bobea (2023), which also studies extreme temperatures' economic impact in two ways. First, Addoum, Ng and Ortiz-Bobea (2023) only accounts for the location of firms' headquarters, while here, I estimate monthly firm-level expected abnormal temperatures based on all the state names that appear in the most recent Form 10-K filing associated with each firm. Second, Addoum, Ng and Ortiz-Bobea (2023) focuses on the effects of both extremely warm and cold weather conditions on earnings within each industry. This study, however, focuses on the effects of only abnormal temperatures on profitability.

Another strand of related literature is those that study the impact of temperature anomalies on the stock market through behavioral channels linked to investors' mood and belief (Cao and Wei, 2005; Jacobsen and Marquering, 2008; Choi, Gao and Jiang, 2020). Choi, Gao and Jiang (2020) show that during abnormally high local temperatures, investors revise their beliefs and avoid climate-unfriendly stocks in the local exchange. This leads to the underperformance of carbon-intensive firms during warmer-than-usual months. In contrast, I focus on stock return performance that is linked to the firm's fundamentals as a function of abnormal temperatures, rather than investors belief. I conduct my returns analysis using the main US exchanges, NYSE, Amex, and Nasdaq, all operating from the same city, to avoid the potential effects of different exchange locations as suggested by Choi, Gao and Jiang (2020).

This paper proceeds as follows. In Section 3.2, I describe the methodology used to construct the measures of state and firm-level abnormal temperatures and report the distribution of values across states and in relation to other firm characteristics. Section 3.3 presents the results on the relation between abnormal temperatures and earnings and earning announcement returns. Section 3.4 reports the portfolio analysis results and the abnormal temperature premium. Section 3.5 examines the correlation between the abnormal temperature premium and climate change concerns. Section 3.6 concludes.

3.2. Data and methodology

The textual data for my analysis comes from Form 10-K filings submitted to the SEC by 12,184 public firms from January 1994 to December 2021. This adds up to a total of approximately 107,000 Form 10-K filings in our analysis. The filings are collected using the Electronic Data Gathering, Analysis, and Retrieval ($EDGAR$) system.[4]

Following Garcia and Norli (2012), I gather the unique state names from each filing and remove the filings with no state names from the sample. Hence, the sample constructed includes firm-year observations, with each observation containing the unique names of the states that appeared in the filing. Fig.3.1 presents the distribution of state names across the filings in the sample. As can be observed, the distribution of state names across filings is skewed positively, with more than half of the filings containing less than ten unique state names. Since I use the state names in the filings as a proxy for the firm's presence across different states, the density in Fig.3.1 suggests more than half of the firms in the economy have a presence in less than ten states.

[4] I use the parsed documents publicly available on the Loughran-McDonald website at: https://sraf.nd.edu/loughranmcdonald-master-dictionary/

To construct the firm-level abnormal temperature, which is a proxy for exposure to abnormal temperatures, I first define and calculate the monthly abnormal temperature exposure at the state level using a methodology similar to Choi, Gao and Jiang (2020). For each state-month observation, abnormal temperature is the deviation of that month's temperature from the historical averages for the same month. abnormal temperature exposure for state s, at month m, and year t is defined as the current month's temperature minus its historical average in the same state and for the same month:

$$\textit{Abnormal Temperature}_{\mathbf{s,m,t}} = \textit{Temperature}_{\mathbf{s,m,t}} - \frac{\mathbf{1}}{\mathbf{t-1}} \sum_{\mathbf{y=1}}^{\mathbf{t-1}} \textit{Temperature}_{\mathbf{s,m,y}} \tag{3.1}$$

The historical monthly state-level temperature data goes as far back as 1895 and is publicly available on the National Centers for Environmental Information (NOAA) website.[5] Fig.3.2 presents the average exposure to abnormal temperature for each state by color, and the distribution of exposure to abnormal temperatures at the state level from 1994 to 2021 is presented in Fig. 3.3. As can be observed in Fig. 3.2 and Fig. 3.3, southern states, which are typically known to have higher average temperatures, are not necessarily exposed to larger abnormal temperatures. This is because the underlying mechanisms behind abnormal temperatures depend on the specific atmospheric circulation patterns, which are a function of the region's land surface features and not necessarily the average temperatures (Yoon et al., 2018; Wu et al., 2023).

Fig. 3.4 plots the distribution of abnormal temperatures across all the states by year. Since abnormal temperatures are estimated using the average temperatures up until the previous year, a particularly hot year must push the average temperatures upward. Since the abnormal temperatures in the subsequent years are calculated as the deviation from historical averages, it is expected that the median value of abnormal temperatures to be close to zero or negative following a year with abnormally high temperatures. However, as can be seen in Fig.3.4 this is not the case. The median values of abnormal temperature distribution across states are almost always above zero in recent years. This indicates that the median of temperature deviations from average temperatures has been consistently positive in recent years, which is strong evidence supporting global warming.

Using monthly state-level abnormal temperatures, I estimate the firm-level exposure to abnormal temperatures for each month by averaging the abnormal temperatures across the states that appeared in the firm's most recent Form 10-K filing, submitted to the SEC in the past twelve months. Therefore, I assume that firms change their presence across states once a year. Another implicit assumption is that firms have an equal presence across the states that appear in their Form 10-K filing, while in reality, firms' revenues may depend on some states more than others. I assume equal presence since the exact revenues and profitability data linked to each state are unavailable at the firm level. abnormal temperature for firm i, at month m, and year t is defined as the average state-level abnormal temperature across N states where firm i has a presence in:[6]

[5]The National Centers for Environmental Information is a US government agency that manages one of the world's largest archives of atmospheric, coastal, geophysical, and oceanic data. The data can be found at: NOAA (2024)

[6]N is the same as geographic dispersion, the measure developed by Garcia and Norli (2012) which is the number of unique states that appear in a Form 10-K filing.

$$\textit{Abnormal Temperature}_{\mathbf{i,m,t}} = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{s=1}}^{\mathbf{N}} \textit{Abnormal Temperature}_{\mathbf{s,m,t}} \tag{3.2}$$

Using the equation above, I calculate the monthly firm-level exposure to abnormal temperatures in my sample. A limitation of this measure is that the degree of granularity for a firm's locations is limited to the geographical borders of the US states. Therefore, regardless of the exact location of a firm within a given state, the average weather conditions across various regions of that state are assumed to impact the firm's profitability equally. Although not ideal, this design is due to the limitation of the textual methodology that only extracts state names in Form 10-K filings. I use the firm-level abnormal temperature to examine the effects of abnormal temperatures on earnings. In particular, for each firm-year observation, I average the firm-level abnormal temperature over the twelve months of the fiscal period and use it as the proxy for exposure to abnormal temperatures in that period.

3.3. Earnings and abnormal temperatures

My analysis of the economic impact of abnormal temperatures starts by examining how exposure to abnormal temperatures, averaged across the states where firms have a presence, affects their earnings. My analysis is motivated by the observations on the sensitivity of profitability to weather conditions such as extreme temperatures (see Addoum, Ng and Ortiz-Bobea, 2023). I conduct my analysis using earnings at both annual and quarterly levels. It is expected that abnormal temperatures would impact earnings differently depending on the season in which they occur. Addoum, Ng and Ortiz-Bobea (2023) show that abnormal temperatures can have both a positive and negative effect on earnings depending on its occurrence in the first and third quarters.

I use the following panel regression to estimate the effects of exposure to abnormal temperatures on earnings:

$$\text{EPS}_{i,t} = \boldsymbol{\beta_0} + \boldsymbol{\beta_{temp}} Abnormal\,Temperature_{i,t} + \boldsymbol{\beta} X_{i,t-1} + \delta_{ind} + \theta_{time} + \epsilon_{i,t} \tag{3.3}$$

The dependent variable $EPS_{i,t}$ is either annual or quarterly earnings per share scaled by the beginning of the fiscal period stock price. $Abnormal\,Temperature_{i,t}$ is the monthly firm-level abnormal temperatures calculated based on eq.3.2 averaged over twelve or three months that cover the fiscal year or quarter t. $X_{i,t-1}$ are the control variables based on Fama and French (2000) that include the firm-level book-to-market, size, indicators of loss and dividend, book value of leverage, and dividend yield for firm i at the beginning of the fiscal period. δ_{ind} and θ_{time} capture the forty-eight Fama-French industry and time fixed effects.

I test the effects of exposure to abnormal temperatures on quarterly earnings by running a pooling regression based on eq. 3.3 for each calendar quarter in my sample. The results are reported in Table 3.1 for the sample period between 1994 and 2021. The coefficient of abnormal temperature β_{temp} is positive for the first quarter, suggesting that exposure to larger abnormal temperatures boost earnings during the winter months. β_{temp} is insignificant for the second quarter and negative for the third and fourth quarters, which corresponds to months during the summer and fall. The results are in line with the expectation that abnormal temperatures in the first and second quarters are based on temperatures that are considered extreme relative to average temperatures during cold months yet still considered to be within the normal range relative to average temperatures in warmer months. Therefore, exposure to abnormal temperatures in colder months affects operating costs to a lesser extent since such costs rise with the energy consumption needed to keep the temperature within the normal range suitable for operations. Accordingly, a particularly hot winter month may not only raise energy consumption for cooling purposes but may even lower energy consumption for heating purposes, which typically adds to operating costs in colder months. The magnitude of the coefficient for the first quarter suggests that a standard deviation increase in abnormal temperatures corresponds to a rise in earnings per share that is as large as 4% of its sample average. abnormal temperatures are larger during warmer months and push the temperatures to the upper extremes of historical temperature distribution. For this reason, abnormal temperatures during warmer months are expected to have a bigger economic effect on earnings relative to older months. This can be seen from the magnitude of coefficients in the third quarter, which suggests that one standard deviation increase in abnormal temperatures corresponds to a drop in earnings per share as large as 11% of its sample average.

Examining the effect of exposure to abnormal temperatures on annual earnings also shows that the negative economic effect of abnormal temperatures dominates across quarters. The results from annual earnings per share reported in Table 3.2 show that the coefficient of abnormal temperature β_{temp} is significant and negative, suggesting that larger abnormal temperatures averaged over the course of a fiscal year predicts lower earnings. The magnitude and significance of β_{temp} do not vary drastically after controlling for other predictors of earnings discussed in the literature. The result in column (4) shows that a standard deviation increase in the magnitude of abnormal temperatures corresponds to a drop in earnings per share that is as large as 5% of its sample average. This suggests that the average effect of exposure to abnormal temperatures across quarters is negative on profitability.

The overall results from Table 3.2 and Table 3.1 suggest that despite the positive effects of abnormal temperatures on profitability in the first quarter, on average, exposure to abnormal temperatures is a predictor of lower earnings in the economy. A firm that operates in a state with exposure to large abnormal temperatures may benefit from the anomalies during the winter months. Yet, the overall effect of presence in that state is negative for its profitability. Since firms are not expected to alter their presence across states every quarter, I argue that the effect of exposure to abnormal temperatures on annual earnings is more important from investors' perspective.

3.3.1 abnormal temperature and earnings across industries

To break down the channels through which abnormal temperatures impact earnings, I investigate their effect on revenues and operating costs within industries based on Fama and French twelve industry classifications. I use the regression specified in eq. 3.3 to estimate the effect of abnormal temperatures on annual earnings within each industry. The results are reported in Table 3.3. The first column in the table shows that abnormal temperatures have a significant negative impact on annual earnings only among five industries, including energy, business equipment, utilities, health care, and 'other' industries.

Starting from the energy industry, the results in the last column show that abnormal temperatures negatively impact earnings by raising selling, general, and administrative (SG&A) components of operating expenses. Since utility bills are reported under SG&A, this indicates that abnormal temperatures lead to an increase in energy consumption to maintain temperatures within the normal range. The same effect of abnormal temperatures on SG&A can be observed for business equipment and healthcare industries. The positive effect of abnormal temperatures on SG&A in business equipment industry, which includes technology firms, can be linked to rising electricity bills during warmer months in California, where most technology firms have a presence. Technology firms rely heavily on electricity consumption for cooling purposes associated with heat–generating operations in data centers and large-scale computational facilities.

abnormal temperatures also negatively impact revenues in utilities and 'other' industries. The effect of abnormal temperatures on revenues is linked to their impact on consumer demand. Revenues of businesses that heavily rely on local demands are subject to temperature anomalies in the regions of operation. A downward shock to consumer demand is due to consumers' preference for indoor activities during abnormal temperatures (see Starr-McCluer, 2000; Addoum, Ng and Ortiz-Bobea, 2023). Hotels are an example of such businesses that fall under 'other' industries category with revenues negatively impacted by abnormal temperatures, according to my results. Similarly, revenues in the utilities industry depend on energy consumption, which is linked to temperature conditions. Addoum, Ng and Ortiz-Bobea (2023) shows that abnormal temperatures do not significantly increase revenues for firms in electric utilities but negatively impact revenues for firms in gas utilities.

Overall, the results from this section help identify industries that are negatively susceptible to abnormal temperatures. This has important implications for investigating the presence and size of the abnormal temperature premium among stocks from different industries with and without sensitivities to abnormal temperatures, which will be discussed in later chapters.

3.3.2 Geographic dispersion effect

In this section, I examine whether the degree of geographic dispersion influences the magnitude of economic effects resulting from exposure to abnormal temperatures. Geographic dispersion, or the number of unique states that appear in Form 10-K filings, is inversely related to the measure of abnormal temperature in eq.3.2. It is expected that a larger portion of the earnings of firms with lower geographic dispersion (more local) is impacted by the occurrence of abnormal temperatures in the states in which they have a presence. This implies that having a presence across a larger number of states (more geographically dispersed) leads to a diversification effect that mitigates the risk of exposure to larger abnormal temperatures.

I test this by separately estimating the coefficients from eq.3.3 for the sample of firms with varying degrees of geographic dispersion. I pick the median value of geographic dispersion in the sample as the breaking point to identify firms with low versus high geographic dispersion. The results are reported in Table 3.4. As predicted, the negative economic effect of exposure to abnormal temperatures is larger among firms with low geographic dispersion. Both the magnitude and significance of β_{temp} are larger for firms with low geographic dispersion. Although the negative economic impacts of abnormal temperatures are significant for both low and high geographic dispersion firms, the effects are twice as large among firms with low geographic.

The larger economic effect of abnormal temperatures among low geographic dispersion firms may also arise due to the possibility that the variable $Abnormal\,Temperature_{i,t}$ in eq.3.3 varies more across firms with low geographic dispersion. This may arise from the inverse relation between the measure of exposure to abnormal temperatures and geographic dispersion based on eq.3.2, which can cause the firm-level measure of exposure to abnormal temperatures to vary more across firms with low geographic dispersion which leads to a larger estimated β_{temp} for these firms. However, the distribution of the variable $Abnormal\,Temperature_{i,t}$ in Fig.3.5 shows that its magnitude and variation are similar amongst both low and high geographic dispersion firms. This confirms that the results from Table 3.4 suggest that the economic effect of abnormal temperatures is larger and grows in a non-linear fashion among firms with low geographic dispersion.

The effect of geographic dispersion on the interaction between abnormal temperatures and earnings is important for verifying my hypothesis on the presence of the abnormal temperature premium. Assuming that the abnormal temperature premium exists and arises from the negative effects of abnormal temperature on profitability, then a larger premium is expected to be observed among firms with low geographic dispersion.

3.4. The abnormal temperature premium

I examine the returns of a factor mimicking portfolio constructed based on the expected abnormal temperature exposure to investigate the presence of the premium associated with exposure to abnormal temperatures. At the beginning of each month, the expected abnormal temperature for each state is available and calculated as the average of historical abnormal temperatures in that state up until last year

$$Expected\ Abnormal\ Temperature_{\mathbf{s,m,t}} = \frac{\mathbf{1}}{\mathbf{t-1}}\sum_{\mathbf{y=1}}^{\mathbf{t-1}} Abnormal\ Temperature_{\mathbf{s,m,y}} \tag{3.4}$$

Similarly, the firm-level expected abnormal temperature is defined as the following based on the state-level expected abnormal temperature exposure:

$$Expected\ Abnormal\ Temperature_{\mathbf{i,m,t}} = \frac{\mathbf{1}}{\mathbf{N}}\sum_{\mathbf{s=1}}^{\mathbf{N}} Expected\,Abnormal\ Temperature_{\mathbf{s,m,t}} \tag{3.5}$$

I follow the long-short strategy to construct HMC, a value-weighted portfolio that takes long and short positions in the top and bottom 20^{th} percentile of stocks in the sample based on the normalized firm-level measure of expected abnormal temperature.[7] I use a sample of NYSE, Amex, and Nasdaq stocks, with data available from the Center for Research in Security Prices (CRSP). The long or hot leg and short or cold leg represent stocks of firms expected to experience the highest and lowest abnormal temperatures in the economy. The portfolio is re-balanced monthly and uses information about the firm's geographical locations extracted from its most recent Form 10-K filing, submitted to the SEC in the past twelve months. Firm-year observations without any state names in the Form 10-K filing are dropped from the sample. The information about the geographical location of firms is updated once every year, corresponding to each Form 10-K filing that is submitted to the SEC annually. The firm-level expected abnormal temperatures are,

[7] I normalize the firm-level expected abnormal temperature every month based on its values across firms.

however, updated every month since the state-level abnormal temperature values vary on a monthly basis. The sample period starts from April 1994 and continues until December 2021.

The cumulative value-weighted returns for hot and cold legs of HMC are separately displayed in Fig.3.6. The hot and cold value-weighted portfolios take only long positions in the highest and lowest 20^{th} percentile of stocks in the market based on monthly observations of expected abnormal temperature across firms. The cumulative returns show that hot stocks outperform cold stocks for most of the sample period. In addition, the return difference between hot and cold stocks has grown significantly in recent years. Fig.3.7 plots the distribution of HMC's monthly returns within each quarter over the sample period. The quarterly distribution of portfolio returns shows that the median returns are above zero for the third and fourth quarters, unlike the first two quarters. This is the first evidence that links the abnormal temperature premium to the negative economic effect of abnormal temperatures. The results in Table3.1 show that quarterly earnings are only negatively correlated with abnormal temperatures in the third and fourth quarters. Therefore, it is expected to observe a larger investor demand for the premium linked with exposure to abnormal temperatures during the third and fourth quarters of the year.

The cumulative return of HMC in recent years is displayed in Fig.3.8 against the cumulative shocks to climate news based on the $MCCC$ index. I also include the cumulative return of green–minus–brown (GMB) portfolio from Pástor, Stambaugh and Taylor (2022) which takes long and short positions based on the firm's environmental, social, and governance (ESG) criteria.[8] As can be observed, the cumulative return of HMC trends downward until near the end of 2012, indicating a reversal in the sign of the abnormal temperature premium. Similarly, the cumulative climate news shocks and GMB trend upward near the end of 2012, indicating a significant increase in climate concerns in that period. This sudden rise in climate concerns coincides with The Doha United Nations Climate Change Conference in November 2012. In addition, the cumulative return of HMC eventually trended upward in November 2015 around the Paris Climate Accord, when climate concerns reach their highest values since 2012.

Pástor, Stambaugh and Taylor (2022) argue that GMB' return is associated with elevated levels of investor concerns following bad news about climate change. Accordingly, following a rise in climate news shocks after 2012, increasingly concerned investors demand greener stocks, which causes GMB's strong performance. I argue that the elevated investor climate concerns also drive HMC's strong performance after 2015. In the following chapters, I offer a more detailed analysis of HMC's returns for the period before and after 2015 and further investigate whether the rising climate concerns drive HMC's return in recent years.

3.4.1 Returns analysis

The average return, Sharpe ratio, and alpha of HMC are reported in Table 3.5 for the period between 1994 and 2021 and separately for the sub-periods before and after 2015. The return for the full sample period is positive but insignificant. However, for the period after 2015 the abnormal temperature premium grows larger and becomes significant, reaching 62 bps per month. The portfolio's alpha reported in Table3.5 is estimated over the Fama and French (2015)

[8] GMB is a long-short portfolio that selects stocks based on their greenness score developed by Pástor, Stambaugh and Taylor (2022), which is calculated using the firms' environmental, social, and governance (ESG) criteria. The upward trend in GMB starts around the year 2012, which coincides with the rise in cumulative climate news shocks based on the $MCCC$ index.

three factors plus the momentum factor from Carhart (1997). The alpha for the period after 2015 is 50 bps per month, with the alpha of the hot and cold legs being 24.4 bps and -25.7 bps per month. The Sharpe ratio of the portfolio for the period after 2015 is 0.25, with the hot and cold legs having a Sharpe ratio of 0.32 and 0.19, respectively.

The return results from Table 3.5 confirm the observation from Fig.3.8 showing that the abnormal temperature premium has strengthened significantly in recent years. Also, despite the positive reported returns for each leg of HMC, the alphas indicate that the cold leg, which is constructed using the stocks of firms with the lowest expected abnormal temperature, commands a negative risk premium when accounting for the common risk factors. Similarly, the positive alpha for the hot leg indicates that investors demand a premium for stocks of firms with the highest expected abnormal temperatures in recent years. Therefore, the results confirm the presence of the abnormal temperature premium in the market, but only for the period after 2015.

With a focus on the period after 2015, I further examine HMC's alpha by including other common risk factors discussed in the literature. I use the following Fama and French regression to estimate the alphas

$$\begin{gathered}\mathbf{R_t} = \boldsymbol{\alpha} + \boldsymbol{\beta_{MKT}}\mathbf{MKT_t} + \boldsymbol{\beta_{SMB}}\mathbf{SMB_t} + \boldsymbol{\beta_{HML}}\mathbf{HML_t} + \boldsymbol{\beta_{RMW}}\mathbf{RMW_t} + \\ \boldsymbol{\beta_{CMA}}\mathbf{CMA_t} + \boldsymbol{\beta_{UMD}}\mathbf{UMD_t} + \boldsymbol{\epsilon_t}\end{gathered} \tag{3.6}$$

where R_t is the return of HMC_t in month t, and α is the intercept that captures the risk-adjusted returns. MKT_t, SMB_t, HML_t, RMW_t, CMA_t, and UMD_t are the returns of the market, size, value, profitability, investment, and momentum portfolios all taken from Ken French's website. The standard errors are estimated using Newey and West (1987), which allows for serially correlated and heteroskedastic error terms.

The alphas are reported in Table 3.6 for the period between January 2015 and December 2021. Alphas remain statistically significant for all specifications, including the Fama and French (2015) three and five factors plus the momentum factor from Carhart (1997). The alpha is also significant in column (4) when I include the traded liquidity factor, LIQ, from Pástor and Stambaugh (2003). I include LIQ since firms with low geographic dispersion tend to be smaller and less liquid. In the final column, I also include the premium for geographic dispersion, which is the returns of the portfolio that is sorted by taking long and short positions in the bottom and top 20^{th} of stocks based on geographic dispersion. This is done to address the documented premium associated with stocks of local firms (low geographic dispersion) in the literature (see Garcia and Norli, 2012). [9] Since the measure of exposure to abnormal temperatures is inversely related to geographic dispersion in eq.3.2, higher values of exposure to abnormal temperatures sometimes indicate lower values of geographic dispersion among local firms. Therefore, the abnormal temperature premium may be a compensation for low geographic dispersion with less recognition as predicted by Merton et al. (1987). However, the significant alpha in column (5) shows that the premium for geographic dispersion does not explain the abnormal temperature premium that emerged in recent years.

The coefficients for risk factors included across various specifications in Table 3.6 are mostly insignificant. The

[9] The premium for low geographic dispersion is an implication of the equilibrium model developed by Merton et al. (1987). In this model, in a market where investors are not aware of all the stocks, stocks of firms that receive less investor recognition offer higher expected returns to the few investors who hold the stock as compensation for insufficient diversification in their portfolio. Garcia and Norli (2012) document the premium for local firms since they are more likely to receive less attention from investors.

lowest alpha stands at 50 bps per month and is reported in column (2). The highest R^2 is reported in column (3), where the coefficients for RMW and CMA are both negative and significant. This suggests that HMC takes long positions in stocks of firms with relatively low profitability and higher investments, which results in the portfolio's negative exposure to RMW and CMA factors.

The positive alphas reported in Table 3.6 suggest that hot stocks have outperformed cold stocks in recent years, similar to the recent outperformance of green stocks relative to brown stocks documented by Pástor, Stambaugh and Taylor (2022). They show that green stocks outperform brown stocks due to a sudden rise in climate concerns after 2012. Following a similar observation for HMC's cumulative returns in recent years (see Fig. 3.8), in the following chapters, I examine investors' climate concerns as a potential candidate for explaining HMC's strong performance after 2015.

3.4.2 The abnormal temperature premium within industry groups

The results reported in Table 3.3 show that the negative economic effects of abnormal temperatures are concentrated among firms that belong to the five industries, including energy, business equipment, utilities, health care, and 'other' industries. This indicates that the premium for abnormal temperature is expected to be concentrated among the stocks of firms belonging to these industries. To test this, I construct HMC separately using only the stocks from industries with and without earnings sensitivity to abnormal temperatures based on the results in Table 3.3. Comparing the portfolio's returns would reveal whether investors' demand for the abnormal temperature premium is associated with its negative economic effects on earnings. Since exposure to abnormal temperatures does not economically impact firms in industries with no earnings sensitivity to abnormal temperatures, it is expected to observe no premium for exposure to abnormal temperatures among the stocks of these firms.

