upf.

TESI DOCTORAL UPF / 2021

Juan Felipe Imbet Jiménez

Essays in Financial Economics

TESI DOCTORAL UPF / 2021

upf.

upf. Universitat Pompeu Fabra Barcelona

Essays in Financial Economics

Juan Felipe Imbet Jiménez

Essays in Financial Economics

Juan Felipe Imbet Jiménez

TESI DOCTORAL UPF / Year 2021

THESIS SUPERVISOR
Dr. Javier Gil-Bazo
Department Department of Economics and Business

Sometimes Science is More Art Than Science, Morty.

Rick Sanchez

A mi abuelo Jairo.

Acknowledgements

The outcome of this dissertation would have not been possible without the support of many people across two continents. I thank my advisor Javier Gil-Bazo for his support and guidance, which started back when I was pursuing my master's degree. He always went beyond his duties and responsibilities as an advisor and taught me valuable lessons that I will take with me for the rest of my life. I thank him for his constant encouragement to ask deep research questions, and for always reminding me that academic research is a service to society and therefore requires the highest code of ethics.

I thank Filippo Ippolito, Roberto Steri, and Winston Dou for their academic and personal support. Despite none of them being an official thesis coadvisor, they always showed an incredible interest and commitment to help me improve my research. I thank Filippo for serving as the academic bridge between UPF and the University of Pennsylvania which allowed me to visit the Wharton School in 2020, and for funding my last months in the doctorate. I thank Roberto Steri for introducing me to the world of structural estimation, and for all his help during the last years of my Ph.D. while coming up with a job market paper. I also thank him for help me finance my last years of Ph.D. through the University of Luxembourg. Finally, I thank Winston Dou for having sponsored my academic visit to the Wharton School at UPenn. I am thankful to Winston for constantly meeting me, even after the unfortunate COVID situation forced me to leave Philadelphia. My job market paper improved substantially after every discussion we had.

I thank my family, Pito, Mita, Alejandro, Locky and Ginebra for always encouraging me to pursue my dreams, and specially for their support in difficult times. Being away from my family has always been hard, but knowing that I could always count with them made everything orders of magnitude easier. I also thank Muñe, Marlon, Tita, Dani, Pau, Luli, and all my close relatives that supported me during these last seven years abroad.

I thank Inês Xavier, Kinga Tchórzewska, and Natalia Perry for being my second family in Barcelona, and specially to Camille DeFrancq and her family in France who have been a personal a professional lighthouse during these last years. I thank my closest friends Raffaele Manini, Menna El Hefnawy, Dagny Pawlak, Anna Porta, Karolis Liaudinskas, Gianmarco Ruzzier, Sandra Kaya, Thomas Woiczyk, Sampreet Goraya, Adil Ismailov, Ilja Kantorovich, and Chris Evans as well as the wonderful members of the PhD rock band White Noise: Derrick Kanngiesser, Niko Schoell, Shohei

Yamamoto, and Giaccomo Caracciolo for all the amazing gigs and time we played together.

I thank all the talented people I met during these last years in the PhD program, Angelo Gutierrez, Andrea Fabiani, Milena Djourelova, Sebastian Ellingsen, Josep Gisbert, Lukas Hoesch, Andre Souza, as well as to the faculty who supported me, Xavier Freixas, Eulalia Nualart, Andrea Polo, Mircea Epure, Javier Gomez, Dmitry Kuvshinov, Bjorn Richter, and Victoria Vanasco. I also thank all the marvelous and talented PhD students and Faculty I met at Wharton, specially Andreas Brogger and Alexander Kronies who were an incredible office companion. I am particularly thankful to the faculty at ESADE Business School for their support specially to Vicente Bermejo, Anna Bayona, and Carlo Sala. I thank the administrative team at UPF and the Barcelona GSE, specially Marta Araque, Laura Agustí, and Marta Ledesma for always helping me understand the bureaucratic sea of paperwork which is studying abroad. I thank my friends from the master's in finance program at the Barcelona GSE, Berenice Ramirez, Jaime Lopez, Jelena Skaric, Octavi Castells and Nuria Mata. I also thank my friends in Yerevan - Armenia, specially Sona Ghahramanyan.

Last but not least I thank all the students in the Barcelona GSE that I had the opportunity to be a TA for. I owe them my passion for teaching, and it fills me with joy that today despite the distance I can call some of them my friends.