Cumulative returns of HMC constructed within industry groups with and without sensitivity to abnormal temperatures are shown in Fig. 3.9 for the period after 2015. An investment of one dollar in the portfolio of stocks with sensitivity to abnormal temperatures generates a cumulative return of 80% from January 2015 to December 2021. In comparison, the same investment in the portfolio of stocks without sensitivity to abnormal temperatures generates a cumulative return of roughly 20% for the same period. Also, cumulative returns for the portfolio of stocks with sensitivity to abnormal temperatures follow a steady growth after 2015 without periods of significant drops in returns. This, however, is not the case for the portfolio that is constructed using the stocks without sensitivity to abnormal temperatures, as for most of the period, the returns fluctuate and cumulative returns remain below 5%.

The estimation of HMC's alphas over various specifications for the stock with and without sensitivity to abnormal temperatures is reported in Table. 3.7. The results show that HMC's alphas are only significant for the portfolio of stocks with sensitivity to abnormal temperatures. The lowest reported alpha is 74 bps per month over the four-factor specification, which is larger than the alpha of 50 bps per month reported Table 3.6 for the same specification. This shows that a portfolio sorting strategy based on the degree of exposure to abnormal temperatures performs better when limited to only the stocks belonging to the five industries with earnings sensitivity to abnormal temperatures. In addition, alphas are not significantly different from zero in any specifications for the portfolio of stocks that belong to industries without sensitivity to abnormal temperatures, even in recent years. Overall, the results suggest that the abnormal temperature premium is closely linked with the negative effects of abnormal temperatures on annual earnings, as reported earlier in this paper.

3.4.3 The abnormal temperature premium and geographic dispersion

The results reported in Table 3.4 show that the negative economic effects of abnormal temperatures are stronger among firms with low geographic dispersion. This indicates that the premium for abnormal temperature is expected to be larger among the stocks of firms with low geographic dispersion for their higher sensitivity to local temperature anomalies. I test this by constructing HMC separately among the stocks of firms with low and high geographic dispersion. I report the alphas of HMC for high and low geographic dispersion firms in Table 3.8 for the period after 2015.

The results show that HMC's alphas are only significant among firms with low geographic dispersion. However, the alphas are not significantly different from zero for firms with high geographic dispersion, even in recent years. The reported alphas for low geographic dispersion firms is 1% per month over the four-factor specification, which is larger than the alpha of 50 bps per month reported Table 3.6 for the same specification.

The larger magnitude of the abnormal temperature premium among low geographic dispersion firms is likely due to the amplifying effect of low geographic dispersion on the negative economic impact of exposure to abnormal temperatures. This is further evidence supporting the hypothesis that links the presence of the abnormal temperature premium to its negative effects on earnings. Moreover, the effects of geographic dispersion in exposure to abnormal temperatures have further implications about the economic advantages of higher geographic dispersion under warmer climate conditions. With the physical risks of extreme weather events varying across different locations, firms can benefit from extending their presence across more states and, as a result, diversifying the source of their revenue streams.

3.4.4 The abnormal temperature premium and technology stocks

Based on the results in Table3.3 business equipment industry, which includes technology stocks, is among the industry groups with earnings negatively impacted by exposure to abnormal temperatures. Technology firms are concentrated in California, which is one of the states with exposure to the largest abnormal temperatures according to Fig.3.3. Therefore, it is plausible that in the construction of HMC, the long positions are more tilted toward technology stocks than the short positions. This would explain the strong performance of HMC in recent years by linking it to the outperformance of technology stocks in the past decade.

To examine the presence of such an effect, I construct HMC separately using only the stocks from business equipment industry and the other four industries with earnings sensitivity to abnormal temperatures, which include energy, utilities, health care, and the 'other' industries. Further, within each industry group, I construct the portfolio separately for low and high geographic dispersion firms to account for the effect of geographic dispersion on the abnormal temperature premium within each industry group. Thus, for the business equipment industry, which includes technology stocks, two portfolios are constructed that correspond to low and high geographic dispersion firms. The same procedure is followed to construct two portfolios for the stocks in the other four industries with earnings sensitivity to abnormal temperatures. The alphas for these four portfolios are reported in Table 3.9.

The results show that the abnormal temperature premium is present among both groups of stocks, although only among low geographic dispersion firms. When business equipment stocks are removed from the sample, and the portfolio is constructed using only the stocks of the other four industries with sensitivity to abnormal temperatures, the alpha is still positive and significant. Additionally, the portfolio's positive and significant alpha using only business equipment stocks suggests that even within the technology sector, the portfolio has generated a significant and positive alpha in recent years.

3.5. The abnormal temperature premium and climate concerns

So far, the results have established that the abnormal temperature premium is linked to the negative effects of exposure to abnormal temperatures on earnings. The results in Table3.10 show that the negative effect of abnormal temperatures on earnings is present both before and after 2015. Thus, the momentum in the abnormal temperature premium cannot be explained entirely by the negative economic effect of exposure to abnormal temperatures. In this section, I address this concern and closely examine how climate concerns may have contributed to the growth in the abnormal temperature premium in recent years. I argue that a rise in climate concerns is expected to amplify investors' uncertainty about the magnitude of abnormal temperatures, which, in turn, would lead to an increase in investor demand for stocks with lower exposure to abnormal temperatures. This explains their recent outperformance against stocks with higher exposure to abnormal temperatures by linking it to the rising climate concerns.

I test the effect of rising climate concerns on the abnormal temperature premium using the error terms from the AR(1) model applied to the $MCCC$ index from Ardia et al. (2022) as the proxy for shocks to climate concerns. I use the following regression to examine the effects of shocks to climate concerns on the abnormal temperature premium. I include the climate news shock and its lag in the regression to control for the possible slow reaction of stocks to an increase in climate concerns:

$$\mathbf{R_t} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{News\,Shock_t} + \boldsymbol{\beta'}\mathbf{News\,Shock_{t-1}} + \boldsymbol{\beta_x}\mathbf{X_t} + \boldsymbol{\epsilon_t} \tag{3.7}$$

where R_t is the return of HMC_t in month t, $NewsShock_t$ and its lag are the shocks to the $MCCC$ index in the current and previous months, and X_t are other risk factors that are predictors of returns.

The results from the regression are reported in Table3.11 covering the period from December 2007 to December 2021, during which the $MCCC$ index values were accessible. The results show that the news shocks from the previous month predict an increase in the abnormal temperature premium in the current month. The coefficients remain significant after controlling for the common predictors of returns. Further examination of each legs of HMC reveals that only the cold leg of the portfolio or stocks with the lowest exposure to abnormal temperatures react to climate news shocks. Therefore statistically significant and negative reaction of this leg of the portfolio reveals that HMC's returns are mainly driven by overvaluation of stocks with the lowest exposure to abnormal temperatures in the aftermath of climate news shocks.

I estimate the response function for HMC following a shock in climate news up to the last twelve months using a vector autoregressive (VAR) model. Fig. 3.10 plots the response function for HMC for stocks from industries with and without sensitivity to abnormal temperatures. As can be observed, HMC rises one month after the realization of climate news shock and falls in the following months only for the stocks from industries with sensitivity to abnormal temperatures. The results strongly suggest that the abnormal temperature premium is driven by shocks to climate concerns, and as expected, its response to a rise in climate concerns is only present when HMC is constructed using the stocks from industries with earnings sensitivity to abnormal temperatures based on the results in Table 3.3.

I also estimate the response function for HMC to climate news shocks separately for the first half of the year (quarters one and two) and the second half of the year (quarters three and four). The second half of the year is known to be warmer across the continental US , thus climate news shocks during the warmer half of the year are expected to have a more significant effect on HMC's returns due to an already elevated investor concerns about abnormal temperatures. The results in Fig. 3.11 are from the estimation of response function for the first versus the second half of the year. As can be observed, HMC only positively responds to climate news shocks during the warmer half of the year.

The results in Table 3.11 also suggest that stocks in HMC react to news shocks with a one-month delay. A similar slow reaction to climate change risks is documented by Hong, Li and Xu (2019). They show that the stocks of publicly traded food companies are slow to react to information about drought trends as one of the most important indicators of climate change risk. The delay in response to news shocks has also been documented in the literature, although only for stocks of small firms (see Lo and MacKinlay, 1990; Pástor, Stambaugh and Taylor, 2022).

Next, I estimate eq. 3.7 separately for HMC when constructed using the stocks from industry groups with and without earnings sensitivity to abnormal temperatures. The results are reported in Table 3.12 and indicate that an unexpected increase in climate concerns lead to a higher premium only for industries with sensitivity to abnormal temperatures. This is in line with the findings on the effects of exposure to abnormal temperatures on earnings and confirms that the abnormal temperature premium although driven by the climate concerns, is associated with the potential negative economic effects of abnormal temperatures.

Overall the results from this section support the hypothesis that rising climate concerns is driving the abnormal temperature premium. This explains the strong outperformance of stocks with the highest exposure to abnormal temperatures against those with the lowest after 2015.

3.6. Conclusion

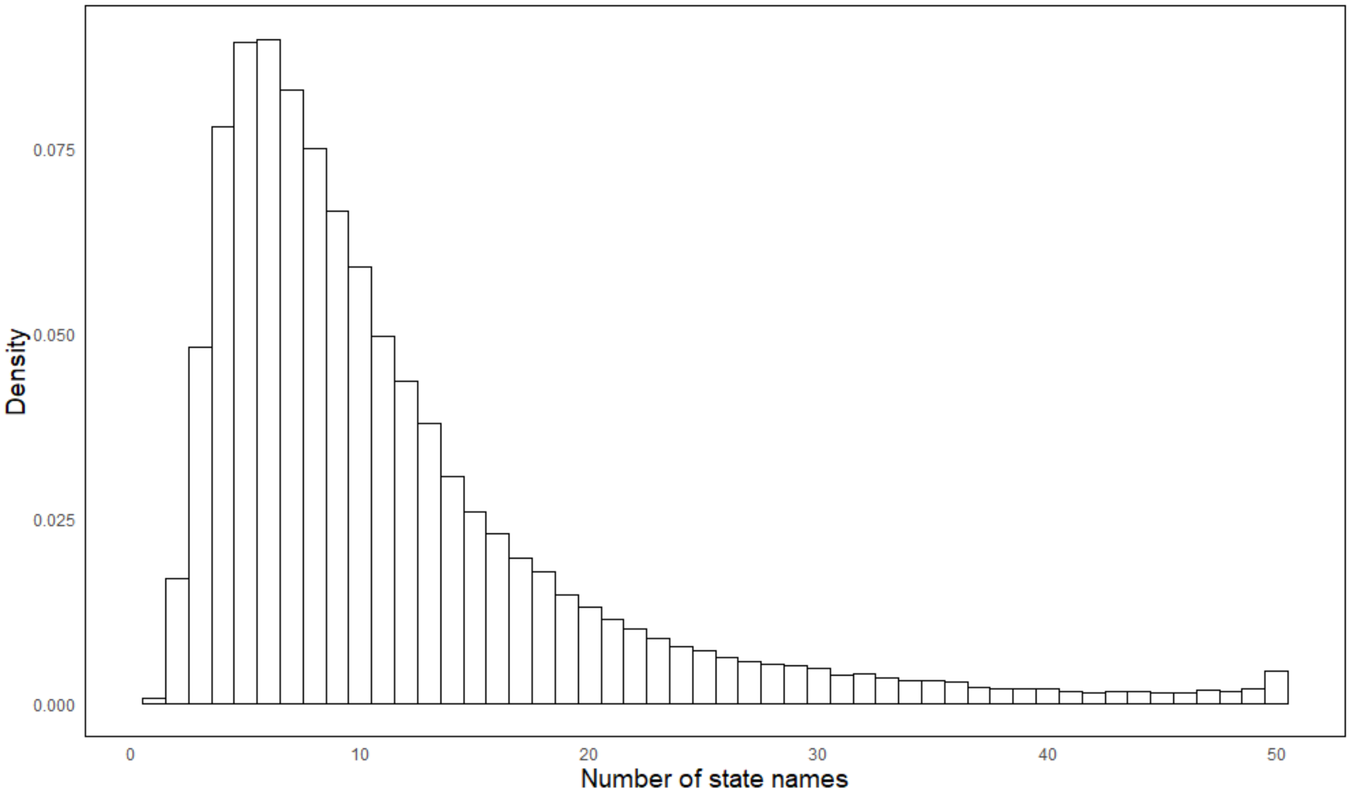
My first contribution is constructing a firm-level measure of exposure to abnormal temperatures based on the textual information on firms' presence across the US states extracted from Form 10-K filings. I show that the effect of exposure to abnormal temperatures on earnings is negative. This negative effect only exists among five industries based on the

Fama and French twelve industries classification. I show that exposure to abnormal temperatures affects earnings through revenues and operating expenses depending on the industry. The results also suggest that firm's geographic dispersion (degree of locality) amplifies the negative economic effects of exposure to abnormal temperatures.

My second contribution is showing that there exists a premium associated with exposure to abnormal temperatures. I select stocks based on the highest and lowest values of exposure to abnormal temperature and form a long-short portfolio. The portfolio's positive alpha indicate the presence of a abnormal temperature premium, which has grown significantly larger in the period after the Paris Agreement. The sudden rise in the abnormal temperature premium in recent years is comparable to that of ESG-based investment strategies and follows the documented rise in climate concerns for this period. Consistent with my earlier analysis on earnings, I find that the abnormal temperature premium is concentrated among firms that belong to the five industries with negative earnings sensitivity to abnormal temperatures. Further, I show that the abnormal temperature premium in the industries with earnings sensitivity to abnormal temperatures is concentrated among firms with low geographic dispersion (more local). The amplifying role of low geographic dispersion in the economic effects of abnormal temperatures suggests that higher geographic dispersion can serve as a diversification strategy by firms to mitigate the rising physical climate risks.

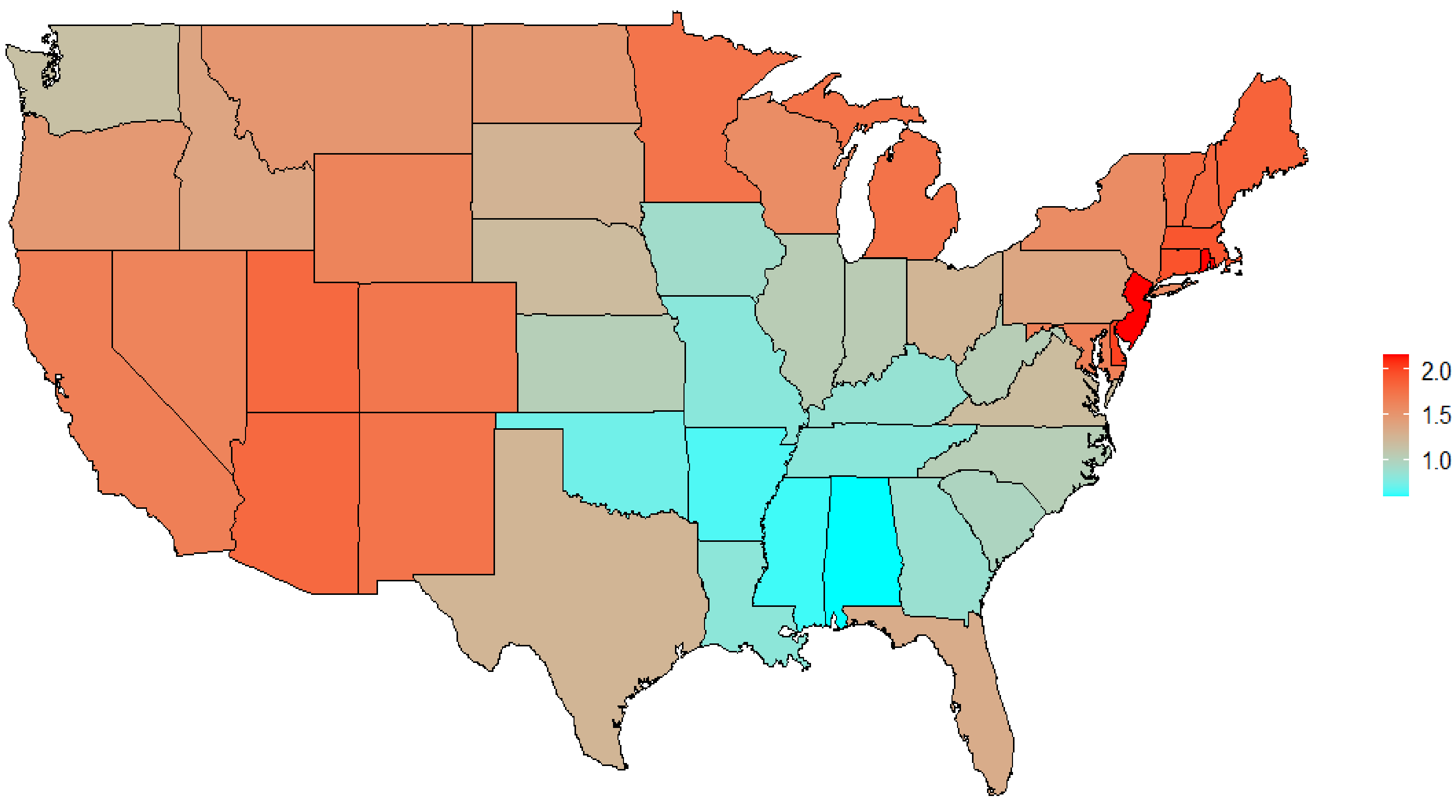
I examine the effect of an unexpected rise in climate concerns on the abnormal temperature premium and show that the premium positively responds to climate news shocks, although with a one-month delay. This supports the idea that investors demand a higher premium for exposure to abnormal temperatures after periods of elevated concerns about climate change. I conclude that the abnormal temperature premium is primarily driven by the expectation of its negative economic effects and the recent rise in climate concerns.

Figure 3.1: Geographic Dispersion



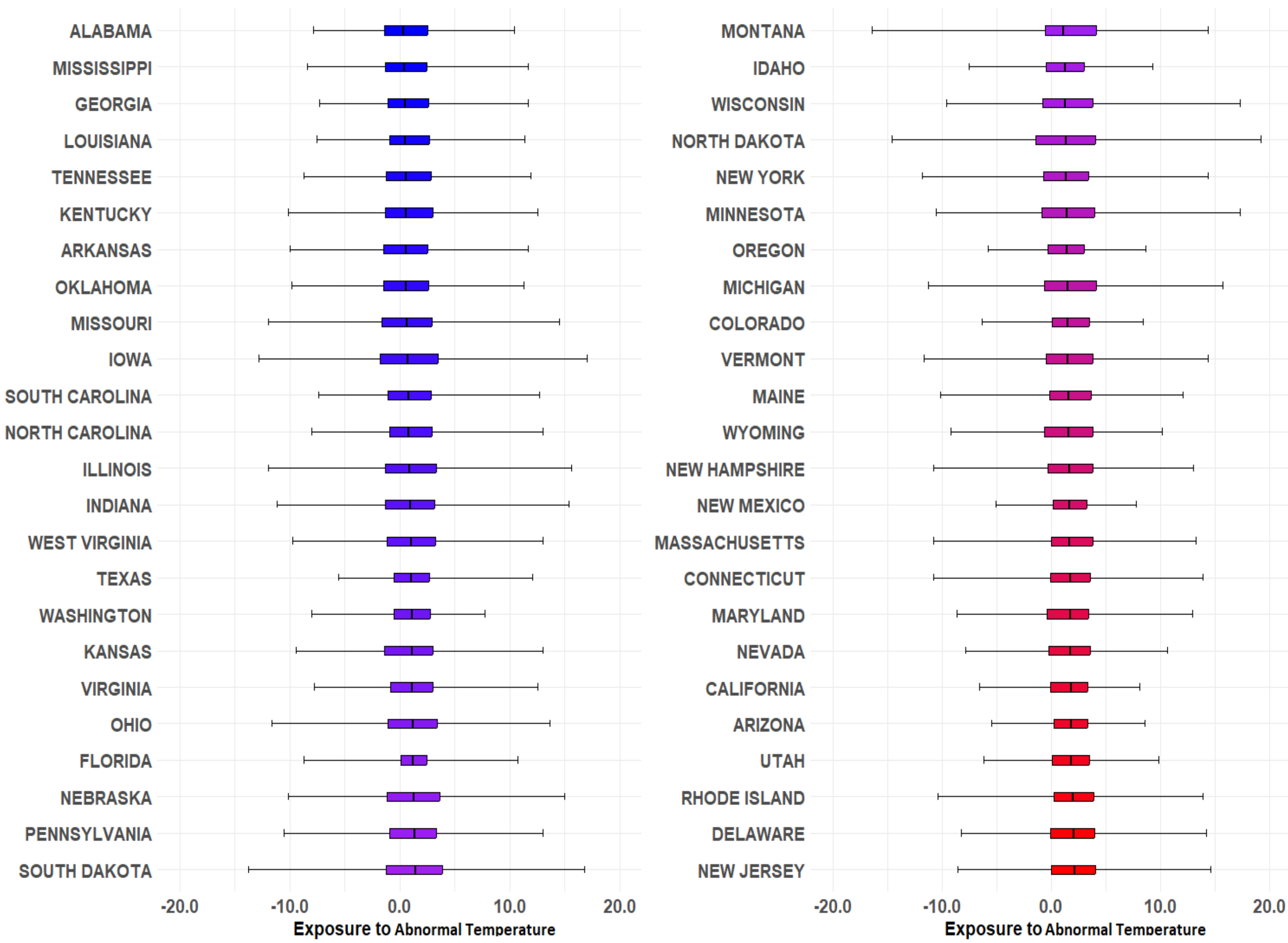
The figure above illustrates the distribution of geographic dispersion which is the number of US states mentioned in firms' 10-K filing submitted to the SEC. I removed the 10-K filings with no mentioning of any US states from the sample. The sample of 10-K filings used in this study starts from 1994 to 2021.

Figure 3.2: Average abnormal temperature across the US states



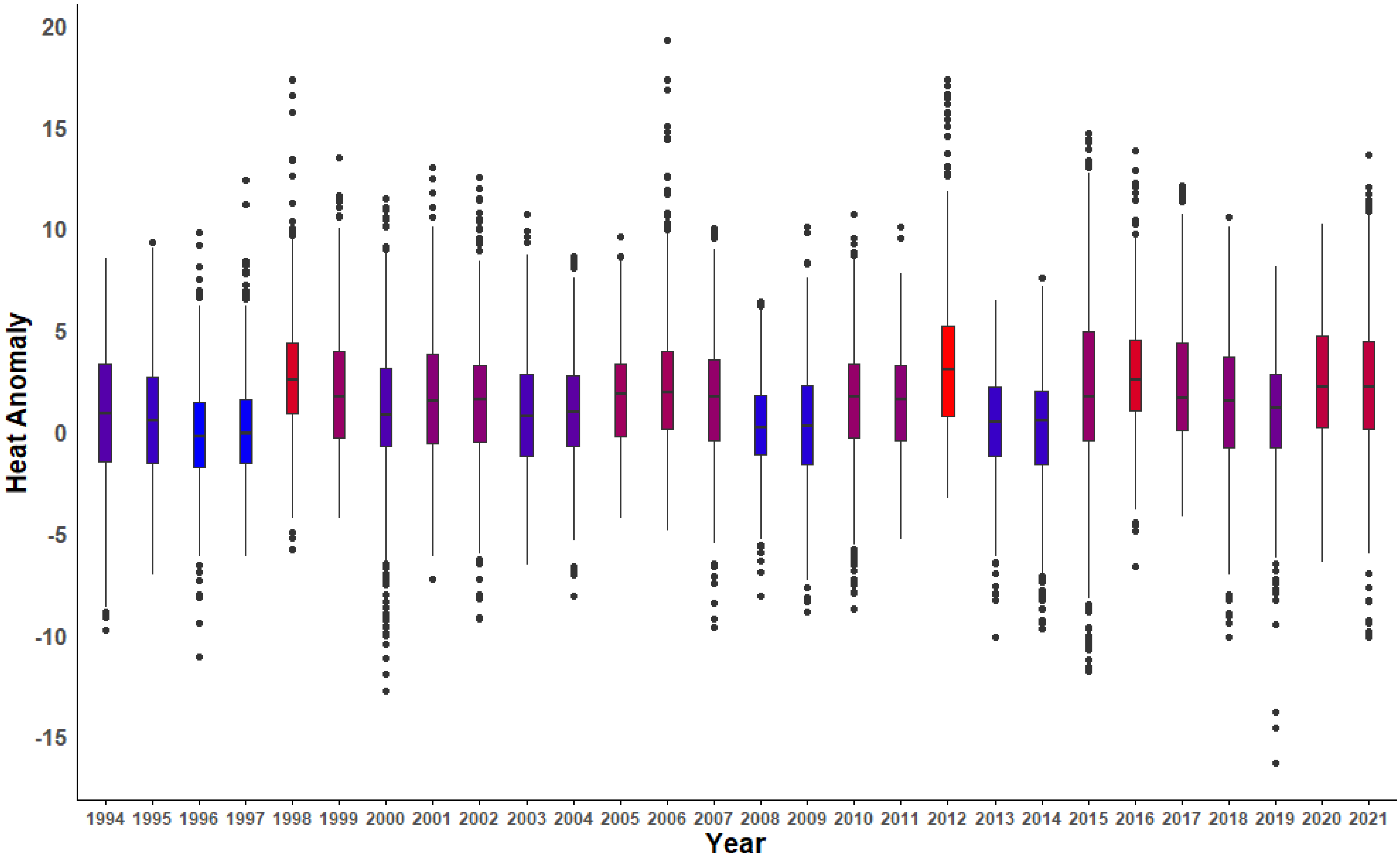
The figure above plots the US map with colors representing the average abnormal temperatures in each state between 1994 and 2021.

Figure 3.3: Distribution of abnormal temperatures by state



The figure above shows the distribution of monthly abnormal temperatures by US states. The states are ordered from the lowest (Alabama) to the highest (Rhode Island) by average abnormal temperature between 1994 and 2021. The monthly state-level temperature data is acquired from the National Centers for Environmental Information website.

Figure 3.4: Distribution of abnormal temperatures by year (across all states)



The figure above shows the distribution of monthly abnormal temperatures every year across all the US states. The sample is between 1994 and 2021. The monthly state-level temperature data is acquired from the National Centers for Environmental Information website.

Table 3.1: Quarterly earnings per share and abnormal temperatures

	Dependent variable: EPS (scaled by stock price)			
Calendar quarter:	Q1	Q2	Q3	Q4
Constant	−0.014*	0.001	−0.001	0.011
	(−1.710)	(0.141)	(−0.098)	(1.271)
abnormal temperature	**0.0004***	0.0003	**−0.001***	**−0.001**
	(2.829)	(1.609)	(−4.287)	(−2.409)
log(Size)	0.006***	0.005***	0.005***	0.006***
	(25.099)	(21.992)	(19.928)	(23.011)
Book–to–market	−0.013***	−0.015***	−0.012***	−0.018***
	(−13.013)	(−13.788)	(−11.894)	(−15.463)
Loss indicator	−0.089***	−0.095***	−0.091***	−0.110***
	(−95.225)	(−96.993)	(−92.216)	(−98.400)
Book leverage	−0.065***	−0.064***	−0.058***	−0.078***
	(−33.137)	(−33.074)	(−30.497)	(−34.966)
Dividend yield	0.937***	0.965***	0.838***	0.884***
	(13.937)	(13.836)	(11.970)	(11.527)
Dividend indicator	0.007***	0.006***	0.005***	0.003***
	(8.129)	(7.539)	(6.568)	(3.273)
Observations	112,410	110,696	106,508	106,928
Adjusted R^2	0.307	0.309	0.302	0.328

In this table, I report the results from pooling panel regression based on equation 3.3:

$$EPS_{i,t} = \boldsymbol{\beta_0} + \boldsymbol{\beta_{temp}} Abnormal\,Temperature_{i,t} + \boldsymbol{\beta} X_{i,t-1} + \delta_{ind} + \theta_{year/quarter} + \epsilon_{i,t}$$

where $EPS_{i,t}$ is the quarterly earnings per share scaled by the beginning of the fiscal quarter stock price. $Abnormal\,Temperature_{i,t}$ is the monthly average of firm-level abnormal temperatures for the three months that cover the target fiscal period t. $X_{i,t-1}$ are the control variables based on Fama and French (2000) that include the firm-level book-to-market, size, indicators of loss and dividend, book value of leverage, and dividend yield for firm i at the beginning of the fiscal period. t-statistics are reported in parenthesis. The error terms are clustered at the year-quarter and firm level, and year-quarter and industry fixed effects (Fama and French 48 industries) are accounted for. All the variables are winzorized at 1% from above and below. The sample is annual and from 1994 to 2021.

Table 3.2: Annual earnings per share and abnormal temperatures

	Dependent variable: EPS (scaled by stock price)			
	(1)	(2)	(3)	(4)
Constant	-0.040^{*}	-0.599^{***}	-0.614^{***}	-0.141^{***}
	(-1.925)	(-28.876)	(-29.184)	(-7.393)
abnormal temperature	$-\mathbf{0.007}^{***}$	$-\mathbf{0.017}^{***}$	$-\mathbf{0.017}^{***}$	$-\mathbf{0.007}^{***}$
	(-5.670)	(-14.795)	(-14.775)	(-7.038)
log(Size)		0.047^{***}	0.048^{***}	0.023^{***}
		(99.832)	(94.399)	(45.413)
Book–to–market			0.007^{***}	-0.012^{***}
			(4.297)	(-8.466)
Loss indicator				-0.306^{***}
				(-141.036)
Book leverage				-0.203^{***}
				(-58.336)
Dividend yield				0.619^{***}
				(12.320)
Dividend indicator				0.015^{***}
				(5.496)
Observations	98,645	98,643	98,643	98,579
Adjusted R^2	0.036	0.124	0.124	0.318

In this table, I report the results from pooling panel regression based on equation 3.3:

$$EPS_{i,t} = \boldsymbol{\beta_0} + \boldsymbol{\beta_{temp}} Abnormal\,Temperature_{i,t} + \boldsymbol{\beta} X_{i,t-1} + \delta_{ind} + \theta_{year} + \epsilon_{i,t}$$

where $EPS_{i,t}$ is the annual earnings per share scaled by the beginning of the fiscal year stock price. $Abnormal\,Temperature_{i,t}$ is the monthly average of firm-level abnormal temperatures for the twelve months that cover the target fiscal period t. $X_{i,t-1}$ are the control variables based on Fama and French (2000) that include the firm-level book-to-market, size, indicators of loss and dividend, book value of leverage, and dividend yield for firm i at the beginning of the fiscal period. t-statistics are reported in parenthesis. The error terms are clustered at the year and firm level, and year and industry fixed effect (Fama and French 48 industries) is accounted for. All the variables are winzorized at 1% from above and below. The sample is annual and from 1994 to 2021.

Table 3.3: Economic effect of abnormal temperatures by Fama and French 12 industries classification

	Dependent variables (Annual):				
Industry	EPS	Revenue	Operating Expense	Cost of Goods Sold	Selling, General Admin. Expenses
NoDur	0 (0)	-0.185 (-4.845)	-0.166 (-4.475)	-0.158 (-5.056)	-0.009 (-0.976)
Dur	-0.002 (-0.476)	-0.124 (-2.683)	-0.114 (-2.578)	-0.105 (-2.908)	-0.009 (-0.898)
Manuf	-0.003 (-1.079)	-0.088 (-4.192)	-0.07 (-3.369)	-0.078 (-4.511)	0.004 (0.802)
Energy	**-0.037***** (-5.365)	0.011 (0.265)	0.08 (1.808)	0.056 (1.488)	**0.028***** (3.227)
Chems	-0.008 (-1.294)	-0.116 (-2.8)	-0.103 (-2.401)	-0.079 (-2.376)	-0.014 (-1.191)
Buseq	**-0.01***** (-4.026)	-0.016 (-1.153)	0.007 (0.463)	-0.014 (-1.207)	**0.015***** (3.379)
Telcm	-0.005 (-0.547)	0.013 (0.265)	0.023 (0.466)	0.013 (0.367)	0.011 (0.598)
Utils	**-0.005**** (-2.088)	**-0.124***** (-5.658)	**-0.112***** (-5.846)	-0.11 (-5.905)	-0.004 (-0.813)
Shops	0.003 (0.872)	-0.179 (-4.448)	-0.168 (-4.202)	-0.141 (-4.301)	-0.021 (-2.426)
Hlth	**-0.026***** (-6.704)	-0.056 (-3.583)	0.022 (1.139)	**-0.029**** (-2.253)	**0.031***** (3.948)
Finance	0.003 (1.473)	-0.107 (-6.486)	-0.107 (-6.703)	-0.099 (-7.002)	-0.018 (-3.688)
Other	**-0.006**** (-2.022)	**-0.078**** (-2.854)	**-0.062**** (-2.296)	**-0.069***** (-3.097)	0.006 (0.915)

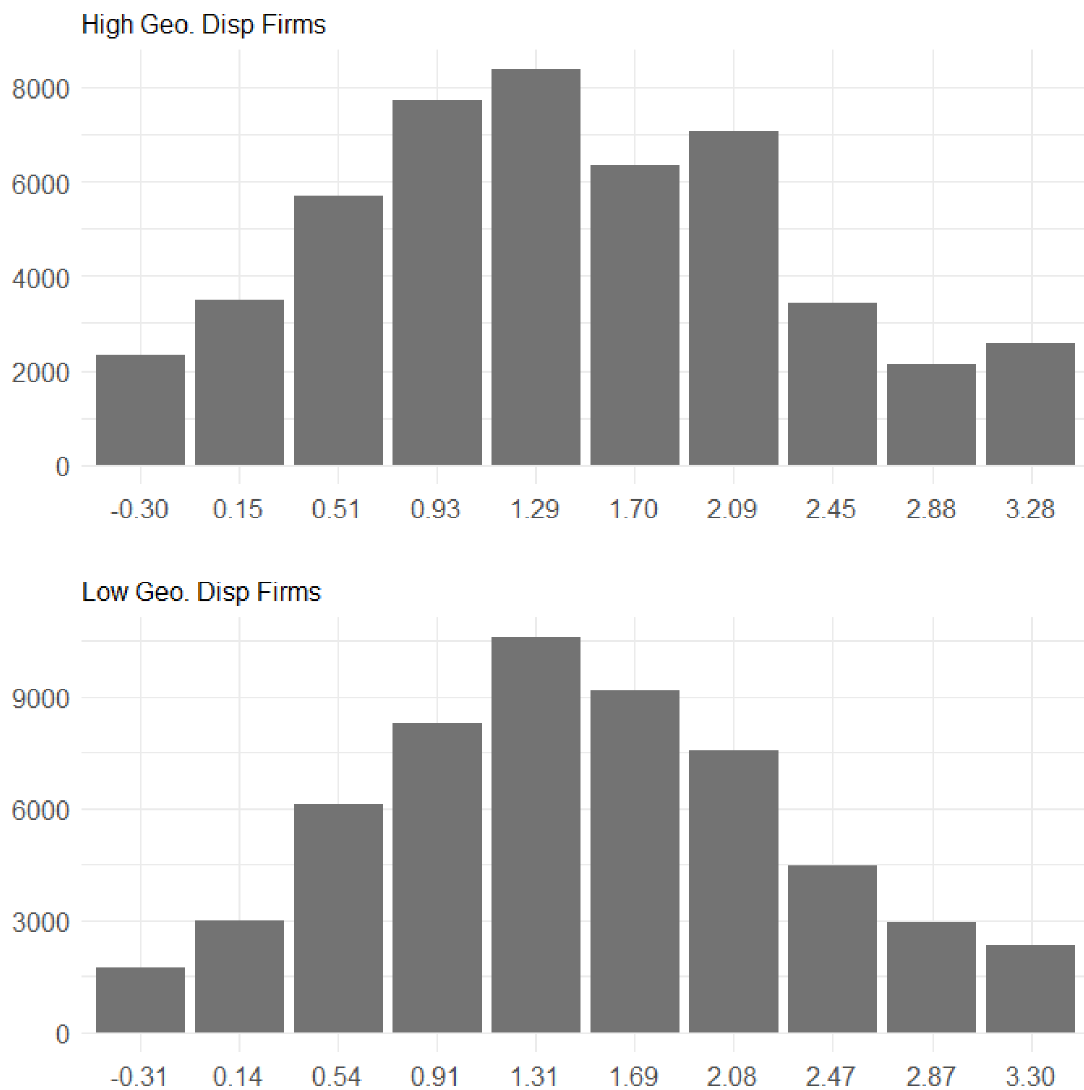
In this table, I report the pooling panel regression results based on equation 3.3, separately estimated for the twelve industries based on Fama and French twelve industries classification, and using a set of dependent variables, including annual EPS, revenue, operating cost, cost of goods sold, and selling general and administrative costs. All dependent variables are values per share and scaled by the stock price at the beginning of the fiscal year. I only report the coefficient for $Abnormal\,Temperature_{i,t}$ for expositional purposes. t-statistics are reported in parenthesis. Control variables are included in the regressions.

Table 3.4: Annual earnings per share and abnormal temperatures by industry group and geographic dispersion

	Industries sensitive to abnormal temperatures		*Industries not sensitive to abnormal temperatures*	
	Low Geo. Disp	High Geo. Disp	Low Geo. Disp	High Geo. Disp
abnormal temperature	**−0.020*****	**−0.014*****	−0.002	0.001
	(−8.364)	(−5.599)	(−1.154)	(0.470)
log(Size)	0.032***	0.029***	0.014***	0.018***
	(19.044)	(14.827)	(9.084)	(11.243)
Book–to–market	−0.025***	−0.0003	−0.014**	−0.001
	(−3.507)	(−0.051)	(−2.363)	(−0.104)
Loss indicator	−0.277***	−0.292***	−0.346***	−0.331***
	(−58.925)	(−47.668)	(−47.777)	(−47.720)
Book leverage	−0.196***	−0.245***	−0.178***	−0.177***
	(−17.839)	(−14.951)	(−13.022)	(−13.449)
Dividend yield	1.046***	0.633***	0.410***	0.430***
	(7.762)	(5.465)	(4.982)	(4.948)
Dividend indicator	0.039***	0.020***	−0.002	0.0004
	(6.043)	(3.578)	(−0.322)	(0.098)
Observations	28,447	20,552	21,154	24,202
Adjusted R^2	0.276	0.318	0.338	0.348

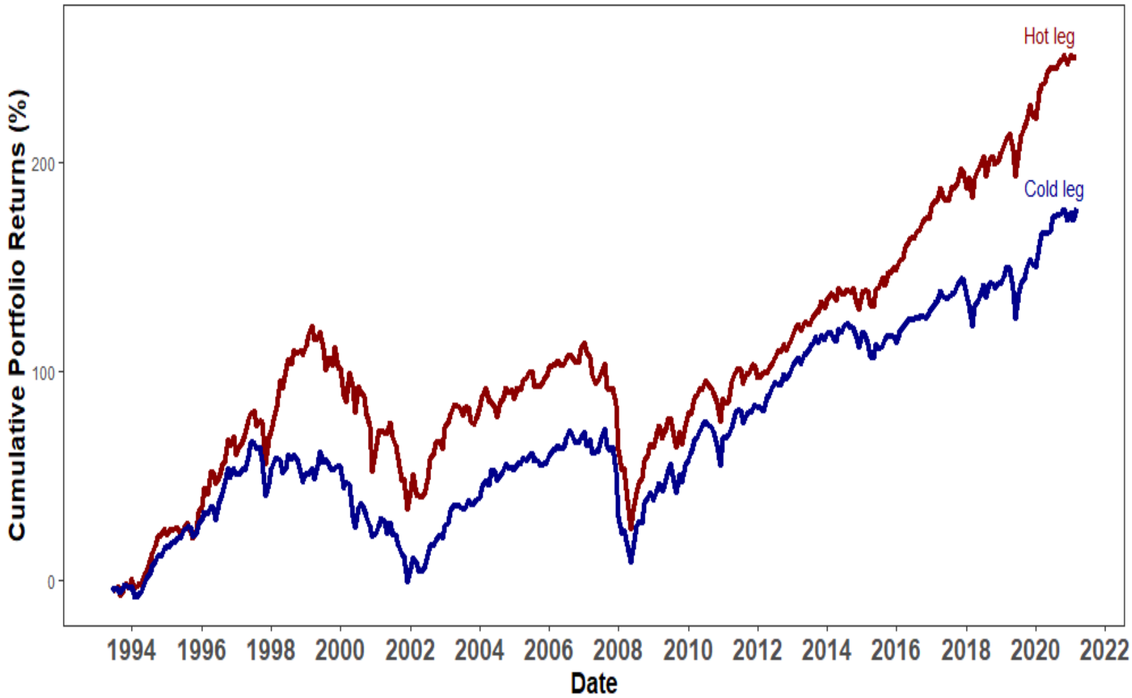
In this table, I report the pooling panel regression results based on eq.3.3 for firms with high and low geographic dispersion (Number of states) belonging to industry groups with and without sensitivity to abnormal temperatures based on the results in Table 3.3. I select the median value of geographic dispersion (9 states) as the threshold to separate low and high geographic dispersion firms in the sample. t-statistics are reported in parenthesis. The error terms are clustered at the year and firm level, and year and industry fixed effect (Fama and French 48 industries) is accounted for. All variables are winzorized at 1% from above and below. The sample is from 1994 to 2021.

Figure 3.5: Distribution of abnormal temperatures by geographic dispersion



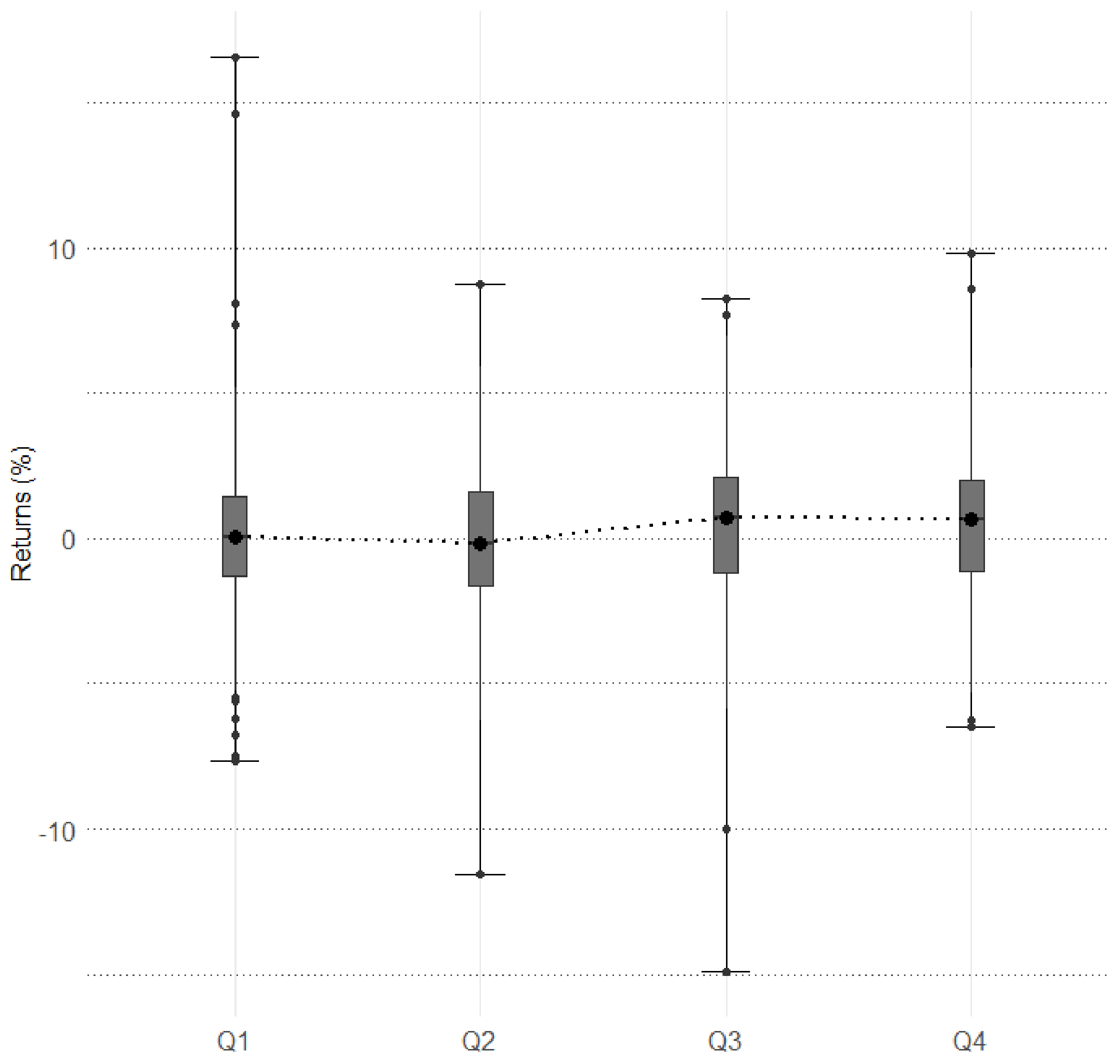
The figure above shows the distribution of the variable $Abnormal\,Temperature_{i,t}$ from eq.3.3 for high and low geographic dispersion firm groups.

Figure 3.6: Cumulative returns for hot and cold legs of HMC



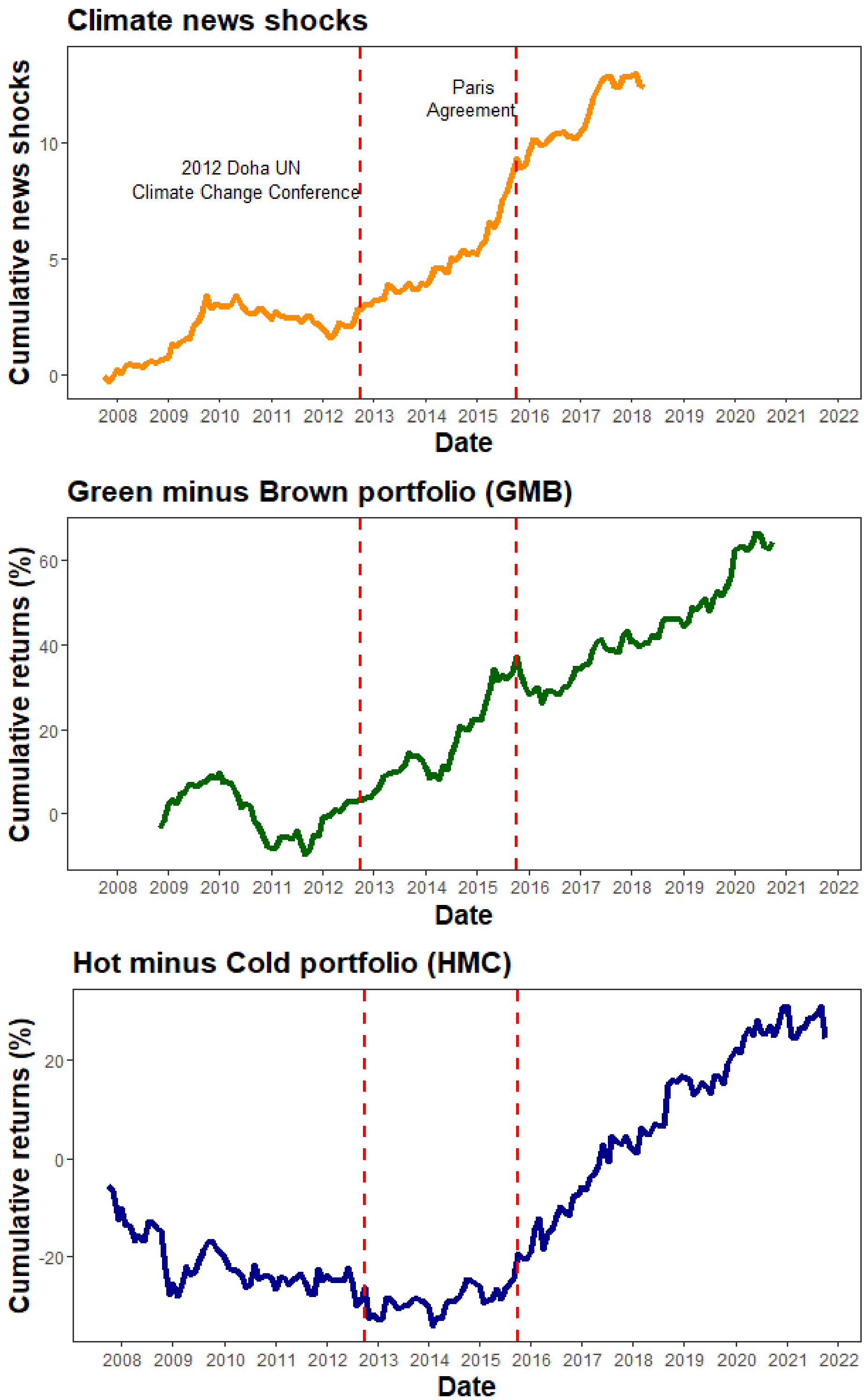
The figure above is the cumulative monthly excess returns (over risk-free rate) of the value-weighted portfolios that are re-balanced every month and take only long positions in the stocks of firms in the top and bottom 20th percentile of expected abnormal temperatures. Hot is the portfolio that holds the stocks with the highest expected abnormal temperature, and Cold is the portfolio of stocks with the lowest expected abnormal temperature.

Figure 3.7: Distribution of HMC's returns by quarter



The figure above represents the distribution of monthly returns for HMC within each quarter. The sample is from 1994 to 2021.

Figure 3.8: Cumulative climate news shocks and portfolio returns



The figure above includes the cumulative climate-concern shocks using the Media Climate Change Concerns Index ($MCCC$) from Ardia et al. (2022). Monthly news shocks are prediction errors from rolling AR(1) models fitted to the monthly $MCCC$ index. Green minus brown (GMB) portfolio is cumulative monthly returns of the portfolios sorted based on stocks greenness' measure from Pástor, Stambaugh and Taylor (2022). Hot minus cold (HMC) is the cumulative monthly returns of the value-weighted portfolios sorted based on stocks' expected abnormal temperatures.

Table 3.5: Performance statistics for the portfolio sorted based on expected abnormal temperatures or Hot minus Cold (HMC)

		Hot minus. Cold (HMC)	Hot leg.	Cold leg.
Full Sample (1994-2021)	$\alpha(\%)$	0.186 (1.127)	0.098 (0.765)	-0.088 (-1.148)
	Avg. Ret (%)	0.26 (1.62)	**0.9** (2.9)	**0.63** (2.61)
	σ	3.31	5.41	4.42
	Sharpe	0.0785	0.166	0.142
Before 2015	$\alpha(\%)$	0.117 (0.545)	0.067 (-0.522)	-0.050 (-0.555)
	Avg. Ret (%)	0.15 (0.73)	0.72 (1.85)	0.57 (1.88)
	σ	3.54	5.66	4.43
	Sharpe	0.04	0.13	0.13
After 2015	$\alpha(\%)$	**0.501***** (2.657)	**0.244**** (2.163)	$\mathbf{-0.257}^{**}$ (-2.190)
	Avg. Ret (%)	**0.62** (3.67)	**1.46** (4.5)	**0.84** (2.47)
	σ	2.51	4.55	4.42
	Sharpe	0.25	0.32	0.19

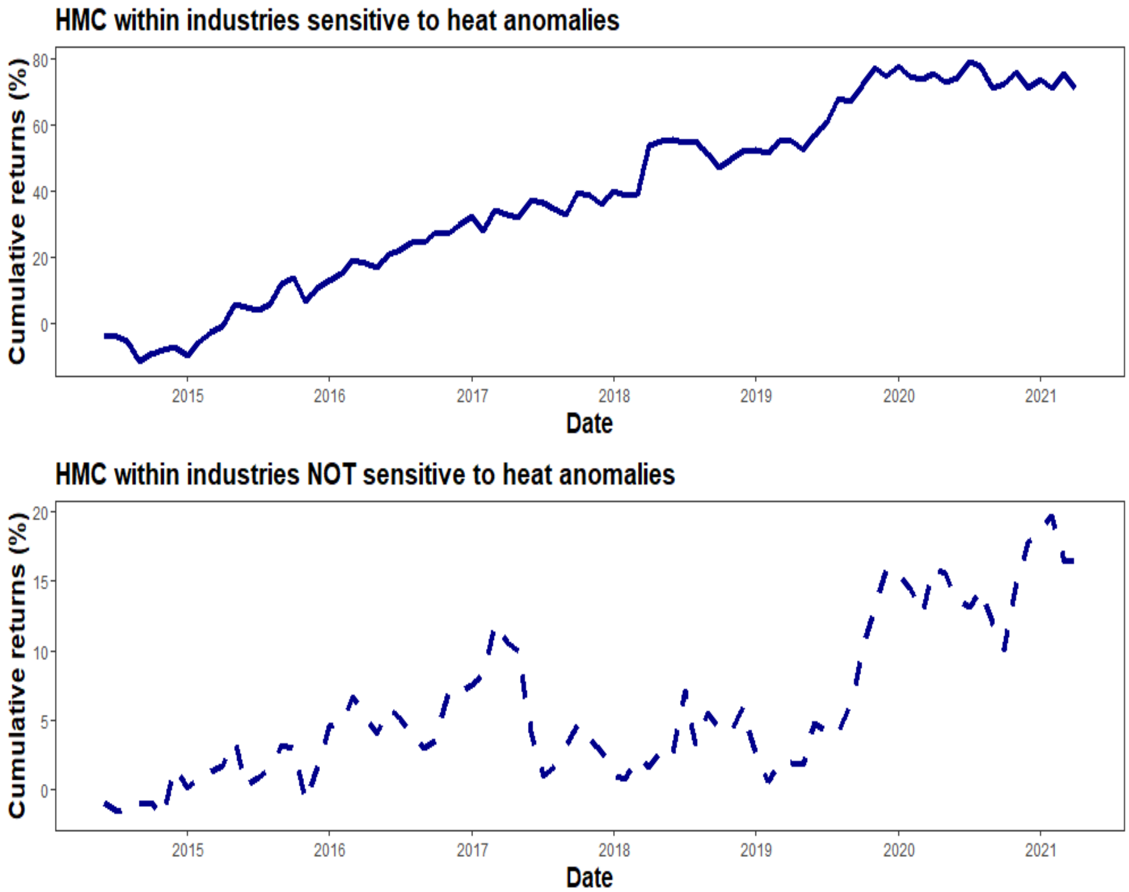
This table summarizes the monthly returns, abnormal returns (α), and risk associated with the portfolio sorted based on expected abnormal temperature (HMC) and its long (Hot leg) and short legs (Cold leg). HMC is the value-weighted portfolio re-balanced every month based on normalized expected abnormal temperatures (normalized across firms every month) calculated using the latest 10-K filing in the previous twelve months. The portfolio takes a long position in stocks of firms in the top 20% of expected abnormal temperatures and takes a short position in the bottom 20%. Hot leg is the excess returns (over risk-free rate) of the long leg of HMC, and Cold leg is the excess returns of the short leg of HMC. Alphas are estimated by regressing the portfolio returns on Fama and French (2015) three factors plus momentum factor from Carhart (1997). The values in parentheses are Newey-West T-statistics that test the difference in the means from zero. The returns are in percent per month.

Table 3.6: HMC's performance after 2015

	Dependent variable: HMC				
	(1)	(2)	(3)	(4)	(5)
$\alpha(\%)$	**0.585*****	**0.501*****	**0.627*****	**0.627*****	**0.573****
	(3.453)	(2.657)	(2.835)	(2.824)	(2.220)
Mkt-RF	0.025	0.046	0.030	0.031	0.025
	(0.763)	(0.848)	(0.730)	(0.743)	(0.561)
SMB		0.046	−0.079	−0.075	−0.083
		(0.467)	(−0.618)	(−0.540)	(−0.650)
HML		−0.144***	−0.002	−0.003	0.022
		(−2.857)	(−0.041)	(−0.053)	(0.365)
RMW			−0.284*	−0.285*	−0.265
			(−1.824)	(−1.853)	(−1.570)
CMA			−0.342**	−0.347*	−0.265
			(−2.011)	(−1.972)	(−1.482)
UMD		0.035	0.002	0.002	0.011
		(0.376)	(0.020)	(0.024)	(0.125)
LIQ				−0.734	−1.068
				(−0.112)	(−0.150)
GEO					−0.108
					(−0.524)
Observations	84	84	84	84	84
Adjusted R^2	−0.010	−0.002	0.065	0.052	0.050

This table reports risk-adjusted returns (alphas) under different specifications, focusing on the period between January 2015 to December 2021. In each regression, the dependent variable is the returns of HMC, which is the value-weighted portfolio re-balanced every month based on the normalized expected abnormal temperatures (normalized across firms every month) calculated using the latest 10-K filing in the previous twelve months. The portfolio takes a long position in stock of firms exposed to abnormal temperature that is in the top 20% percentile and goes short in stock of firms in the bottom 20% percentile. Columns (1) through (3) includes Fama and French (2015) three factors and momentum factor from Carhart (1997). Columns (4) and (5) include the liquidity factor LIQ from Pástor and Stambaugh (2003) and premium for geographic dispersion GEO from Garcia and Norli (2012), respectively. I include Newey-West T-statistics. The returns are percentages per month.

Figure 3.9: Cumulative HMC after 2015 by industry groups with and without sensitivity to abnormal temperatures



The figures above are the cumulative monthly returns of HMC after 2015. The top and bottom figures represent HMC that is constructed using only the stocks belonging to industry groups with and without earning sensitivity to abnormal temperatures according to the results in Table 3.3.

Table 3.7: HMC after 2015 by industry groups with and without sensitivity to abnormal temperatures

	HMC within industries sensitive to abnormal temperatures			*HMC within industries Not sensitive to abnormal temperatures*		
	(1)	(2)	(3)	(1)	(2)	(3)
$\alpha(\%)$	**0.743****	**0.924*****	**0.869****	0.116	0.163	0.110
	(2.378)	(3.151)	(2.448)	(1.253)	(1.574)	(1.170)
Mkt-RF	0.061	0.087	0.074	0.049	0.054	0.040
	(0.546)	(1.084)	(0.766)	(0.655)	(0.628)	(0.554)
SMB	0.013	−0.242	−0.276	0.022	−0.043	−0.083
	(0.088)	(−1.375)	(−1.643)	(0.343)	(−0.533)	(−0.746)
HML	−0.238*	−0.083	−0.051	−0.133**	−0.090	−0.058
	(−1.816)	(−0.525)	(−0.321)	(−2.051)	(−0.898)	(−0.671)
RMW		−0.630***	−0.605***		−0.160*	−0.134
		(−3.079)	(−2.675)		(−1.781)	(−1.615)
CMA		−0.224	−0.097		−0.070	0.066
		(−1.196)	(−0.447)		(−0.419)	(0.301)
UMD	0.026	−0.024	−0.018	−0.083	−0.097	−0.091
	(0.175)	(−0.173)	(−0.123)	(−1.048)	(−1.211)	(−1.184)
LIQ			4.857			6.277
			(0.438)			(0.787)
GEO			−0.119			−0.116
			(−0.374)			(−1.350)
Observations	84	84	84	84	84	84
Adjusted R^2	0.007	0.082	0.065	0.012	0.007	0.008

This table reports alphas under different specifications, focusing on the period between January 2015 to December 2021. In the first three columns, the dependent variable is the returns of HMC constructed using the stocks of industries with earnings sensitivity to abnormal temperatures based on the results from Table 3.3. In the last three columns, the dependent variable is the returns of HMC constructed using the stocks of industries with earnings that is not impacted by abnormal temperatures. I include Newey-West T-statistics. The returns are percentages per month.

Table 3.8: HMC after 2015 by geographic dispersion

	HMC within industries with sensitivity to abnormal temperatures					
	Low Geo. Dispersion			*High Geo. Dispersion*		
	(1)	(2)	(3)	(1)	(2)	(3)
$\alpha(\%)$	**1.038*****	**1.101*****	**1.132*****	−0.392	−0.275	−0.281
	(2.685)	(2.908)	(2.880)	(−1.018)	(−0.776)	(−0.692)
Mkt-RF	−0.100	−0.060	−0.043	−0.023	−0.048	−0.031
	(−0.978)	(−0.686)	(−0.515)	(−0.192)	(−0.471)	(−0.278)
SMB	0.024	−0.125	−0.083	0.369**	0.218	0.283*
	(0.189)	(−0.884)	(−0.438)	(2.017)	(1.562)	(1.988)
HML	0.068	0.119	0.097	−0.164**	0.036	0.024
	(0.314)	(0.547)	(0.446)	(−2.339)	(0.233)	(0.141)
RMW		−0.383**	−0.406**		−0.392	−0.408
		(−2.360)	(−2.398)		(−1.199)	(−1.129)
CMA		0.026	−0.092		−0.517**	−0.627***
		(0.139)	(−0.386)		(−2.566)	(−3.457)
UMD	0.064	0.040	0.040	0.243*	0.207	0.214
	(0.341)	(0.209)	(0.212)	(1.865)	(1.443)	(1.311)
LIQ			−8.059			−13.536
			(−0.532)			(−1.107)
GEO			0.073			0.010
			(0.346)			(0.047)
Observations	81	81	81	79	79	79
Adjusted R^2	−0.033	−0.038	−0.061	0.059	0.095	0.079

This table reports alpha for HMC portfolio constructed for low and high geographic dispersion firms belonging to the industries with earnings sensitivity to abnormal temperatures based on Table 3.4. In the first three columns, the dependent variable is the returns of HMC constructed using the stocks of firms with low geographic dispersion. In the last three columns, the dependent variable is the returns of HMC constructed using the stocks of firms with high geographic dispersion. I select the median value of geographic dispersion in the sample (9 states) as the threshold to separate low and high geographic dispersion firms. I include Newey-West T-statistics in parentheses. The sample is monthly and is from January 2015 to December 2021. The returns are in percent per month.

Table 3.9: HMC after 2015 by the technology industry and other industries

	HMC within industries with sensitivity to abnormal temperatures			
	Technology industry		*Other industries*	
	Low Geo. Disp	High Geo. Disp	Low Geo. Disp	High Geo. Disp
$\alpha(\%)$	**0.973*****	-0.449	**1.153*****	-0.480
	(2.773)	(-0.524)	(3.462)	(-1.350)
Mkt-RF	-0.263^{**}	-0.194	-0.180^{*}	-0.085
	(-2.037)	(-1.065)	(-1.803)	(-0.857)
SMB	0.062	0.330	-0.132	0.289^{**}
	(0.260)	(1.456)	(-0.600)	(2.392)
HML	0.081	0.383^{**}	-0.204	-0.355^{***}
	(0.316)	(2.056)	(-1.475)	(-2.830)
RMW	-0.436^{**}	-0.014	-0.148	-0.043
	(-2.025)	(-0.035)	(-0.522)	(-0.318)
CMA	-0.090	0.082	-0.693^{***}	-0.676^{***}
	(-0.242)	(0.243)	(-3.800)	(-2.999)
UMD	-0.309	0.385^{**}	0.176	-0.111
	(-1.418)	(2.568)	(0.972)	(-1.088)
LIQ	-17.969	-7.447	0.597	-3.652
	(-1.018)	(-0.317)	(0.052)	(-0.356)
GEO	0.043	-0.512^{*}	0.328^{**}	0.331^{**}
	(0.152)	(-1.815)	(2.370)	(2.010)
Observations	78	75	77	75
R^2	0.121	0.148	0.334	0.186
Adjusted R^2	0.019	0.045	0.255	0.087

This table reports alpha for HMC portfolio constructed for low and high geographic dispersion firms belonging to the technology industry and other industries with earnings sensitivity to abnormal temperatures based on Table 3.3. In the first two columns, the dependent variable is the returns of HMC constructed using the stocks of firms with low and high geographic dispersion within the technology industry or business equipment based on the Fama and French twelve industry classification. In the last two columns, the dependent variable is the returns of HMC constructed using the stocks of firms with low and high geographic dispersion within other industries with earnings sensitivity to abnormal temperatures based on Table 3.3. I include Newey-West t-statistics in parentheses. The sample is monthly and is from January 2015 to December 2021. The returns are in percent per month.

Table 3.10: Annual earnings per share and abnormal temperatures for different periods

	EPS from industries with sensitivity to abnormal temperatures		
	Full Sample 1994-2021	Before 2015	After 2015
abnormal temperature	**−0.015*****	**−0.004*****	−0.009*
	(−8.994)	(−2.823)	(−1.842)
log(Size)	0.030***	0.023***	0.058***
	(23.850)	(18.125)	(19.380)
Book–to–market	−0.016***	−0.016***	−0.012
	(−3.209)	(−3.270)	(−1.029)
loss indicator	−0.282***	−0.285***	−0.241***
	(−73.374)	(−71.030)	(−30.707)
Book leverage	−0.215***	−0.206***	−0.237***
	(−23.554)	(−20.235)	(−13.126)
Dividend yield	0.800***	0.617***	1.581***
	(9.078)	(6.765)	(8.074)
Dividend indicator	0.033***	0.022***	0.071***
	(7.374)	(5.007)	(7.047)
Observations	51,450	39,262	12,188
Adjusted R^2	0.292	0.302	0.302

In this table, I report the results from pooling panel regression based on equation 3.3 for the subperiods before and after 2015 and within industries with sensitivity to abnormal temperatures based on the results in Table3.3:

$$EPS_{i,t} = \boldsymbol{\beta_0} + \boldsymbol{\beta_{temp}} Abnormal\, Temperature_{i,t} + \boldsymbol{\beta} X_{i,t-1} + \delta_{ind} + \theta_{year} + \epsilon_{i,t}$$

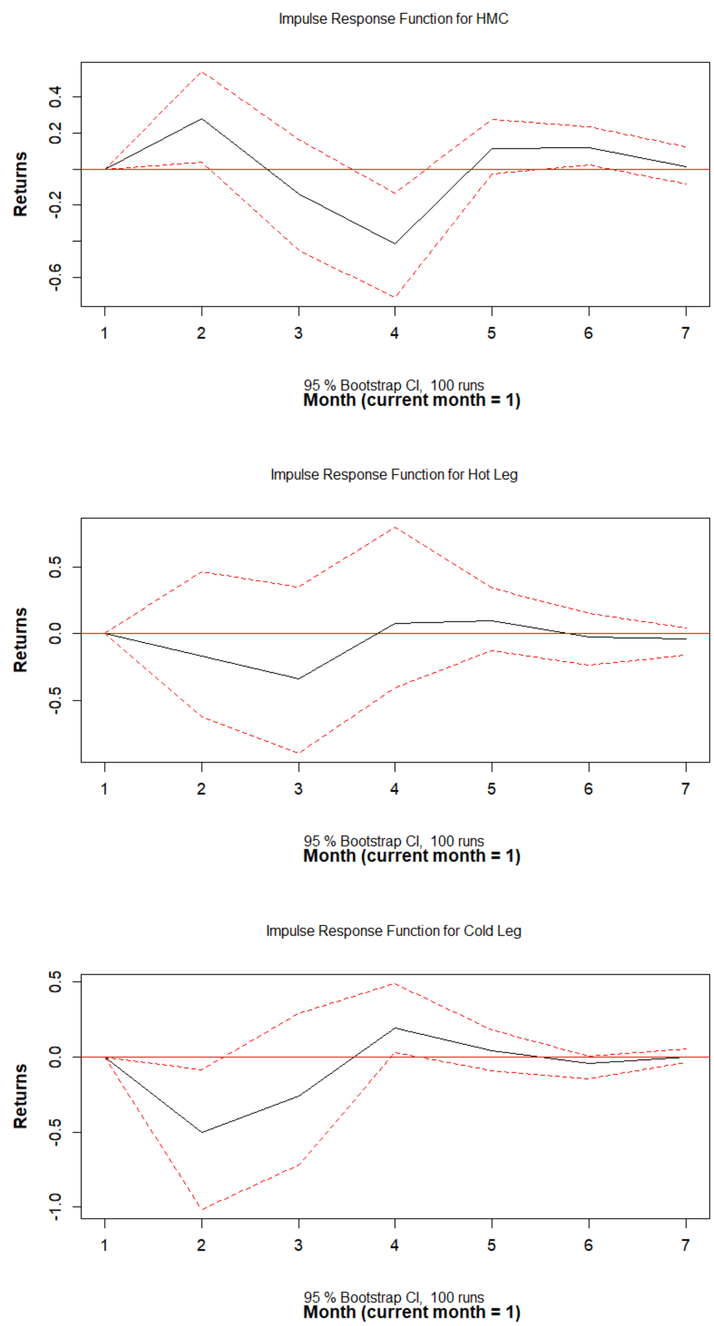
where $EPS_{i,t}$ is the annual earnings per share scaled by the beginning of the fiscal year stock price. $Abnormal\, Temperature_{i,t}$ is the monthly average of firm-level abnormal temperatures for the twelve months that cover the target fiscal period t. $X_{i,t-1}$ are the control variables based on Fama and French (2000) that include the firm-level book-to-market, size, indicators of loss and dividend, book value of leverage, and dividend yield for firm i at the beginning of the fiscal period. t-statistics are reported in parenthesis. The error terms are clustered at the year and firm level, and year and industry fixed effect (Fama and French 48 industries) is accounted for. All the variables are winzorized at 1% from above and below. The sample is annual and from 1994 to 2021.

Table 3.11: News shocks and HMC

	HMC (1)	HMC (2)	HMC (3)
Constant	0.174	0.194	0.176
	(1.148)	(1.357)	(1.248)
News Shock	−0.169	−0.043	−0.101
	(−0.453)	(−0.109)	(−0.306)
Lag News Shock		**1.072*****	**1.038*****
		(3.000)	(2.649)
Observations	227	226	226
Adjusted R^2	−0.004	0.021	0.145
	Hot Leg (1)	*Hot Leg* (2)	*Hot Leg* (3)
Constant	1.049***	1.057***	0.012
	(3.054)	(3.114)	(0.103)
News Shock	0.453	0.453	−0.074
	(0.558)	(0.550)	(−0.338)
Lag News Shock		0.070	0.376
		(0.114)	(1.626)
Observations	227	226	226
Adjusted R^2	−0.003	−0.008	0.905
	Cold Leg (1)	*Cold Leg* (2)	*Cold Leg* (3)
Constant	0.875***	0.863***	−0.164*
	(3.147)	(3.106)	(−1.934)
News Shock	0.622	0.495	0.027
	(1.012)	(0.803)	(0.142)
Lag News Shock		−1.003	−**0.662*****
		(−1.582)	(−2.989)
Observations	227	226	226
Adjusted R^2	−0.001	0.003	0.909

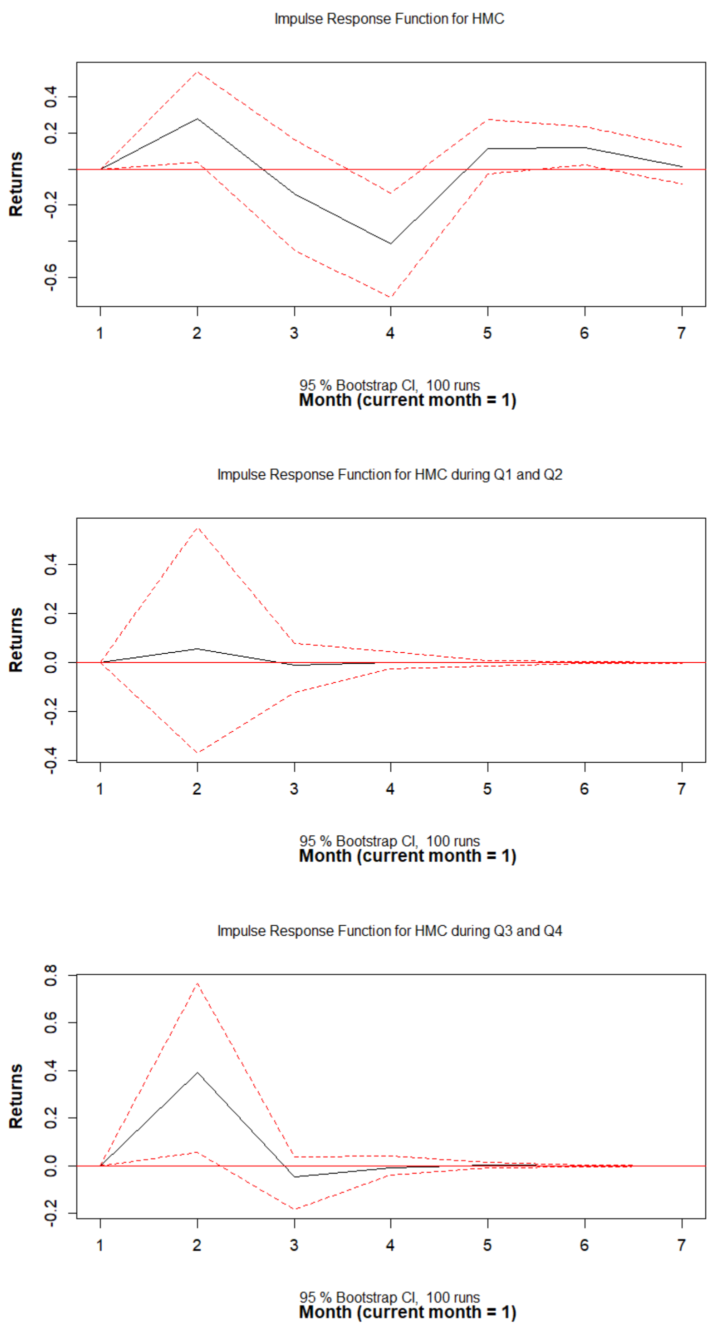
This table reports the time-series regressions of HMC and its legs against the extreme temperature category of climate news shocks. Climate news shock is the prediction error from the AR(1) model applied to the $MCCC$ index from Ardia et al. (2022). News shock is only based on the extreme temperature category of climate news shock series. The lag of news shocks is the shock from the previous month. Column (3) in both tables contains all the control variables from Table 3.6 but are not reported for expositional limitations. I include Newey-West t-statistics in parentheses. The sample is monthly from December 2007, when the $MCCC$ index data is available, to December 2021. The returns are in percentages.

Figure 3.10: HMC Response Function to Shocks in Extreme Temperature Category of Climate News



The plots above show the response function of HMC to the extreme temperature category of climate news shocks. The second and third plots show the response function of HMC's long and short legs. The shocks happen in the current month (month = 1). The red dotted lines represent the 95% confidence interval.

Figure 3.11: HMC Response Function to News Shocks Across Quarters



The plots above show the response function of HMC to the extreme temperature category of climate news shocks. The second and third plots show the response function of HMC during the first and second half of the year. The shocks happen in the current month (month = 1). The red dotted lines represent the 95% confidence interval.

Table 3.12: News shocks and HMC

	All industries		
	(1)	(2)	(3)
Constant	0.174	0.194	0.176
	(1.148)	(1.357)	(1.248)
News Shock	−0.169	−0.043	−0.101
	(−0.453)	(−0.109)	(−0.306)
Lag News Shock		**1.072*****	**1.038*****
		(3.000)	(2.649)
Observations	227	226	226
Adjusted R^2	−0.004	0.021	0.145

	Within industries with sensitivity to abnormal temperatures		
	(1)	(2)	(3)
Constant	0.126	0.151	0.092
	(0.482)	(0.598)	(0.376)
News Shock	−0.253	−0.103	−0.504
	(−0.456)	(−0.180)	(−1.010)
Lag News Shock		**1.344*****	**1.457*****
		(3.019)	(3.189)
Observations	226	225	225
Adjusted R^2	−0.004	0.015	0.181

	Within industries with no sensitivity to abnormal temperatures		
	(1)	(2)	(3)
Constant	0.161	0.164	0.158
	(1.253)	(1.256)	(1.400)
News Shock	−0.057	−0.110	0.051
	(−0.222)	(−0.396)	(0.187)
Lag News Shock		−0.371	−0.365
		(−1.464)	(−1.298)
Observations	226	225	225
R^2	0.0001	0.006	0.113
Adjusted R^2	−0.004	−0.003	0.071

This table reports the time-series regressions of HMC from different industry groups (with and without sensitivity to abnormal temperatures) against the extreme temperature category of climate news shocks. Climate news shock is the prediction error from the AR(1) model applied to the $MCCC$ index from Ardia et al. (2022). News shock is only based on the extreme temperature category of climate news shock series. The lag of news shocks is the shock from the previous month. Column (3) in both tables contains all the control variables from Table 3.6 but are not reported for expositional limitations. I include Newey-West t-statistics in parentheses. The sample is monthly from December 2007, when the $MCCC$ index data is available, to December 2021. The returns are in percentages.

Chapter 4

Common Sources of Uncertainty According to the SEC Filings

This paper re-examines the role of uncertainty in business decisions using a textual analysis approach that measures the uncertainty perceived by firms through quarterly and annual reports submitted to the SEC. Since firms discuss a range of topics in their reports, with the right methodology, it is possible to measure the textual uncertainty in the context in which it is expressed. Accordingly, I develop a categorical measure of aggregate uncertainty in the economy by identifying the six most common sources of uncertainty across 10-Q and 10-K filings. This is done by aggregating the most discussed topics in association with uncertain words across all filings at each point in time. The six common sources of uncertainty identified in this study are related to monetary policy, tax, financial market, exchange markets, financial intermediation, and oil & gas. I then measure the effect of uncertainty shocks under each category on capital investment at the firm level to answer the following question: What sources of uncertainty are important with respect to the delaying effect of uncertainty on investments?

In summary, my methodology first collects the terms associated with uncertain words - occur with uncertain words in the same sentence - across filings and identifies the terms with the highest frequency. Next, I demonstrate that there are six major topics that appear the most in association with uncertain tone in the SEC filings. I then document a negative correlation between quarterly shocks in my measure of aggregate uncertainty and investments in the subsequent quarter. However, the negative correlation is present only for shocks linked to particular sources of uncertainty, while I find no evidence of the delaying impact of uncertainty shocks from other sources.

The results point to only two sources of uncertainty with delaying effects on investment, namely monetary policy and oil & gas markets. I show that a standard deviation rise in uncertainty levels associated with monetary policy predicts a 0.037 standard deviation fall in investment across firms. Accounting for the mean and standard deviation of investment in my sample, the fall in investments is then equivalent to about 7% of average investment levels across firms. I record quarterly rises of up to 2.7 standard deviations under monetary policy category in the quarters leading to the 2001 recession. A standard deviation rise under oil&gas related textual uncertainty predicts a drop in investments of about 3%, with quarterly rises recorded up to 3.5 standard deviations in the wake of the recent pandemic period. Although partially in line with the literature, this indicates that not all uncertainty shocks have delaying effects on investments. In addition, I find no evidence for the delaying effects of uncertainty shocks associated with the financial market, exchange markets, and tax-related topics.

The literature on investment-uncertainty relation, which builds on the real options framework, establishes that investment activities respond to uncertainty in the presence of irreversibility of investment. The effect of uncertainty on the valuation and optimal exercise of real options is extensively discussed in the literature (see McDonald and Siegel, 1986; Dixit, 1989; Dixit and Pindyck, 2012). In parallel to the theoretical framework, the empirical evidence on the link between uncertainty and the contraction of investment activities is abundant and examined under various measures of uncertainty. Volatility-based measures, along with the recently developed text-based measures of uncertainty, all demonstrate a strong and negative relationship with investment (see Bulan, 2005; Gulen and Ion, 2016; Leahy and Whited, 1996; Alfaro, Bloom and Lin, 2016).

Furthermore, firm-level dynamics between investment and uncertainty may prove to be insightful in understanding business cycles, as uncertainty is countercyclical and rises in recessions. Bloom et al. (2018) develop a DSGE model with heterogeneous firms and show that exogenous uncertainty shocks lead to a significant drop in GDP and productivity growth due to a reduction in the degree of reallocation in the economy. This suggests that uncertainty shocks may play an important role in driving business cycles. Similarly, Gilchrist, Sim and Zakrajšek (2014) points to the effect of uncertainty as an exogenous factor on real activity to explain business cycle dynamics in the presence of financial frictions. While both studies assume the exogenous source of uncertainty shocks, the endogenous fluctuations in uncertainty as a result of economic downturns are also up for debate (Bachmann, Moscarini et al., 2011; Carriero, Clark and Marcellino, 2018). Bachmann, Moscarini et al. (2011) propose the idea that economic downturns induce firms to change their strategy for survival, which contributes to a rise in uncertainty as a result of riskier decisions.

In my view, examining the sources of uncertainty may clarify the origins of uncertainty shocks that drive business cycles as opposed to those that occur as a response to shifts in business cycles. Exploring the sources of uncertainty shocks, be it in relation to economic factors, regulations, financial markets, and others, is an important step in distinguishing between different scenarios under which economic activity and uncertainty interact. Accordingly, analyzing the text produced by firms helps develop a more nuanced view of how firms perceive uncertainty over time and make investment decisions that ultimately shape business cycles. Therefore, this study first and foremost contributes to the investment-uncertainty literature by proposing a measure of uncertainty constructed to reflect the perspective of firms.

Another related branch of literature includes studies that tackle the challenge of measuring uncertainty. Besides the text-based methods of measuring uncertainty, most studies on the uncertainty-investment relationship do not distinguish uncertainty from volatility and apply the concepts interchangeably (see Brainard et al., 1980; Ferderer, 1993; Pindyck and Solimano, 1993; Leahy and Whited, 1996; Bulan, 2005). Nevertheless, there are some novel and non-text-based methods that use informational content of the market to measure firms uncertainty level. David, Hopenhayn and Venkateswaran (2016), uses the degree of correlation between firms' decisions and market information tied to their stock price movements to estimate the uncertainty within the firm. The more a firm's decisions co-vary with its stock movements, the higher the uncertainty. Another approach to defining uncertainty is presented in David and Veronesi (2013), which applies the Bayesian updating rule to agents' beliefs about the future state of economic fundamentals to model uncertainty shocks. Such shocks are linked to updating of distribution parameters representing the agent's belief (belief shock) about the future, as the new information arrives(David and Veronesi (2001)). A growing body of literature has adopted similar approaches and expanded on the Bayesian updating of distribution parameters by suggesting that agents also revise their tail risk estimations with the arrival of new information (Kozlowski, Veldkamp and Venkateswaran, 2020; Izhakian, Yermack and Zender, 2021). Measuring textual uncertainty, on the other hand, is not an accurate statistical measure that can gauge the shifts in tail risk or the rate at which the distribution parameters update. Instead, it solely aims to measure the uncertainty in text using methodologies borrowed from the information retrieval literature. Textual analysis methods have evolved over time to optimally extract informational content of textual material. Textual measurement of uncertainty is a relatively recent practice in the economics and finance literature and only gained popularity over the past few years. Baker, Bloom and Davis (2016) is a seminal study that demonstrates the power of text-based measurement of uncertainty at the macro-level for the first time. In essence, their method relies on the occurrence of a set of manually selected words [1] across news articles. These words are particularly selected to reflect the uncertainty associated with government policies and legislation in the top ten US newspaper outlets.

In my analysis, I focus on the filings published by firms rather than news to capture the sources of uncertainty from the perspective of firms. Given the richness of future-oriented text in the SEC filings, surprisingly, they have received little attention in the literature when it comes to developing text-based measures of uncertainty. I develop such a measure using the Loughran-McDonald (LM) master dictionary that was initially developed in conjunction with their influential paper Loughran and McDonald (2011). LM dictionary is built using 10-Q, 10-K filings, newspaper articles, press releases, and investment message boards and has been constantly updated ever since. I use the uncertain words category in LM dictionary among various sentiment categories presented in the dictionary. Some examples of uncertain words in LM dictionary are *'unforseen', 'anticipate', 'believe', 'conditional', 'contingent'*[2]. my method of choice for textual analysis is bag-of-words used by Loughran and McDonald (2011) to estimate quantitative measures from textual content of documents. One major shortcoming of bag-of-words is that it parses text into its building blocks (words), and thus obtains a quantitative measure that is only based on word frequency, while disregarding the grammar, order of words, and in general, the context in which words are used. This may cause a significant loss of information when comparing documents with similar word content and yet different sequencing of words. My method combines bag-of-words with an approach to locate uncertain words in the text along with their neighboring terms in the same sentence and paragraph, thus capturing the context of the expressed uncertainty. The results are six distinct series of textual uncertainty shocks representing each source of uncertainty identified across firms.

To investigate the impact of uncertainty on investment I conduct a panel data analysis using firm-level data and regress the firm-quarter capital investment against uncertainty shock series associated with each source. Since the

[1] The words are: "economic", "economy", "uncertain", "uncertainty", "congress", "deficit", "Federal Reserve", "legislation", "regulation", "White House".

[2] For the full list of uncertain words and other categories in LM dictionary visit: Loughran (2021)

textual uncertainty series are correlated across various sources this proposes a challenge in measuring the isolated impact of uncertainty shocks from each source on investment. To mitigate this concern, I first disentangle the variations in each textual uncertainty shock series from others before implementing my panel regression analysis.

The paper proceeds as follows. Section 2 describes the data and methodology and discusses the process to construct the aggregate measure of uncertainty in depth with examples. Section 3 discusses the causality concerns in investigating and distinguishing the impact of uncertainty shocks from various sources on investment. Section 4 reports the main results and compares the predictive power of my measure with other measures in the literature. Section 5 further tests the sources of uncertainty with seemingly no delaying effects on investment for robustness check. Section 6 concludes.

4.1. Data & Methodology

4.1.1 Common sources of uncertainty

Before starting the textual analysis I must prepare the raw filings in order to filter out unimportant words and other components within the text. I used the parsed documents that are publicly available at Loughran (2021). These documents are pulled using the EDGAR system and already parsed using the algorithms mentioned in Loughran and McDonald (2011). In the parsing process, numbers, tables, and figures are removed, and the raw text is prepared for textual analysis. My sample includes 10-K and 10-Q filings from 1994-2021, with about 527 thousand quarterly and annual documents on 13,646 public firms.

To measure the uncertainty in quarter t, I aggregate the filings published in quarter t across all firms. Next, I locate all the words belonging to the uncertainty category based on the LM dictionary within each filing, along with their neighboring terms that show up both before and after the target uncertain words in the text. The technical term for any sequence of neighboring terms is n-grams.[3] In this study, I only focus on n-grams of size 2 or bigrams. I choose bigrams since relying on unigrams or single words in many cases does not allow an accurate identification of the source of uncertainty in the filings.[4] Yet, I do not expand my analysis beyond a sequence of two neighboring words in the vicinity of uncertain words due to time limitations associated with compute power needed to work with a sequence of words larger than two. The following example shows how I locate the target uncertain word "exposure" and the bigram "foreign exchange" associated with it in the paragraph from a filing:

"*Generally, the Company's practice is to hedge a majority of its material foreign exchange exposures, typically for 3 to 6 months. However, the Company may choose not to hedge certain* ***foreign exchange*** $\boxed{\textbf{\textit{exposures}}}$ *due to immateriality, prohibitive economic cost of hedging particular exposures, and limited availability of appropriate hedging instruments (Apple, 2008, July 10-Q).*"

In the paragraph above, the bigram "foreign exchange" is followed immediately by the target uncertain word "exposures". However, this is not always the case as shown in the following example:

[3] n-gram is a continuous sequence of n items from a given sample of text.

[4] For instance, if I locate the word "credit" as the unigram associated with uncertain words, I cannot distinguish whether it refers to a tax credit or credit risk. Similarly, "federal" may refer to any branches of the federal government and not necessarily the Federal Reserve.

*"the effective **tax rate** could **fluctuate** significantly on a quarterly basis and could be adversely affected to the extent earnings are lower than anticipated in countries that have lower statutory rates and higher than anticipated in countries that have higher statutory rates. (Alphabet, 2016, Nov 10-Q).*"

Here the bigram "tax rate" is separated from the target uncertain word "fluctuate" and thus is not detected by the algorithm that limits its searching boundary to a sequence of only two words. To mitigate this, I remove punctuations and stop words to boil down the text into only meaningful words and bigrams.[5] After modifications the paragraph above is transformed into the following paragraph which allows the algorithm to detect "tax rate" as the neighbouring bigram of "fluctuate":

*"effective **tax rate** **fluctuate** significantly quarterly basis adversely affected extent earnings lower anticipated countries lower statutory rates higher anticipated countries higher statutory rates.* "

Note that the size of the paragraph above shrinks significantly after removing the punctuations and stop words which are computationally favorable as well. I implement the procedure discussed for all the filings; Next, I gather all uncertain words and neighboring bigrams in a pool which is a vector of all the collected uncertain words and their neighboring bigrams. The unique bigrams that are used repeatedly used across filings are then identified. I count the number of times each unique bigram appears in the pool and assign a frequency to it accordingly. Finally, by arranging these unique bigrams based on their frequency, I find the most frequent bigrams as shown in Table 4.1.

I do not rely on automated algorithms to determine the topic of each bigram collected in the pool. Instead I manually browse through and assign a category to bigrams in the pool. However, since the total number of unique bigrams in the pool is 1,660,665 I refrain from manually browsing through the entire pool due to time limitations and rely on the power law feature of natural languages to limit the scope of my my search.[6]

The frequency of top bigrams in the pool constitutes a large portion of the sum of all frequencies across bigrams. In my sample, the top thousand most frequent bigrams account for up to 25% of the total frequency, while accounting for only about 0.006% (1000/1660665) of the total number of bigrams collected in the pool. Fig 4.1 illustrates this power law behavior of bigrams in the pool. The exponential decay in the frequency of bigrams based on their rank, allows me to focus on the top few thousand bigrams and identify the topics they are associate with. Taking into account that the frequency of bigrams drops exponentially with their rank, even if there are other sources of uncertainty to be discovered in the range of lower frequency bigrams, they are unlikely to be among the main sources of uncertainty across firms as evidenced by their low frequency.

When browsing through the bigram pool my goal is to find those bigrams, among the most frequent ones, that are

[5]Examples of stop words in English are "a", "the", "is", "are", "and", "could", etc.

[6]Words in natural languages roughly follow a power law distribution known as *Zipf's law*. The frequency of a few words is very high while the majority of words have very low frequencies on average (See Newman, 2005) Bigrams are constructed using words and as it appears here they also follow a power law distribution, while it is not clear whether this power law distribution exactly corresponds with Zipf's law. According to Zipf's law, the frequency of words is proportional to the inverse of their ranking: $\mathbf{f(r)} \propto \frac{\mathbf{1}}{\mathbf{r}^{\alpha}}$, with $\alpha \approx 1$ (See Mandelbrot, 1961).

sources of uncertainty beyond the firm's influence. These are sources of uncertainty that are not tied to a particular firm's business or financial concerns. I therefore pass on bigrams representing uncertainty related to topics such as sales, cash flows, supply chain, operational issues, and accounting items (such as intangible assets, fair value, etc.).[7] However, I must acknowledge that uncertainty about sales, cash flows, and supply chain may stem from underlying macroeconomic or industry-level uncertainty shocks. Yet since it is not clear whether the primary source of uncertainty about sales, cash flows, etc. is the result of firm-level decisions or shocks imposed on the firm from its environment, I am unable to link them to a particular source of uncertainty.

Following the steps described above, I find top bigrams associated with six major sources of uncertainty. Table 4.2 shows each category along with notable bigrams representing it. The rest of the frequent bigrams that fall outside of these categories are primarily accounting items or describe the uncertainty regarding business and financial practices within the firms, and in many cases do not represent any particular topics.

4.1.2 Aggregate textual uncertainty

To construct the quarterly aggregate uncertainty series for each source, I sum up the frequency of all bigrams belonging to a category and divide it by the total number of firms in each quarter identified through their unique CIK number. Thus I construct the average frequency series for each category by aggregating the frequency of bigrams in quarter t, belonging to category k and dividing it by the total number of firms to obtain the average frequency for category k at time t:

$$\textit{Average frequency}_{\mathbf{k},\mathbf{t}} = \frac{\sum(\textit{bigram} \quad \textit{frequency}_{\mathbf{k},\mathbf{t}})}{\textit{total} \quad \textit{number} \quad \textit{of} \quad \textit{firms}_{\mathbf{t}}}$$

The denominator ensures that I account for the fluctuating number of firms over the years and quarters to obtain an average quarterly frequency. Table 4.3 reports the correlation of quarterly average frequency series for each source with Baker, Bloom and Davis (2016) Economic and Policy Uncertainty index (BBD), CBOE Volatility Index (VIX) and Jurado, Ludvigson, and Jurado, Ludvigson and Ng (2015) total financial uncertainty index (JLN). After constructing the average frequency series for each category I normalize the series by their standard deviation.

Next, I apply the Hodrick-Prescott (HP) filter from Kaiser and Maravall (1999) to separate the cyclical from growth component in average frequency series, since I am interested in the quarterly shocks in uncertainty. Thus, the quarterly series $SECUX_k$, which tracks the textual uncertainty shocks with respect to source k is defined as:

[7] With respect to accounting terms I find that uncertain words such as ”approximate” or ”assume” among others are usually associated with bigrams like ”fair value” and ”intangible assets”, which in my view do not represent the uncertainty about accounting standards, but rather an inherent uncertainty regarding the valuation of intangible assets.

$$\textit{SECUX}_{\mathbf{k}} \equiv \textit{ Average frequency series}_{\mathbf{k}} - \textit{Average frequency series}_{\mathbf{k}}^{\mathbf{trend}}$$

And finally, I scale down the $SECUX$ values across all six categories to be between 0 and 1. Table 4.4 reports summary statistics of quarterly average frequency and $SECUX$ series.

As can be seen in Fig. 4.2 the uncertainty associated with monetary policies peaks during the quarter leading to the 2001 recession while in the other two recessions, it does not reach to its highest levels. In the case of tax uncertainty, the peak occurs before the 2008-09 crisis. According to my analysis, the high levels of uncertainty associated with taxes in this period can mostly be attributed to the high frequency of the bigram ”tax positions” in the pool. The uncertainty associated with the financial market peaks during all recessions with its highest values during the quarters leading the 2001 recession. It also peaks in quarters leading to 2007 only to fall and rise again in the 2008-2009 financial crisis. The uncertainty associated with exchange markets also peaks similarly during the 2001 recession and 2008-09 financial crisis, while also reaching its highest levels around the 2016 presidential election. It also once again peaks after the start of the pandemic and falls as the recession rolls in by the beginning of the year 2020. Financial intermediation uncertainty series peaks in the first quarter of 2010 after a sharp rise that starts during the 2008-09 crisis. Similar to financial intermediation uncertainty, the uncertainty associated with oil & gas markets does not seem to peak during the 2001 and 2008-09 recessions, however, it sets its record high levels during the most recent recession.

4.1.3 Pandemic driven uncertainty

In this section I address the most recent source of uncertainty shock associated with COVID-19 pandemic and discuss its implications for my results. During the recent pandemic, the global economy has received an uncommon yet massively disruptive uncertainty shock. The most immediate impacts of the pandemic on the economy targeted worldwide supply chains and caused unprecedented disruptions. Soon after, this led to economic downturns and the most recent recessionary period. The first two series in Fig 4.3, illustrate the magnitude of the downward shock to the aggregate investment and the real GDP growth in the aftermath of the pandemic. Using my methodology I have constructed the supply chain related textual uncertainty series to demonstrate the sharp rise in aggregate concerns about supply chain disruptions across the US firms.

As mentioned earlier, I concentrate on sources of uncertainty shocks that are the most common throughout my selected sample period. The pandemic as a source of uncertainty shock, although dominating the narratives from news to the SEC filings over the past two years, is still a relatively uncommon event. In my bigram pool, the bigrams associated with the pandemic period and COVID-19 in general, when aggregated over all quarters and firms, still fall far under the other six common sources of uncertainty identified so far. Therefore, I do not to include the pandemic as a common source of uncertainty due to its relatively recent and unprecedented impact on the economy.

The pandemic period and its impact on economic activity especially investments across firms may distort the results of my analysis and mistakenly attribute the downturn in economic activity to one of my identified sources. I will address this concern when I further explain the details of panel regression design to ensure that I capture the true firm-level impact of uncertainty shocks associated with each source.

4.1.4 Firm-level financial data

I use COMPUSTAT database to construct firm-level variables from 1994-2021 for 10,790 public firms. The total number of firm-quarter observation is 656,448. These variables are investment level, Tobin's Q, cash flows, and sales growth. Summary statistics of COMPUSTAT variables is presented in Table 4.5. I match the current firm-level investments with the values of $SECUX$ and other aggregate variables in the previous calendar quarter.

4.2. Causality concerns

There are two causality concerns in relation to the analysis of the effects of uncertainty on investment. First, is the case of reverse causality when analyzing how investment responds to uncertainty shocks. As firms decide to invest in projects and attempt to raise capital using various financing options, their forward looking tone changes in the text. Uncertainty in the filings are then inevitably rise after committing to investments. In addition, in many cases raising debt to fund the project adds another dimension to the uncertainty about the future. One should note that measuring textual uncertainty using firms' documents may always reflect the uncertainty caused by firms decisions in the past, rather than the other way around. However, by construction, my regression analysis goes around this problem by examining the impact of aggregate uncertainty on a single firm's investment activities in the cross-section of firms and over time. This is because the variations in $SECUX$ categories is solely driven from sources beyond any single firm's circle of influence and rooted in the macroeconomic environment. Yet it is still likely that events such as financial market volatility shocks or credit crises are to some extent influenced by the aggregation of firm-level decisions. Ultimately the aggregate impact of firms' investment activities contributes to macroeconomic cycles and thus uncertainty shocks that ensue. Considering this concern, I investigate the impact of current levels of aggregate uncertainty on the next quarter's firm-level investments to ensure that causality runs only in one direction.

The second causality concern is linked to the correlation among different sources of macroeconomic uncertainty shocks. This is an issue since my aim is to capture the effect of uncertainty on investment, separately for each source of uncertainty. This poses a concern that goes beyond just the dynamics between firms' investments and macroeconomic uncertainty and involves the government, markets, and possibly factors beyond the economy of a single country (e.g. shocks in oil and currency markets). For instance, a sudden market crash that causes an uncertainty shock about the future of the economy can in turn lead to rising uncertainty about the Federal Reserve's imminent decisions to mitigate the situation. Therefore, regardless of what the initial source of uncertainty shock is (e.g. the government, financial market, credit market, etc.), it is likely that the uncertainty associated with other sources also experience an unexpected shock. While disentangling different categories of $SECUX$ series is manageable following linear regression methods, describing the dynamics between macroeconomic uncertainty shocks from different sources is beyond the scope of this paper. Therefore I first control for potential correlations among $SECUX$ categories and then proceed to test their effects on investment activities separately. Accordingly, for the rest of my analysis I assume that the independent variations in each $SECUX$ category (from other $SECUX$ categories) reflects the changes in aggregate uncertainty that is solely associated with that category - at least from the perspective of the firm.

4.2.1 Correlations among textual uncertainty series

Before moving on to test the predictive power of each $SECUX$ category with respect to future firm-level investment, here I discuss the correlation among the categories. The rise and fall in some of $SECUX$ category may coincide with others and as a result, make it difficult to pinpoint the source of fluctuations in textual uncertainty. Table 4.6 reports the results from six time series regressions to determine the correlation amongst categories of $SECUX$. As expected all $SECUX$ series are strongly correlated with at least one other series from other categories. The portion of variations in $SECUX_{monetary\, policy}$ that covaries with other series is the largest reported, followed by $SECUX_{financial\, market}$ and $SECUX_{exchange\, markets}$.

With the high correlation among some categories according to the results in Table 4.6, there is always the risk of multicollinearity influencing the sign and significance of the coefficients. Therefore, I adopt a different approach by dividing my analysis into two parts. First, I use the time series results in Table 4.6 to orthogonalize each $SECUX_k$ with respect to other series. In this step, I collect both the residuals and the fitted values from the regressions in Table 4.6 and call them $SECUX_k\ perp$ and $SECUX_k\ fitted$.[8] Second, I use both $SECUX_k\ fitted$ and its orthogonal component $SECUX_k\ perp$ in my panel regressions to capture both the effects of the dependent and independent variations in each category on investments. This way if $SECUX_k$ is not the predictor of investment on its own and only proxies for the effect of other series, this should be reflected by comparing the coefficient of its orthogonal and fitted components in investment–uncertainty panel regressions.

4.3. SECUX & firm-level investment

To examine the predictive power of each categories of $SECUX$ with respect to firm-level investments, I conduct a panel data regression analysis and report the results in this section. The data used in my panel analysis is quarterly firm-level fundamentals from Compustat, over the period starting from January 1994 to December 2021. I empirically test the predictive power of all categories of $SECUX$ separately, with respect to the firm-level investments, leading one quarter ahead. To accomplish this, I design the baseline regression similar to the commonly used ones in the investment literature (see Gulen and Ion (2016)):

[8] Note that I only keep the significant coefficients in the regressions from Table 4.6 when collecting the fitted values and residuals.

$$\frac{\mathbf{CAPX_{i,t+1}}}{\mathbf{TA_{i,t}}} = \boldsymbol{\alpha_i} + \boldsymbol{\beta_1}\mathbf{SECUX_{k,t}} + \boldsymbol{\beta_2}\mathbf{TQ_{i,t}} + \boldsymbol{\beta_3}\mathbf{SG_{i,t}} + \boldsymbol{\beta_4}\frac{\mathbf{CF_{i,t}}}{\mathbf{TA_{i,t-1}}}$$
$$+\boldsymbol{\gamma}\mathbf{C_t} + \mathbf{FQ_{i,t}} + \boldsymbol{\epsilon_{i,t+1}}$$

Here, i indexes firms, t indexes calendar quarter, α_i are firm fixed effects, and $FQ_{i,t}$ are the fiscal quarter dummies to control for potential seasonality in the firm-level variables pulled from Compustat. Standard errors are all clustered at the firm level. The industry fixed effects are also included in the regressions to capture the varying degree of investment irreversibility associated with the nature of businesses across industries. The irreversibility of investment across various projects can influence the firm's investment decisions when faced with uncertainty. The literature identifies factors such as assets redeployability and the portion of physical assets rented among others (see Kim and Kung, 2017; Kessides, 1990), as determinants of the degree of investment irreversibility across industries. Here I assume that these factors do not vary over time and only across industries. Accordingly, I capture the effect of varying levels of irreversibility across industries by including the industry-fixed effects.

All the measures of uncertainty used in the regression above (including $SECUX$, BBD, VIX, JLN) are aggregate series and therefore do not vary in the cross-section of firms in my panel-data analysis. Hence following Gulen and Ion (2016), I do not include time-fixed effects in my regression since it absorbs the explanatory power of variables that are constant across firms but change over time, namely $SECUX$ series and other uncertainty measures. To control for the potential macroeconomic factors that change over time and thus may impact investment opportunities across firms, I include a macroeconomic control variable, once again following the methodology in Gulen and Ion (2016), which is represented as C_t in the regression model. To proxy for the changes in business investment opportunity set over time I use the quarterly Business Confidence Index (BCI), which monitors output growth and anticipates turning points in economic activity and confidence in the near future business performance.[9]

If the textual uncertainty shocks track the real uncertainty shocks, then I expect a $\beta_1 < 0$ due to the delaying effect of uncertainty shocks on investment in the presence of irreversibility of investment that is discussed in the literature. However, it is also expected to observe different magnitudes and significance levels for the coefficients across the six categories of $SECUX$, as it is not yet clear that the delaying effect of uncertainty on investment depends on the source of uncertainty.

4.3.1 Baseline regression results

Before presenting the panel regression results I must address once again the pandemic period and its impact on investments. The impact of worldwide supply chain disruptions and the broader macroeconomic impact of the pandemic on demand caused downward shocks to economic activity. The sharp drop in investment activities can potentially distort the captured effect of uncertainty on investments presented in my results since the fall in investment levels are likely

[9]For the detailed discussion of BCI visit https://data.oecd.org/leadind/business-confidence-index-bci.htm

to be a response to the pandemic that falls outside the six categories I identified so far. To address this issue, I run my entire analysis twice, once including and once excluding the pandemic period from my sample to ensure that the delaying effects of uncertainty under different categories is not ,in fact, the effect of the pandemic on investments.

The firm-level variables used as predictors are Tobin's Q, cash flows, and sales growth. These variables are all normalized by their sample standard deviation and their estimated coefficients in all the regressions are significant, confirming that they are all strong predictors of investment at the firm-level. The positive correlation of these variables with investment is in line with the results presented in the literature (see Gulen and Ion, 2016); Alti (2003); Fazzari, Hubbard and Petersen (1987). The coefficient for BCI which proxies for the quarterly investment opportunities is also significant and positively correlated with future investments. This is expected since higher values of the index suggest optimism and confidence in the near future business performance. Further, the sign of the coefficients for all the variables in the regression, including $SECUX$, do consistently remain the same across all regressions, indifferent to the inclusion of the pandemic period. This indicates that the captured effects of uncertainty on investments are present regardless of the shocks to economic activity due to the pandemic.

Reported in Table 4.7, the coefficients for $SECUX_k\,fitted$ in all categories are all negative and significant, except for $SECUX_{tax}\,fitted$. The latter is due to the negative correlation of $SECUX_{tax}$ with other series (see Table 4.6). For all other series, the results for $SECUX_k\,fitted$ coefficients in Table 4.7 along with the positive and strong correlations among the series reported in Table 4.6 implies that in at least one category of $SECUX$ the expected delaying effects on investment is present. However, this may also lead to the negative coefficients for other $SECUX_k\,fitted$ that are correlated with it.

To find the category(s) of $SECUX$ that are primarily responsible for the reported delaying effect of uncertainty on investment, I now turn to the reported coefficients for $SECUX\perp$ across categories. The orthogonal textual uncertainty ($SECUX\perp$) associated with monetary policy and oil and gas prices are the strongest predictors of delaying in investment activities, followed by the uncertainty associated with financial intermediation. A standard deviation increase in $SECUX_{monetary\,policy}$predicts a 0.036 (-0.487 $\times$ 0.075) standard deviation decline in investment activity across firms. The standard deviation of investment in my sample is about 0.045. Thus the total effect on investment is around 0.002 (0.036 $\times$ 0.0405). Accounting for the investment mean of about 0.029 across my sample, the fall in investment is equivalent to 6.9% of average investments. Next category of uncertainty in terms of magnitude of its delaying effect on investments is $SECUX_{oil\,\&\,gas}$, with a fall in investment equivalent to 3.1% of average investment as a result of one standard deviation rise in the measure - note that the coefficient of $SECUX_{oil\,\&\,gas}\perp$ is used for this interpretation as well.

Contrary to my expectations, however, tax, financial market, and currency-related textual uncertainty appear to have no delaying effects on investments. Meanwhile, the positive and significant coefficients of the orthogonal components under these categories may have multiple interpretations. Since the orthogonal parts in all the series have, by construction, zero correlation with their fitted parts then the positive coefficients of the former can hint at a positive or boosting effects of uncertainty on investment under these categories. However it may also be the case that the orthogonal components of $SECUX$ categories that are positively correlated with investments, although not correlated with their fitted counterparts, are still negatively correlated with the orthogonal parts of the following categories, $SECUX_{monetary\,policy}$, $SECUX_{oil\,\&\,gas}$, and $SECUX_{financial\,intermediation}$ which are based on the results predict lower future investments. This implies that the positive correlation with future investment may only capture the boost in investments due to lower values of textual uncertainty in categories demonstrating the delaying effects on investment. For instance, the positive correlations with investment in the case of $SECUX_{financial\,market}\perp$ can be linked to the fact that it has a negative correlation with uncertainty under the moderate policy category which has delaying effects on investment.

The results so far establish that textual uncertainty shocks measured through 10-Q and 10-K filings have a delaying effect on investments as predicted by real options literature. More importantly, the results also suggest that the source of uncertainty plays a role in the delaying impact on investment. This suggests that the delaying effect of uncertainty on investment depends on the source of uncertainty. I provide further evidence for the lack of response of firm-level investment to uncertainty shocks in the following chapters.

The results in Table 4.8 compare the predictive power of $SECUX$ categories against three major measures of uncertainty in the literature, namely CBOE volatility index or VIX, BBD, and JLN index. Each section in Table 4.8 reports panel regression results corresponding to the comparison of each category of $SECUX$ with VIX, BBD, and JLN index. As observed in the results for the sample that includes the pandemic period and otherwise, BBD index does not explain the predictive power of $SECUX_{monetary\,policy}\perp$. The reason that BBD index is particularly important here is that it partially measures the newspaper uncertainty about the Federal Reserve's policies, while $SECUX_{monetary\,policy}$ reflects the uncertainty about the Federal Reserve's policies from the perspective of the firm. $SECUX_{monetary\,policy}\,fitted$ seems to lose its delaying effect on investment in the presence of BBD index. This indicates that BBD index tracks the movements in $SECUX_{monetary\,policy}$ that correlates with other $SECUX$ categories. Moving on to the next uncertainty measure, when I include VIX in the regressions in the sample that excludes the pandemic period, I observe VIX's predictive power to be absorbed by $SECUX_{monetary\,policy}$. This is in line with my previous results that firm-level investments do not respond to the financial market uncertainty shocks that are independent of $SECUX_{monetary\,policy}$ and $SECUX_{oil\,\&\,gas}$.

Next is the second category of $SECUX$ in terms of its delaying effects on investment, namely $SECUX_{oil\,\&\,gas}$. The results are similar to the case of $SECUX_{monetary\,policy}$, as the fitted component of $SECUX_{oil\,\&\,gas}$ seem to lose its delaying effect in the presence of BBD index, while its orthogonal variations still remain significantly and negatively correlated with investment. The coefficient for VIX loses its significance when I include $SECUX_{oil\,\&\,gas}$ in the regressions (for the sample without the pandemic). This is similar to the case of $SECUX_{financial\,market}$ which correlates strongly with $SECUX_{monetary\,policy}$ and $SECUX_{oil\,\&\,gas}$. And in the presence of the latter uncertainties in the regression, no evidence for the delaying effects of $SECUX_{financial\,market}$ was found, as it is the case with VIX here. The implications again point to the same conclusion that financial market uncertainty shocks do not have, at least a direct, impact on investment activities across firms.

Another category of $SECUX$ with a delaying impact on investments according to the results in Table 4.7 is $SECUX_{financial\,intermediation}$. When compared with BBD index, the orthogonal component of $SECUX_{financial\,intermediation}$

loses its predictive power. This is expected as $SECUX_{financial\,intermediation}$ demonstrated a very small delaying effect on investment and has a high correlation with BBD index among other $SECUX$ categories. For other categories of $SECUX$, the results in Table 4.8 are the same as Table 4.7.

The findings so far first point to the strong delaying impact of $SECUX$ on investment under the monetary policy, oil & gas markets, and financial intermediation categories. The effect is strong and present even after including other measures of financial and economic uncertainty in regressions, except for $SECUX_{financial\,intermediation}$ against BBD index. Also, I find no evidence for the delaying effect of uncertainty on investment predicted by real options literature in the cases of financial markets, exchange markets, and tax-related uncertainty. There is no theoretical framework that captures the effects of uncertainty on investment while accounting for the source of uncertainty. One reason for this is the researchers' inability to identify all the possible sources of uncertainty in the firm's environment. However, I believe that current methodologies, made possible by the advancements in textual analysis, offer an opportunity to identify and quantify common concerns from the perspective of firms.

4.4. Uncertainty sources with no delaying effects?

Here I return to specific categories of $SECUX$ under which no evidence is found to support the delaying impact of uncertainty on investment. This may imply that uncertainty shocks under these categories do not elicit a wait-and-see response across firms. It is beyond the scope of this study to speculate on why that might be the case. However, here I test the alternative hypothesis that $SECUX$ values under these categories deviate from tracking the actual macroeconomic uncertainty, and thus, my results do not reflect the effects of real uncertainty shocks. If this proves to be the case, then using other measures of uncertainty in the literature that correspond to the $SECUX$ categories should reveal their true impact on investment that is not captured by my measure. I test this using other well-established measures of uncertainty such as VIX, BBD index, and its topical categories.

4.4.1 The case of $SECUX_{financial\,market}$

The most popular measure of market uncertainty in the literature is VIX which is widely used in academic and industry circles. I select VIX to test the impact of financial market uncertainty on firm-level investment. There are other major sources of uncertainty that move VIX, such as disaster events (wars, natural disasters, etc.), government policies, and financial intermediation, which are identified in a study by Manela and Moreira (2017). The news volatility index (NVIX) developed by Manela and Moreira (2017) constructed using the front-page coverage of the Wall Street Journal (WSJ) follows a methodology that identifies words and n-grams in the text with predictive power about variations in VIX. The measure is shown to perform well in predicting VIX values out of sample and more importantly breaks down the sources of variations in VIX into five categories. They report that the largest portion of variations in the NVIX (around 52%) comes from the stock market category with the top n-grams such as stock, market, industry. A very similar set of n-grams to my selected bigrams used to measure $SECUX_{financial\,market}$. Thus I use VIX as the alternative measure of financial market uncertainty in my panel analysis and filter out other sources of uncertainty that account for the rest of the variations in it. I account for their effects using BBD index as a proxy for general uncertainty in the economy - BBD index also accounts for uncertainty with respect tax codes. This way I plan to capture the isolated impact of only the financial market uncertainty shocks measured by VIX, on investment at the level of the firm.

Table 4.9 reports the panel regression results, including VIX as an uncertainty measure. I only include the sample with the pandemic period in my regressions since I found no significant impact associated with it in my results reported previously. As can be observed in column 1, when included by itself, VIX captures the delaying impact of uncertainty on investment. This, however, changes after adding BBD index (column 3), which absorbs the delaying effects of VIX. With the knowledge that the BBD index does not measure the financial market uncertainty, at least directly, the result here is in line with $SECUX_{financial\,market}$ coefficient in panel regressions before.[10] The implications are that variations in VIX that are not correlated with the general economic and political uncertainty are not a predictor of firm-level investments.

I conclude that the results from Table 4.9 support my previous findings on the lack of investment delaying impact using $SECUX_{financial\,market}$. It is important, however, to note that uncertainty shocks stemming from an event in the financial market can cause a ripple effect throughout the economy that cause credit shortages and raise uncertainty with regard to monetary policy which would, in turn, lead to a fall in investments.

4.4.2 The case of $SECUX_{exchange\,markets}$

Among the sources of uncertainty covered so far, the exposure to foreign exchange fluctuations besides interest rate risks are the most commonly managed sources of risk using derivatives instruments (Bodnar et al. (2011)). The relatively low volatility of exchange rates suggests that firms assume that the rate movements will remain within the predictable limited range and mostly hedge against the direction of the changes in rates (Bartram (2019)). I turn to BBD index and its sub-indices again, this time using its news component with a focus on sovereign debt and currency crises to test the impact of currency uncertainty on investments. The categorical indices of BBD index are derived using the Access World News database of over 2000 US newspapers.[11] Sovereign debt and currency crises topics are both clustered under the same category that I call the $BBD_{currency}$ here and include notable terms such as currency crisis, sovereign debt, and currency devaluation.

Following the methodology in the previous section, I include BBD index itself in the regressions to account for the general uncertainty in the economy in an attempt to isolate the impact of currency-related uncertainty on investment. I acknowledge that the $BBD_{currency}$ does not solely track currency-related uncertainty and partially measures the uncertainty associated with sovereign debt crises. However I assume that the latter source of uncertainty is captured by BBD index itself, as the general measure of uncertainty in the broad economic sense.

Table 4.10 reports the panel regression results using $BBD_{currency}$. As observed in the table, before including

[10] BBD index has three components; First is the news component that targets newspaper articles with terms mentioned in the footnote on page 3. Second, is the tax component, and third captures the forecaster disagreement about the future fiscal and monetary policy (see Baker, Bloom and Davis (2016))

[11] For more details and the terms list used to measure textual uncertainty under each category, visit https://www.policyuncertainty.com/categorical_terms.html

BBD index in the regression, its currency component seem to predict a boost in investment. I speculate that the lack of evidence for the delaying effect of currency risk on investments may be attributed to the relatively low volatility of exchange markets which perhaps indicates to a manageable exposure to currency risks through the extensive use of currency derivatives. An event-based investigation with a focus on a particular target currency may provide more concrete evidence on the delaying impact of currency uncertainty. I thus conclude there is no evidence on the delaying impact of $SECUX_{exchange\,markets}$ on investments across firms in line with the results produced based on $SECUX_{exchange\,markets}$.

4.4.3 The case of $SECUX_{tax}$

To examine the validity of my results on the effect tax related uncertainty on firm-level investment based on $SECUX_{tax}$, I use another component of BBD index, BBD_{tax}, that is constructed from the World Access News database.[12] To test the predictive power of the BBD_{tax}, again I use BBD index as a proxy for the general uncertainty to capture the isolated effects of tax uncertainty on investment. However since the two series are highly correlated, to avoid potential multicollinearity issues, I follow the same strategy as in the previous chapter and break down the BBD_{tax} to its orthogonal and fitted components with respect to BBD index. [13].

Table 4.11 reports the results of my panel analysis using both the $BBD_{tax}\,fitted$ and the $BBD_{tax}\perp$ in one regression. Once again the results are in line with what I found previously using $SECUX_{tax}$. I leave the implications and further investigations in this area to the accounting literature, which may benefit from my methodology to construct text-based measures of uncertainty with a focus on tax-related topics from the perspective of firms.

4.5. Robustness check

First I address the issue of time fixed effect in my panel analysis. I argued that I account for the effect of time on changes in business opportunities in the economy by including the Business Confidence index. There are other variables at the macroeconomic level that may also capture the changes in investment opportunities over time. I test the robustness of my results using two other variables, namely the Consumer Sentiment index and the current quarter's real GDP growth to account for time fixed effects.[14]

My results in Table 4.7 are robust for the coefficients of $SECUX\perp$ under almost all categories after including the Consumer Sentiment index and the real GDP growth as a replacement for the Business Confidence index. The exception to this is the coefficient for $SECUX_{financial\,intermediation}$ which loses its explanatory power after including the Consumer Sentiment index. All the results are also robust when no variables are included in the regressions to account for time-fixed effects. The same holds true for the coefficients of $SECUX\perp$ reported in Table 4.8.

The results in Tables Table 4.9, Table 4.10, and Table 4.11 are also robust to the inclusion of the Consumer

[12]The list of terms used for BBD_{tax}: taxes, tax, taxation, taxed

[13]The correlation between BBD index and its tax component is 0.9. Note that the issue of correlation was not the case for the $BBD_{currency}$ with a correlation of only 0.16

[14]Note that both of these variables are also used in Gulen and Ion (2016) to account for time fixed effects.

Sentiment index and the real GDP growth. Further, one can argue that VIX and BBD index both can account for the change in investment opportunity across firms and thus there is no need to include other variables to account for time-fixed effects. The results again remain unchanged after excluding the Consumer Sentiment and Business Confidence indices along with the real GDP growth.

4.6. Conclusion

I develop a methodology that relies on the neighbouring terms of uncertain words in the SEC filings with the objective to identify the topics associated with uncertainty discussed in the filings. A few of these neighbouring terms occur repeatedly across my large sample of the SEC filings, over time. I argue that this implies the topics associated with such terms represent the most common sources of uncertainty from the perspective of firms due to their exponentially higher frequency. After identifying the most frequent terms and identifying their topics, I construct an aggregate textual measure of uncertainty that tracks the aggregate number of the identified neighboring terms on a quarterly basis. Accordingly, the most common sources of uncertainty are monetary policy, especially interest rate, tax, financial market, exchange markets, financial intermediation, and oil & gas markets.

I then move on to examine the predictive power of my measure under each category with respect to future investment at the level of firm. Consistent with the literature, the shocks to my aggregate measure of uncertainty ($SECUX$) is found to be a strong and negative predictor of future investments, however conditional on the source of uncertainty. The top two sources of uncertainty shocks with the largest delaying effects are monetary policy and oil & gas markets. The coefficient for financial intermediation source of uncertainty shocks also suggests a delaying effect yet with a much lower impact size. My results are robust to the inclusion of the pandemic period in the sample.

The uncertainty shocks associated with the financial market, exchange markets, and tax on the other hand demonstrate a positive and strong correlation with future investment. I argue that the strong positive coefficient may just reflect a boost in investments in the periods where uncertainty levels are low under the categories with delaying effects, namely monetary policy and oil & gas markets. To further test my findings under these categories I use three other measures of uncertainty from the literature, namely VIX, $BBD_{currency}$, and the BBD_{tax}. I find similar results which only support my previous findings.

I compare the predictive power of each $SECUX$ category with BBD index, VIX, and JLN index from the literature. The results show that $SECUX$ under the monetary policy and oil & gas markets categories remains a strong and predictor of lower investment in the presence of well-established measures of uncertainty in the literature. Thus indicating that $SECUX$, captures a portion of fluctuations in macroeconomic uncertainty beyond the current measures in the literature.

The main implications of my results are the following. First, the textual uncertainty measure constructed using the SEC filings demonstrates a strong predictive power with respect to future investments. Second, although uncertainty shocks may arise from various sources in the economy, the average investment across firms primarily responds to uncertainty associated with the Federal Reserve's policy and the oil & gas market conditions.

Tables and figures

Table 4.1: Bigram Pool

Bigram	Frequency	Bigram	Frequency	Bigram	Frequency
fair value	106680	company will	28285	will continue	17714
lived intangible	95444	market risk	27220	quarter quarter	17616
interest rate	88768	useful lives	26485	adopting signature	17544
interest entities	66354	income taxes	26136	per share	17367
intangible assets	55558	assets liabilities	25801	assets indefinite	17356
interest rates	53136	carrying value	25351	assets sfas	17154
foreign currency	45642	based upon	23158	useful life	16999
fully paid	44093	asset retirement	22851	cash flow	16711
rate debt	43336	purchase price	22726	company abilities	16176
raise substantial	43030	results operations	22227	minimize use	16018
forward looking	40839	tax rate	21703	balance sheet	15563
tax positions	39939	liabilities readily	21693	acquired liabilities	15466
actual results	37594	primarily due	20956	shall conclusively	15257
financial statements	34751	obligation update	20055	based historical	15226
assets acquired	33994	exchange rate	19883	loan losses	15201
cash flows	33590	approximately million	19317	change control	15173
operating results	33347	company ability	19286	using significant	15120
indefinite lived	32382	million million	19267	core deposit	14986
risks uncertainties	31858	stock price	19116	future events	14944
assets goodwill	31673	involve known	19065	certain liabilities	14374
ability continue	31073	results may	18825	fair values	14312
interest entity	29598	period period	18503	federal statutory	14279
sfas goodwill	29507	sources actual	18256	inputs used	14081
inputs level	29349	typed form	18178	carrying amount	14005
common stock	28347	disclosures market	18127	amortization expense	13786

First 75 bigrams associated with uncertain words pooled from all the documents. Frequency refers to the number of times each unique bigram co-occured with an uncertain word. Bigrams highlighted in red represent various sources of uncertainty.

Figure 4.1: Bigram frequency based on their rank in my constructed pool of bigrams using the SEC filings.

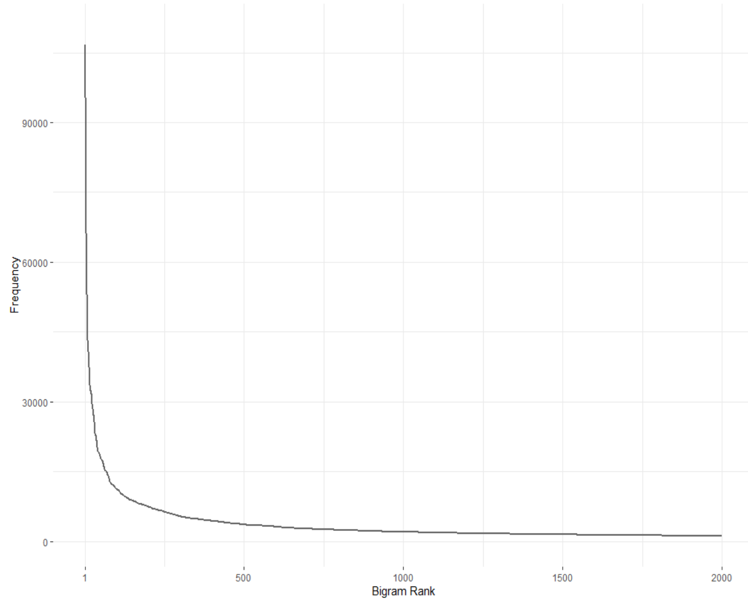


Table 4.2: Top Bigrams by Source

Uncertainty Source	Notable Bigrams
Monetary policy	interest rate, federal reserve, money supply
Tax	tax rate, statutory rate, effective tax
Financial market	stock price, market price, market risk
Exchange markets	currency risk, exchange rate, foreign exchange
Financial intermediation	credit loss, credit risk, loan loss
Oil & gas markets	oil & gas , natural gas, gas price, oil price

Notable bigrams representing the six sources of uncertainty. For an in-depth discussion on the choice of bigrams see Appendix A 1, choice of bigrams.

Table 4.3: Correlation of uncertainty measures

Average frequency$_k$ k:	BBD	VIX	JLN	Correlation among BBD, VIX, JLN	
Monetary policy	-0.01	0.18	0.23	cor(BBD, VIX)	0.44
Tax	0.23	-0.06	-0.08	cor(BBD, JLN)	0.39
Financial market	0.07	0.32	0.37	cor(VIX, JLN)	0.89
Exchange markets	0.30	0.06	0.17		
Financial intermediation	0.46	-0.02	-0.07		
Oil & gas markets	0.59	-0.09	-0.05		

Economic & Policy Uncertainty (BBD) developed by Baker, Bloom and Davis (2016), CBOE Volatility Index (VIX), and total financial uncertainty developed by Jurado, Ludvigson and Ng (2015) (JLN) are quarterly averages of the original monthly series. All series are seasonally adjusted based on the United States Census Bureau X-12 ARIMA-SEATS Seasonal Adjustment Program (henceforth X 12) method.

Table 4.4: Summary statistics - Aggregate textual uncertainty

	Mean	Median	Std. Dev	Min	Max
$Avg\,freq_{monetary\,policy}$	0.2940	0.2946	0.0838	0.1521	0.5531
$Avg\,freq_{tax}$	0.5571	0.6192	0.3163	0.1810	1.8726
$Avg\,freq_{financial\,market}$	0.23173	0.22975	0.0928	0.06216	0.54295
$Avg\,freq_{exchange\,markets}$	0.2977	0.3083	0.0778	0.1112	0.4634
$Avg\,freq_{financial\,intermediation}$	0.7048	0.6973	0.1058	0.5237	1.1770
$Avg\,freq_{oil\,\&\,gas}$	0.10925	0.10360	0.0366	0.05363	0.21978
$Total\,number\,of\,firms(reports)$	4746	4376	1697.8	1418	8156
$SECUX_{monetary\,policy}$	0.22327	0.21800	0.0747	0.03396	0.53790
$SECUX_{tax}$	0.2233	0.2169	0.1293	0.0000	0.9590
$SECUX_{financial\,market}$	0.22327	0.20875	0.0788	0.08063	0.52575
$SECUX_{exchange\,markets}$	0.22327	0.21044	0.0795	0.08803	0.54315
$SECUX_{financial\,intermediation}$	0.22327	0.21078	0.1340	0.02517	1.00000
$SECUX_{oil\,\&\,gas}$	0.22327	0.21611	0.0656	0.06133	0.54034

The total number of quarterly observations are 112 from January 1994 to December 2021. $SECUX$ series are all seasonally adjusted using X 12 method.

Figure 4.2: $SECUX$ series by category

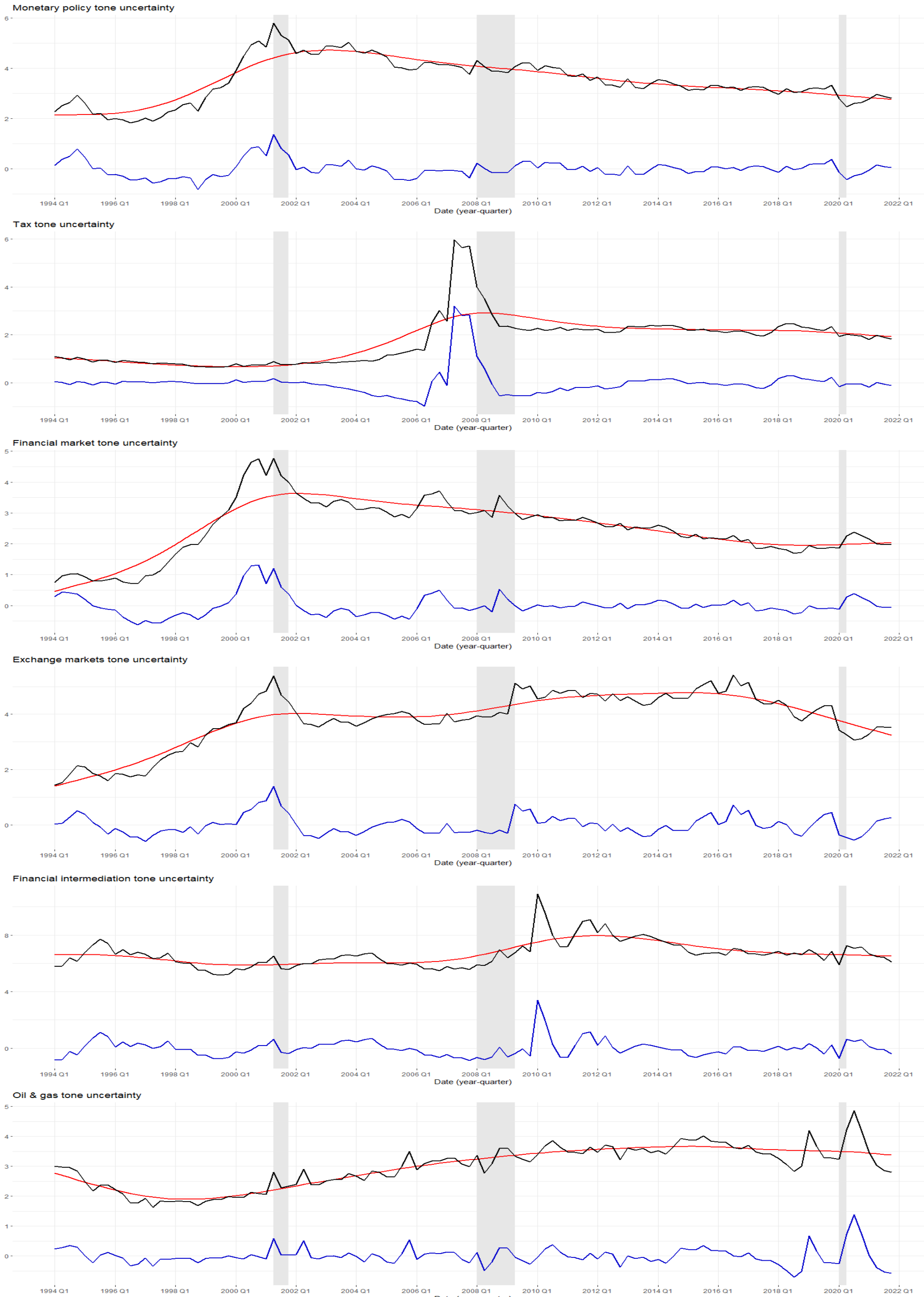


Figure above illustrates removing the growth component from the aggregate uncertainty series using HP filter for monetary policy and tax textual uncertainty series. The series are seasonally adjusted using the X 12 method. Plots in the following belong to other categories of textual uncertainty series. The grey boxes show recessionary periods according to the St. Louis Fed FRED.

Figure 4.3: The pandemic impact on economic activity and supply chain textual uncertainty

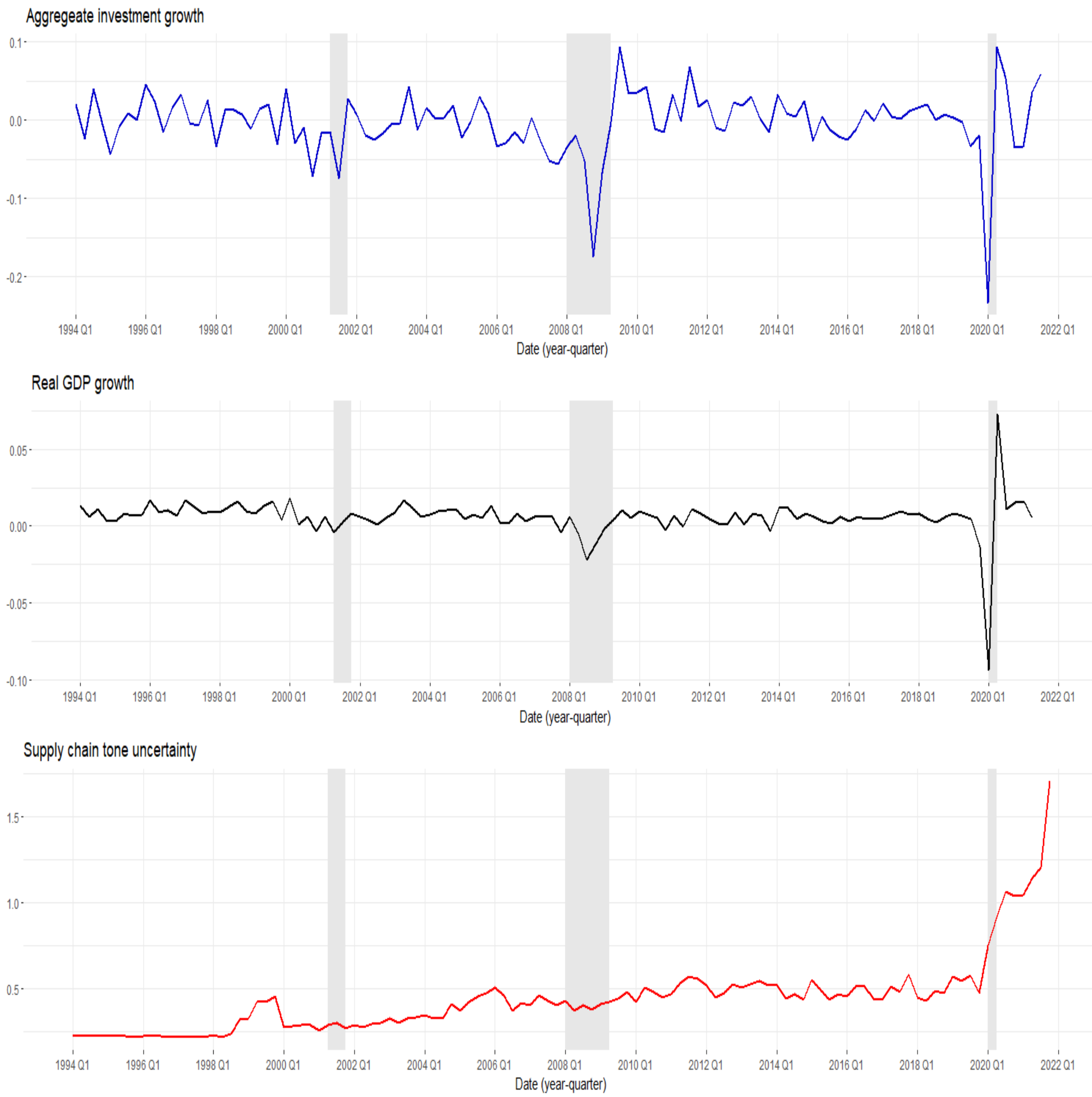


Figure above illustrates using only the bigram "supply chain" to construct the supply chain textual uncertainty series using my methodology. The grey boxes show recessionary periods according to the St. Louis Fed FRED.

Table 4.5: Summary statistics - COMPUSTAT data

	No. Obs	Mean	Std. Dev	Median
$Capital\,exp/lag\,total\,assets$	605093	0.02957	0.04479	0.01327
$Tobin's Q$	572388	1.9244	1.2968	1.4486
$Cash\,flows/lag\,total\,assets$	614021	0.19126	0.24675	0.08720
$Sales\,growth$	569464	0.01987	0.27844	0.02142

Investment is defined as $Capital\,expenditure_{t+1}/Total\,assets_t$. Tobin's Q is constructed as $(Total\,assets+(Equity\,market\,value)-Equity\,book\,value)/Total\,assets$. Cash flows is $Cash\,flows_t/Total\,assets_{t-1}$ and sales growth is $log(Sales_t) - log(sales_{t-1})$. Firm-level variables are winsorized at 1 and 99 percentile and normalized by their standard deviations before included in the panel regressions.

Table 4.6: Time series regressions among $SECUX$ categories

$Dependent variable$:	$SECUX_{monetary\,policy}$	$SECUX_{tax}$	$SECUX_{financial\,market}$	$SECUX_{exchange\,markets}$	$SECUX_{financial\,intermediation}$	$SECUX_{oil\,\&\,gas}$
$SECUX_{monetary\,policy}$	-	0.34958	0.629851***	0.65120 ***	0.58276 *	-0.13586
		(0.26181)	(0.094643)	(0.10329)	(0.26761)	(0.12991)
$SECUX_{tax}$	0.04732	-	0.022951	-0.11667**	-0.29178 **	-0.04037
	(0.03544)		(0.041401)	(0.04310)	(0.09656)	(0.04788)
$SECUX_{financial\,market}$	0.46788***	-0.00158	-	0.12604	-0.37004	0.43935***
	(0.07030)	(0.22721)		(0.10367)	(0.23300)	(0.10414)
$SECUX_{exchange\,markets}$	0.41878 ***	-0.55433**	0.109114	-	-0.36229 .	-0.11872
	(0.06643)	(0.20475)	(0.089749)		(0.21651)	(0.10408)
$SECUX_{financial\,intermediation}$	0.07348 *	-0.27180 **	-0.062809	-0.07103.	-	0.06878
	(0.03374)	(0.08995)	(0.039548)	(0.04245)		(0.04588)
$SECUX_{oil\,\&\,gas}$	-0.07517	-0.16503	0.327240***	-0.10214	0.30183	-
	(0.07188)	(0.19572)	(0.077566)	(0.08954)	(0.20134)	
Adj. R^2	0.632	0.09383	0.5549	0.4945	0.09474	0.1378

***p$<$0.001. Standard errors in parentheses. The results in the table above are from time series regressions where each category of $SECUX$ is regressed against its other categories. All the series are seasonally adjusted using X12 method. The residuals from these regressions are the portion of variations in each $SECUX_k$ series that is independent of other categories. The residuals (only including significant coefficients) are the $SECUX\perp$ series used in my panel data analysis later on.

Table 4.7: Dependent variable: $CAPX_{i,t+1}/TotalAssets_{i,t}$

k:	$Monetary\, policye$		Tax		$Financial\, market$	
	$pandemic\, included$	$pandemic\, not\, included$	$pandemic\, included$	$pandemic\, not\, included$	$pandemic\, included$	$pandemic\, not\, included$
$SECUX_k\, fitted$	-0.11643***	-0.12662***	0.17313***	0.70846***	-0.37279***	-0.37147***
$SECUX_k \perp$	-0.48671***	-0.59841***	0.05956***	0.05486***	0.34777***	0.46110***
$Business\, confidence$	0.02118***	0.01686***	0.01667***	0.01119***	0.01909***	0.014187***
$Tobin's\, Q_i$	0.10410***	0.10630***	0.10529***	0.10761***	0.10372***	0.10586***
$Cash\, flow_i$	0.00425*	0.00530**	0.00481**	0.00562**	0.00444*	0.00566**
$Sales\, growth_i$	0.01638***	0.01671***	0.01662***	0.01714***	0.016600***	0.01693***
Adj. R^2	0.1203	0.123	0.1194	0.1221	0.1209	0.1236
No. Obs	492660	472565	492660	472565	492660	472565
firm fixed effect	yes	yes	yes	yes	yes	yes
financial quarter dummy	yes	yes	yes	yes	yes	yes
Industry fixed effect	yes	yes	yes	yes	yes	yes
k:	$Exchange\, markets$		$Financial\, intermediation$		$Oil\, \&\, mgas$	
	$pandemic\, included$	$pandemic\, not\, included$	$pandemic\, included$	$pandemic\, not\, included$	$pandemic\, included$	$pandemic\, not\, included$
$SECUX_k\, fitted$	-0.35264***	-0.40865***	-0.38517***	-0.41576***	-0.21698***	-0.06941
$SECUX_k \perp$	0.17486***	0.07659***	-0.03571***	-0.02368***	-0.30908***	-0.19187***
$Business\, confidence$	0.01951***	0.01511***	0.01561***	0.01114***	0.01606***	0.01400***
$Tobin's\, Q_i$	0.10449***	0.10693***	0.10531***	0.10779***	0.10495***	0.10745***
$Cash\, flow_i$	0.00437*	0.00516**	0.00514**	0.00615**	0.00508**	0.00590**
$Sales\, growth_i$	0.01626***	0.01656***	0.01664***	0.01700***	0.01658***	0.01686***
Adj. R^2	0.1203	0.1226	0.1197	0.1219	0.1199	0.1216
No. Obs	492660	472565	492660	472565	492660	472565
firm fixed effect	yes	yes	yes	yes	yes	yes
financial quarter dummy	yes	yes	yes	yes	yes	yes
Industry fixed effect	yes	yes	yes	yes	yes	yes

Baseline panel regression results. $^{\cdot}$p<0.1;*p<0.05; **p<0.01; ***p<0.001. Standard errors are clustered at the firm level. The firm and industry fixed effects are assumed to be constant over time and financial quarter dummies are adjusted firm by firm. The industry is defined at 3-digit NAICS code level. It is worth mentioning that not including industry-fixed effects did not change the results so it is not reported in the table above. All the independent variables lag one-quarter behind the dependent variable. Aggregate variables are all seasonally adjusted using the X12 method and are quarterly averages of their respected monthly series. Business confidence is from The Organization for Economic Co-operation and Development (OECD). $SECUX_k \perp$ is the orthogonal variations and $SECUX_k\, fitted$ is the correlated variations in $SECUX_k$ with respect to other categories of $SECUX$. Firm-level variables are divided by their standard deviation for normalization and winsorized at 1 and 99 percentile levels.

Table 4.8: Dependent variable: $CAPX_{t+1}/TotalAssets_t$

	Pandemic included			Pandemic not included		
$SECUX_{monetary\ policy}\ fitted$	0.05815***	-0.10938***	-0.04820***	0.06608	-0.13020***	-0.09855***
$SECUX_{monetary\ policy}\perp$	-0.38532***	-0.49047***	-0.48118***	-0.34031***	-0.59697***	-0.59395***
BBD	-0.00216***			-0.00225***		
VIX		-0.00061***			0.00030.	
JLN			-0.11878***			-0.04772***
Adj. R^2	0.129	0.1203	0.1209	0.1291	0.123	0.123
$SECUX_{tax}\ fitted$	0.56439***	0.21578***	0.19766***	0.48433***	0.72585***	0.70460***
$SECUX_{tax}\perp$	0.09026***	0.06758***	0.076363***	0.08057***	0.05841***	0.06569***
BBD	-0.00227***			-0.00231***		
VIX		-0.00101***			-0.00043***	
JLN			-0.14047***			-0.07901***
Adj. R^2	0.129	0.1195	0.1202	0.1291	0.1221	0.1223
$SECUX_{financial\ market}\ fitted$	-0.11355***	-0.36493***	-0.31238***	-0.11621***	-0.36874***	-0.33399***
$SECUX_{financial\ market}\perp$	0.39555***	0.37536***	0.425130***	0.38083***	0.46990***	0.50119***
BBD	-0.00216***			-0.00220***		
VIX		-0.00110***			-0.00036*	
JLN			-0.13983***			-0.08086***
Adj. R^2	0.1294	0.121	0.1216	0.1295	0.1236	0.1238
$SECUX_{exchange\ markets}\ fitted$	-0.15889***	-0.35126***	-0.30445***	-0.11810***	-0.41103***	-0.39411***
$SECUX_{exchange\ markets}\perp$	0.06367***	0.173111***	0.18436***	0.07262***	0.07793***	0.08102***
BBD	-0.00212***			-0.00225***		
VIX		-0.00038***			0.00050**	
JLN			-0.10057***			-0.02825***
Adj. R^2	0.1286	0.1203	0.1207	0.1287	0.1226	0.1226
$SECUX_{financial\ intermediation}\ fitted$	-0.30646***	-0.40846***	-0.40597***	-0.25737***	-0.41811***	-0.42910***
$SECUX_{financial\ intermediation}\perp$	0.09398***	-0.03362***	-0.03563***	0.08449***	-0.02347**	-0.02317**
BBD	-0.00226***			-0.00238***		
VIX		-0.00094***			-0.00009	
JLN			-0.13507***			-0.07661***
Adj. R^2	0.1289	0.1197	0.1204	0.129	0.1219	0.1221
$SECUX_{oil\ \&\ gas}\ fitted$	0.43197**	-0.18971***	0.03412***	0.35175***	-0.08287***	0.07508***
$SECUX_{oil\ \&\ gas}\perp$	-0.03419.	-0.30990***	-0.34883***	-0.17760***	-0.19093***	-0.21871***
BBD	-0.00223***			-0.00238***		
VIX		-0.00050**			0.00025	
JLN			-0.14590***			-0.08860***
Adj. R^2	0.1286	0.1199	0.1207	0.1289	0.1216	0.1218
No. Obs	492660	492660	492660	472565	472565	472565
firm fixed effect	yes	yes	yes	yes	yes	yes
financial quarter dummy	yes	yes	yes	yes	yes	yes
Industry fixed effect	yes	yes	yes	yes	yes	yes

Panel results including multiple measures of uncertainty $\cdot p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$. Standard errors are clustered at the firm level. All other variables reported in Table 4.7 are included in each of the panel regressions above but not reported in the table for expositional purposes.

Table 4.9: Dependent variable: $CAPX_{t+1}/TotalAssets_t$

VIX	-0.00058***		0.00363***
BBD		-0.00218***	-0.00240***
$Business\,confidence$	0.016629***	-0.00830***	0.000745
$Tobin's\,Q_i$	0.105085***	0.10090***	0.10122***
$Cash\,flow_i$	0.00463**	0.00362*	0.00423*
$Sales\,growth_i$	0.01645***	0.014810***	0.015166***
Adj. R^2	0.1193	0.1284	0.1292
No. Obs	492660	492660	492660
firm fixed effect	yes	yes	yes
financial quarter dummy	yes	yes	yes
Industry fixed effect	yes	yes	yes

Panel results VIX and BBD index $^{\cdot}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$. Standard errors are clustered at firm level and the pandemic period is included in all regressions.

Table 4.10: Dependent variable: $CAPX_{t+1}/TotalAssets_t$

$\text{BBD}_{currency}$	0.00017***		0.00027***
BBD		-0.00218***	-0.00257***
$Business\,confidence$	0.01889***	-0.00830***	-0.01249***
$Tobin's\,Q_i$	0.10566***	0.10090***	0.10089***
$Cash\,flow_i$	0.00501**	0.00362*	0.00385*
$Sales\,growth_i$	0.016699***	0.014810***	0.01477***
Adj. R^2	0.1211	0.1284	0.1331
No. Obs	492660	492660	492660
firm fixed effect	yes	yes	yes
financial quarter dummy	yes	yes	yes
Industry fixed effect	yes	yes	yes

Panel results BBD and its currency component $^{\cdot}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$. Standard errors are clustered firm level and the pandemic period included in all regressions.

Table 4.11: Dependent variable: $CAPX_{t+1}/TotalAssets_t$

$BBD_{tax}\, fitted$	-0.00156***
$BBD_{tax} \perp$	0.00057***
$Business\, confidence$	-0.00775***
$Tobin's\, Q_i$	0.10094.
$Cash\, flow_i$	0.003312***
$Sales\, growth_i$	0.014809***
Adj. R^2	0.1288
No. Obs	492660
firm fixed effect	yes
financial quarter dummy	yes
Industry fixed effect	yes

Panel results BBD tax component $^{\cdot}$p$<$0.1;*p$<$0.05; **p$<$0.01; ***p$<$0.001. Standard errors are clustered firm level and the pandemic period is included in all regressions. $BBD_{tax}\, fitted$ and $BBD_{tax} \perp$ are collected from BBD_{tax} (t)= $\alpha + \beta$BBD$(t) + \epsilon(t)$

Chapter 5

Conclusion

This thesis includes three studies that benefit from textual information in the SEC filings in addressing topics related to the following three branches of the literature: measuring value of intangible assets, physical climate risks, and the effect of uncertainty on investment activities.

The first chapter develops a textual measure of intangibles using 10-K filings and assesses its informativeness about future returns. By constructing a long-short portfolio strategy based on this measure, we find significant positive returns from 1995 to 2020, excluding 2000 and 2001 representing the dot–com bubble burst and the poor performance of intangible intensive technology stocks. Our portfolio outperforms traditional and intangible-augmented value strategies, particularly in recent years. Testing against Fama and French's along with other common risk factors, our measure shows a significant positive alpha, suggesting that intangibles talk provides unique information not fully captured by the previous accounting measure of intangible value. The abnormal returns indicate that investors may not fully price intangibles disclosed in 10-K filings, aligning with similar findings on R&D and employee satisfaction mispricing in the literature. Further, the mispricing is pronounced in stocks with high idiosyncratic volatility (IVOL), supporting the notion that limits to arbitrage exacerbate this mispricing. Our study contributes to the literature by demonstrating that value-relevant information from textual disclosures on intangibles is not fully reflected in stock prices.

The study presented in the second chapter makes two key contributions. First, it presents a firm-level measure of exposure to abnormal temperature using textual data from 10-K filings, revealing that exposure to abnormal temperature negatively affects earnings. This impact on earnings is only present in five industries based on Fama and French twelve industry classification and happens through the effects of abnormal temperature on revenues and operating expenses. The negative economic effect of abnormal temperature is amplified for firms with low geographic dispersion. Second, I identify a premium for exposure to abnormal temperature by constructing a long-short portfolio based on firms' exposure levels, finding that this premium has grown significantly in recent years, mirroring trends in ESG–based investment returns. This premium is particularly pronounced in the same five industries and among firms with low geographic dispersion, suggesting that higher geographic dispersion can mitigate climate risks. Additionally, the abnormal temperature premium responds positively to unexpected climate concern increases, indicating that investor demand for higher premium during periods of elevated climate concerns.

In the third chapter, I develop a methodology that uses the neighboring terms of uncertain words in SEC filings to identify common sources of uncertainty in firm's environment. The most common sources of uncertainty according to this methodology are related to monetary policy, tax, financial markets, exchange markets, financial intermediation, and oil & gas markets. By analyzing the frequency of terms under each source across all filings and every quarter, I construct a textual index of uncertainty. I find that the shocks to this index are a strong predictor of lower future investment at the firm-level, particularly when the uncertainty is about monetary policy and oil & gas markets. Conversely, uncertainty related to financial markets, exchange markets, and tax correlates positively with future investment. These results remain unchanged even after including the pandemic period and using other established measures of uncertainty in the literature, namely VIX, economic policy uncertainty developed by Baker, Bloom and Davis (2016), and index of macroeconomic uncertainty developed by Jurado, Ludvigson and Ng (2015). $SECUX$, particularly for monetary policy and oil & gas markets categories, proves to be a robust predictor of delays in investment beyond existing measures. The key implications of the results from the third chapter are that average firm investment is most responsive to uncertainty about the Federal Reserve policies and the oil & gas market conditions.

In conclusion, the findings from the studies detailed in this thesis underscore the potential of textual analysis as a robust methodology for constructing indices that capture fluctuations in intangible value, geographical presence, and corporate uncertainty both across firms and over time. In addition, the informativeness of intangibles talk regarding future returns suggests that textual information is currently underutilized in stock valuation. This highlights the importance of enhancing the integration of textual data into financial analysis frameworks moving forward.

Bibliography

Aboody, David. 1996. "Recognition versus disclosure in the oil and gas industry." *Journal of Accounting Research* 34:21–32.

Aboody, David and Baruch Lev. 1998. "The value relevance of intangibles: The case of software capitalization." *Journal of Accounting Research* 36:161–191.

Addoum, Jawad M, David T Ng and Ariel Ortiz-Bobea. 2020. "Temperature shocks and establishment sales." *The Review of Financial Studies* 33(3):1331–1366.

Addoum, Jawad M, David T Ng and Ariel Ortiz-Bobea. 2023. "Temperature shocks and industry earnings news." *Journal of Financial Economics* 150(1):1–45.

Ahmed, Anwer S, Emre Kilic and Gerald J Lobo. 2006. "Does recognition versus disclosure matter? Evidence from value-relevance of banks' recognized and disclosed derivative financial instruments." *Accounting Review* 81(3):567–588.

Alfaro, Ivan, Nicholas Bloom and Xiaoji Lin. 2016. The real and financial impact of uncertainty shocks. In *Stanford Institute for Theoretical Economics (SITE) 2016 Workshop*.

Alok, Shashwat, Nitin Kumar and Russ Wermers. 2020. "Do fund managers misestimate climatic disaster risk." *The Review of Financial Studies* 33(3):1146–1183.

Alti, Aydoğan. 2003. "How sensitive is investment to cash flow when financing is frictionless?" *The journal of finance* 58(2):707–722.

Ang, Andrew, Robert J Hodrick, Yuhang Xing and Xiaoyan Zhang. 2006. "The cross-section of volatility and expected returns." *Journal of Finance* 61(1):259–299.

Ardia, David, Keven Bluteau, Kris Boudt and Koen Inghelbrecht. 2022. "Climate change concerns and the performance of green vs. brown stocks." *Management Science* .

Arnott, Robert D, Campbell R Harvey, Vitali Kalesnik and Juhani T Linnainmaa. 2021. "Reports of value's death may be greatly exaggerated." *Financial Analysts Journal* 77(1):44–67.

Asness, Clifford S, R Burt Porter and Ross L Stevens. 2000. "Predicting stock returns using industry-relative firm characteristics." *Social Science Research Network*. "https://papers.ssrn.com/sol3/papers.cfm?abstract_id=213872" .

Aswani, Jitendra, Aneesh Raghunandan and Shivaram Rajgopal. 2023. "Are carbon emissions associated with stock returns?" *Review of Finance, forthcoming* .

Bachmann, Rüdiger, Giuseppe Moscarini et al. 2011. Business cycles and endogenous uncertainty. In *2011 Meeting Papers*. Vol. 36 Society for Economic Dynamics pp. 82–99.

Baker, Scott R, Nicholas Bloom and Steven J Davis. 2016. "Measuring economic policy uncertainty." *The quarterly journal of economics* 131(4):1593–1636.

Bansal, Ravi, Dana Kiku and Marcelo Ochoa. 2016. Price of long-run temperature shifts in capital markets. Technical report National Bureau of Economic Research.

Bansal, Ravi and Marcelo Ochoa. 2011. Welfare costs of long-run temperature shifts. Technical report National Bureau of Economic Research.

Bartram, Söhnke M. 2019. "Corporate hedging and speculation with derivatives." *Journal of Corporate Finance* 57:9–34.

Belo, F, V Gala, J Salomao and MA Vitorino. 2022. "Decomposing firm value (0898-2937)." *Journal of Fnancial Economics 143 (2), 619–639* .

Birru, Justin and Trevor Young. 2020. "Sentiment and uncertainty." *Fisher College of Business Working Paper* (2020-03):010.

Blanc, Elodie and Wolfram Schlenker. 2017. "The use of panel models in assessments of climate impacts on agriculture." *Review of Environmental Economics and Policy* .

Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten and Stephen J Terry. 2018. "Really uncertain business cycles." *Econometrica* 86(3):1031–1065.

Bodnar, Gordon M, Erasmo Giambona, John R Graham, Campbell R Harvey and Richard C Marston. 2011. Managing risk management. In *AFA 2012 Chicago Meetings Paper*.

Bolton, Patrick and Marcin T Kacperczyk. 2021. "Carbon disclosure and the cost of capital." *Available at SSRN 3755613* .

Brainard, William C, John B Shoven, Laurence Weiss, Phillip Cagan and Robert E Hall. 1980. "The financial valuation of the return to capital." *Brookings Papers on Economic Activity* 1980(2):453–511.

Bulan, Laarni T. 2005. "Real options, irreversible investment and firm uncertainty: new evidence from US firms." *Review of Financial Economics* 14(3-4):255–279.

Burke, Marshall, Solomon M Hsiang and Edward Miguel. 2015. "Global non-linear effect of temperature on economic production." *Nature* 527(7577):235–239.

Cao, Jie and Bing Han. 2016. "Idiosyncratic risk, costly arbitrage, and the cross-section of stock returns." *Journal of Banking and Finance* 73:1–15.

Cao, Melanie and Jason Wei. 2005. "Stock market returns: A note on temperature anomaly." *Journal of Banking & Finance* 29(6):1559–1573.

Carhart, Mark M. 1997. "On persistence in mutual fund performance." *Journal of finance* 52(1):57–82.

Carriero, Andrea, Todd E Clark and Massimiliano Giuseppe Marcellino. 2018. "Endogenous uncertainty.".

Chan, Louis KC, Josef Lakonishok and Theodore Sougiannis. 2001. "The stock market valuation of research and development expenditures." *Journal of finance* 56(6):2431–2456.

Choi, Darwin, Zhenyu Gao and Wenxi Jiang. 2020. "Attention to global warming." *The Review of Financial Studies* 33(3):1112–1145.

Choi, Ki-Soon, Eric C So and Charles CY Wang. 2021. "Going by the Book: Valuation Ratios and Stock Returns." *Social Science Network.* "https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3854022.".

Cochrane, JH. 2001. "Asset Pricing Princeton University Press: Princeton.".

Colacito, Riccardo, Bridget Hoffmann and Toan Phan. 2019. "Temperature and growth: A panel analysis of the United States." *Journal of Money, Credit and Banking* 51(2-3):313–368.

Corrado, Carol, Charles Hulten and Daniel Sichel. 2005. Measuring capital and technology: An expanded framework. In *Measuring capital in the new economy*. University of Chicago Press pp. 11–46.

Cuculiza, Carina, Alok Kumar, Wei Xin and Chendi Zhang. 2022. "Climate sensitivity, mispricing, and predictable returns." *Mispricing, and Predictable Returns (November 26, 2022)* .

Curtis, Asher, Sarah McVay and Sara Toynbee. 2020. "The changing implications of research and development expenditures for future profitability." *Review of Accounting Studies* 25:405–437.

Daniel, Kent and Sheridan Titman. 2006. "Market reactions to tangible and intangible information." *Journal of Finance* 61(4):1605–1643.

David, Alexander and Pietro Veronesi. 2001. "Inflation and earnings uncertainty and the volatility of asset prices: an empirical investigation." *Manuscript, Graduate School of Business, University of Chicago* .

David, Alexander and Pietro Veronesi. 2013. "What ties return volatilities to price valuations and fundamentals?" *Journal of Political Economy* 121(4):682–746.

David, Joel M, Hugo A Hopenhayn and Venky Venkateswaran. 2016. "Information, misallocation, and aggregate productivity." *The Quarterly Journal of Economics* 131(2):943–1005.

Davis-Friday, Paquita Y, L Buky Folami, Chao-Shin Liu and H Fred Mittelstaedt. 1999. "The value relevance of financial statement recognition vs. disclosure: Evidence from SFAS No. 106." *Accounting Review* 74(4):403–423.

Deng, Zhen, Baruch Lev and Francis Narin. 1999. "Science and technology as predictors of stock performance." *Financial Analysts Journal* 55(3):20–32.

Deschênes, Olivier and Michael Greenstone. 2011. "Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US." *American Economic Journal: Applied Economics* 3(4):152–185.

Dixit, Avinash. 1989. "Entry and exit decisions under uncertainty." *Journal of political Economy* 97(3):620–638.

Dixit, Robert K and Robert S Pindyck. 2012. *Investment under uncertainty*. Princeton university press.

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

Edmans, Alex, Lucius Li and Chendi Zhang. 2014. Employee satisfaction, labor market flexibility, and stock returns around the world. Technical report National Bureau of Economic Research, Cambridge, MA.

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

Eisfeldt, Andrea L, Edward Kim and Dimitris Papanikolaou. 2020. Intangible value. Technical report National Bureau of Economic Research, Cambridge, MA.

Enache, Luminita and Anup Srivastava. 2018. "Should intangible investments be reported separately or commingled with operating expenses? New evidence." *Management Science* 64(7):3446–3468.

Ewens, Michael, Ryan H Peters and Sean Wang. 2019. Measuring intangible capital with market prices. Technical report National Bureau of Economic Research, Cambridge, MA.

Faccini, Renato, Rastin Matin and George Skiadopoulos. 2023. "Dissecting climate risks: Are they reflected in stock prices?" *Journal of Banking & Finance* 155:106948.

Falato, Antonio, Dalida Kadyrzhanova, Jae Sim and Roberto Steri. 2022. "Rising intangible capital, shrinking debt capacity, and the US corporate savings glut." *Journal of Finance* 77(5):2799–2852.

Fama, Eugene F and James D MacBeth. 1973. "Risk, return, and equilibrium: Empirical tests." *Journal of political economy* 81(3):607–636.

Fama, Eugene F and Kenneth R French. 2000. "Forecasting profitability and earnings." *The journal of business* 73(2):161–175.

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

Fazzari, Steven, R Glenn Hubbard and Bruce C Petersen. 1987. "Financing constraints and corporate investment.".

Ferderer, J Peter. 1993. "The impact of uncertainty on aggregate investment spending: an empirical analysis." *Journal of money, credit and banking* 25(1):30–48.

Filipovic, Zoran and Alexander F Wager. 2019. "The intangibles song in takeover announcements: Good tempo, Hollow Tune." *Social Science Research Network.* "https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3330623. .

Fisher, Anthony C, W Michael Hanemann, Michael J Roberts and Wolfram Schlenker. 2012. "The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather: comment." *American Economic Review* 102(7):3749–3760.

Garcia, Diego. 2013. "Sentiment during recessions." *Journal of Finance* 68(3):1267–1300.

Garcia, Diego and Øyvind Norli. 2012. "Geographic dispersion and stock returns." *Journal of Financial Economics* 106(3):547–565.

Giglio, Stefano, Bryan Kelly and Johannes Stroebel. 2021. "Climate finance." *Annual Review of Financial Economics* 13:15–36.

Gilchrist, Simon, Jae W Sim and Egon Zakrajšek. 2014. Uncertainty, financial frictions, and investment dynamics. Technical report National Bureau of Economic Research.

Goldsmith-Pinkham, Paul, Matthew T Gustafson, Ryan C Lewis and Michael Schwert. 2023. "Sea-level rise exposure and municipal bond yields." *The Review of Financial Studies* 36(11):4588–4635.

Graff-Zivin, J and M Neidell. 2010. "Temperature and the Allocation of Time: Implications for Climate Change. US National Bureau of Economic Research Working Paper No. 15717.".

Gulen, Huseyin and Mihai Ion. 2016. "Policy uncertainty and corporate investment." *The Review of Financial Studies* 29(3):523–564.

Hall, Richard. 2009. The strategic analysis of intangible resources. In *Knowledge and strategy*. Routledge, London, England, pp. 181–195.

Hong, Harrison, Frank Weikai Li and Jiangmin Xu. 2019. "Climate risks and market efficiency." *Journal of econometrics* 208(1):265–281.

Hsiang, Solomon M. 2010. "Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America." *Proceedings of the National Academy of sciences* 107(35):15367–15372.

Hsu, Po-hsuan, Kai Li and Chi-yang Tsou. 2023. "The pollution premium." *The Journal of Finance* 78(3):1343–1392.

Hulten, Charles R and Xiaohui Hao. 2008. What is a Company Really Worth? Intangible Capital and the" Market to Book Value" Puzzle. Technical report National Bureau of Economic Research, Cambridge MA.

Iqbal, Aneel, Shivaram Rajgopal, Anup Srivastava and Rong Zhao. 2023. "Value of internally generated intangible capital." *Social Science Research Network.* "https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3917998. .

Izhakian, Yehuda, David Yermack and Jaime F Zender. 2021. "Ambiguity and the tradeoff theory of capital structure." *Management Science* .

Jacobsen, Ben and Wessel Marquering. 2008. "Is it the weather?" *Journal of Banking & Finance* 32(4):526–540.

Jiang, Fuwei, Joshua Lee, Xiumin Martin and Guofu Zhou. 2019. "Manager sentiment and stock returns." *Journal of Financial Economics* 132(1):126–149.

Jones, Benjamin F and Benjamin A Olken. 2010. "Climate shocks and exports." *American Economic Review* 100(2):454–459.

Jurado, Kyle, Sydney C Ludvigson and Serena Ng. 2015. "Measuring uncertainty." *American Economic Review* 105(3):1177–1216.

Kaiser, Regina and Agustín Maravall. 1999. "Estimation of the business cycle: A modified Hodrick-Prescott filter." *Spanish Economic Review* 1(2):175–206.

Kessides, Ioannis N. 1990. "Market concentration, contestability, and sunk costs." *The Review of Economics and Statistics* pp. 614–622.

Kim, Hyunseob and Howard Kung. 2017. "The asset redeployability channel: How uncertainty affects corporate investment." *The Review of Financial Studies* 30(1):245–280.

Kozlowski, Julian, Laura Veldkamp and Venky Venkateswaran. 2020. "The tail that wags the economy: Beliefs and persistent stagnation." *Journal of Political Economy* 128(8):2839–2879.

Leahy, John V and Toni M Whited. 1996. "The effect of uncertainty on investment: Some stylized facts." *Journal of Money, Credit and Banking* 28(1):64–83.

Lev, Baruch. 2000. *Intangibles: Management, Measurement, and Reporting*. Brookings Institution Press, Washington, DC.

Lev, Baruch. 2005. "Intangible assets: Concepts and measurements. In: Kempf-Leonard, K. (Ed.), Encyclopedia of Social Management." 2.

Lev, Baruch. 2012. *Winning Investors Over: Surprising Truths about Honesty, Earnings Guidance, and Other Ways to Boost Your Stock Price*. Harvard Business Review Press, Boston, MA.

Lev, Baruch and Anup Srivastava. 2022. "Explaining the recent failure of value investing." *Critical Finance Review* 11(2):333–360.

Lev, Baruch and Theodore Sougiannis. 1996. "The capitalization, amortization, and value-relevance of R&D." *Journal of Accounting and Economics* 21(1):107–138.

Lo, Andrew W and A Craig MacKinlay. 1990. "When are contrarian profits due to stock market overreaction?" *The review of financial studies* 3(2):175–205.

Loughran, Tim and Bill McDonald. 2011. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." *Journal of Finance* 66(1):35–65.

Loughran, Tim, McDonald Bill. 2021. "Loughran and McDonald Sentiment Word Lists.".
URL: *https://sraf.nd.edu/textual-analysis/resources/*

Mandelbrot, Benoit. 1961. "On the theory of word frequencies and on related Markovian models of discourse. In:Structure of language and its mathematical aspects." *American Mathematical Society, Providence* 12:190–219.

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

McDonald, Robert and Daniel Siegel. 1986. "The value of waiting to invest." *The quarterly journal of economics* 101(4):707–727.

Merkley, Kenneth J. 2014. "Narrative disclosure and earnings performance: Evidence from R&D disclosures." *The Accounting Review* 89(2):725–757.

Merton, Robert C et al. 1987. "A simple model of capital market equilibrium with incomplete information.".

Newey, Whitney K and Kenneth D West. 1987. "Hypothesis testing with efficient method of moments estimation." *International Economic Review* pp. 777–787.

Newman, Mark EJ. 2005. "Power laws, Pareto distributions and Zipf's law." *Contemporary physics* 46(5):323–351.

NOAA. 2024. "Climate at a Glance Statewide Time Series.". `https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series`, Last accessed on 2024-02-22.

Park, Hyuna. 2019. "An intangible-adjusted book-to-market ratio still predicts stock returns." *Critical Finance Review* 25(1):207–236.

Pástor, L'uboš and Robert F Stambaugh. 2003. "Liquidity risk and expected stock returns." *Journal of Political economy* 111(3):642–685.

Pástor, L'uboš, Robert F Stambaugh and Lucian A Taylor. 2021. "Sustainable investing in equilibrium." *Journal of Financial Economics* 142(2):550–571.

Pástor, L'uboš, Robert F Stambaugh and Lucian A Taylor. 2022. "Dissecting green returns." *Journal of Financial Economics* 146(2):403–424.

Peters, Ryan H and Lucian A Taylor. 2017. "Intangible capital and the investment-q relation." *Journal of Financial Economics* 123(2):251–272.

Pindyck, Robert S and Andres Solimano. 1993. "Economic instability and aggregate investment." *NBER macroeconomics annual* 8:259–303.

Pontiff, Jeffrey. 1996. "Costly arbitrage: Evidence from closed-end funds." *Quarterly Journal of Economics* 111(4):1135–1151.

Schlenker, Wolfram and Michael J Roberts. 2009. "Nonlinear temperature effects indicate severe damages to US crop yields under climate change." *Proceedings of the National Academy of sciences* 106(37):15594–15598.

Seamons, Quinton F and Robert W Rouse. 1997. "MD&A: How to use safe harbor to report "soft" information." *Journal of Corporate Accounting and Finance* 9(1):81–90.

Smith, Tiffany T, Benjamin F Zaitchik and Julia M Gohlke. 2013. "Heat waves in the United States: definitions, patterns and trends." *Climatic change* 118:811–825.

Srivastava, Anup. 2014. "Why have measures of earnings quality changed over time?" *Journal of Accounting and Economics* 57(2-3):196–217.

Srivastava, Anup. 2023. "Trivialization of the bottom line and losing relevance of losses." *Review of Accounting Studies (forthcoming)* .

Stambaugh, Robert F, Jianfeng Yu and Yu Yuan. 2012. "The short of it: Investor sentiment and anomalies." *Journal of Financial Economics* 104(2):288–302.

Stambaugh, Robert F, Jianfeng Yu and Yu Yuan. 2015. "Arbitrage asymmetry and the idiosyncratic volatility puzzle." *Journal of Finance* 70(5):1903–1948.

Starr-McCluer, M. 2000. "The Effects of Weather on Retail Sales, Federal Reserve Board of Governors." *Dostupno na: http://www. federalreserve. gov/pubs/feds/2000/200008/200008pap. pdf* .

Tetlock, Paul C. 2007. "Giving content to investor sentiment: The role of media in the stock market." *Journal of Finance* 62(3):1139–1168.

Wu, Sijia, Ming Luo, Rui Zhao, Jing Li, Peng Sun, Zhen Liu, Xiaoyu Wang, Peng Wang and Hui Zhang. 2023. "Local mechanisms for global daytime, nighttime, and compound heatwaves." *npj Climate and Atmospheric Science* 6(1):36.

Wyatt, Anne. 2008. "What financial and non-financial information on intangibles is value-relevant? A review of the evidence." *Accounting and Business Research* 38(3):217–256.

Yoon, Donghyuck, Dong-Hyun Cha, Gil Lee, Changyong Park, Myong-In Lee and Ki-Hong Min. 2018. "Impacts of synoptic and local factors on heat wave events over southeastern region of Korea in 2015." *Journal of Geophysical Research: Atmospheres* 123(21):12–081.

Appendix

Excerpts from example Form 10-K filings containing intangibles terms:

"Our license and development arrangements with customers typically require significant customization of our ***intellectual property*** *components. As a result, we recognize the revenue from the* ***license*** *and the revenue from the development services as a single performance obligation over the period in which the development services are performed. We measure progress to completion based on actual cost incurred to date as a percentage of the estimated total cost required to complete each project. If a loss on an arrangement becomes probable during a period, we record a provision for such loss in that period." (NVIDIA CORPORATION, Form 10-K, January 26, 2020)*

"We purchase and roast high-quality whole bean coffees that we sell, along with handcrafted coffee and tea beverages and a variety of fresh food items, through company-operated stores. We also sell a variety of coffee and tea products and ***license*** *our* ***trademarks*** *through other channels such as licensed stores, grocery stores, and national food service accounts. In addition to our flagship Starbucks* ***brand****, our portfolio also includes Tazo® Tea, Seattle's Best Coffee®, and Starbucks VIA® Ready Brew." (Starbucks Corporation, Form 10-K, October 2, 2011)*

"At FedEx, it is our people—our greatest asset—that give us our strong ***reputation*** *and stand at the heart of our success. In addition to our superior physical and information* ***networks****, FedEx has an exemplary* ***human network****. Across the globe, our team members are united by our passion to deliver the FedEx Purple Promise—to make every FedEx experience outstanding—and our People–Service–Profit principles. " (FedEx Corporation , Form 10-K, May 31, 2022)*

"We have obtained ***patents*** *in the U.S. and other countries. Because of the fast pace of* ***innovation*** *and product development, and the comparative pace of governments' patenting processes, our products are often obsolete before the* ***patents*** *related to them expire; in some cases, our products may be obsolete before the* ***patents*** *related to them are granted. As we expand our products into new industries, we also seek to extend our patent development efforts to patent such products. " (INTEL Corporation, Form 10-K, December 26, 2015)*

"We also connect consumers with public transportation networks. We use this same ***network, technology****, operational excellence and product expertise to connect shippers with carriers in the freight industry. " (UBER TECHNOLOGIES, INC., Form 10-K, December 31, 2020)*