Juan Felipe Imbet Jiménez, 2021

Abstract

This dissertation studies the role of uncertainty and information in financial markets, and its consequences for firms' and investors' capital allocation decisions. It contains three chapters. Chapter one revisits the relationship between policy uncertainty, investment and stock returns. I find that the uncertainty about future energy policies covaries positively with investment, aggregate consumption growth, and its innovations carry a negative price of risk. Chapter two investigates the use of social media in the asset management industry. The results suggest that managers use social media to persuade investors rather than to alleviate information asymmetries. Chapter three develops a model of information disclosure for a market of mutual funds in which fund managers strategically transmit qualitative information. I find that fund flows as a result, increase with the tone of the signal, while reputation and verification costs affect the probability of funds manipulating information.

Resum

Aquesta tesi doctoral estudia el paper de la incertesa i la informació en els mercats financers, i les seves consequéncies en les decisions d'inversió d'empreses i inversors. El primer capítol estudia la relació entre la incertesa política, la inversió i els retorns de les accions. La incertesa sobre la realització de polítiques energètiques futures es relaciona amb major inversió, major creixement de consum i les seves innovacions comporten un preu de risc negatiu. En el segon capítol s'investiga l'ús de les xarxes socials per part de gestores de capital. Els resultats suggereixen que els directius d'aquestes gestores poden utilitzar les xarxes socials per persuadir inversors en comptes d'utilitzar-les per alleujar problemes d'asimetria d'informació. El tercer capítol desenvolupa un model de divulgació d'informació en un mercat de fons d'inversió en el qual informació de caràcter qualitatiu pot ser divulgada estratègicament. Com a resultat, els costos de reputació i verificació afecten la probabilitat que els fons manipulin la informació que transmeten.

Contents

List of Figures

List of Tables

Introduction

This doctoral dissertation lies at the intersection between Asset Pricing, and Corporate Finance. It aims to improve our understanding of the role of uncertainty and information in financial markets and the consequences for both firms' and investors' capital allocation decisions. It contains three chapters: Chapter 1 studies the reaction of firms and financial markets to the uncertainty about future energy policies using a q-theory approach, and state of the art quantitative methods in asset pricing. Chapter 2 and Chapter 3 study both from a theoretical and empirical perspective how firms strategically disclose voluntary information, and its impact on investors' demand for financial assets.

This dissertation contains both a theoretical and empirical contribution to these fields. From a theoretical perspective, it introduces novel modelling techniques such as the disclosure of qualitative information in a mutual fund market (Chapter 3), or the use of energy as a factor of production in a corporate-based asset pricing model (Chapter 1). From an empirical perspective, this dissertation's contribution consists of collecting, processing and studying novel data to improve our knowledge on how information and uncertainty impact firms' and investors' capital allocation decisions. In Chapter 2 (joint with Javier Gil-Bazo) we create the first asset management database of social media communications and use machine learning (ML) algorithms to analyze their content contributing to a growing literature using ML in Finance (e.g. Gu et al. 2020; Bianchi et al. 2020; DeMiguel et al. 2021).

The dissertation also contributes to developing an energy related policy uncertainty index. This index is the first to explicitly use historical political data such as executive orders and public laws to measure policy uncertainty objectively and multi-dimensionally. It associates policy uncertainty with the difficulty of forecasting policy decisions, and stands apart from the existing text and news based approach to measuring policy uncertainty (Baker et al. 2016), which is highly aggregated and relies on a general perception of uncertainty rather than on the randomness of policy decisions. I use this index of policy uncertainty on Chapter 1 to test the empirical predictions of a q-theory model with capital and brown energy. In this chapter, I am the first to empirically document a positive relation between policy uncertainty and investment. In order to rationalize these findings, I propose and test that firms invest in energy efficient capital when the level of policy uncertainty is high in order to hedge against higher energy costs in the future, in particular in state of natures with high marginal utility. The main contribution of each chapter can be summarized as follows:

Chapter 1 revisits the relation between policy uncertainty, investment and stock returns. In particular I focus on a novel type of policy uncertainty, the uncertainty about future energy policies, which I refer to as energy policy uncertainty. This uncertainty - measured as the uncertainty about the U.S. President signing an energy related executive order in the future - covaries positively with investment and aggregate consumption growth, and its innovations carry a negative price of risk. In order to rationalize my findings, I propose and test a q-theory explanation in which firms can invest in energy-efficient capital in order to hedge against higher energy costs product of a tighter regulation in the future. Consistent with my model, as uncertainty increases, the average firm invests more, and the differences in investment between value and growth firms are amplified. As the benefits to invest increase, aggregate expected consumption growth decreases creating a predictable pattern in the stochastic discount factor and therefore in expected returns. I find that without explicitly including an investment factor in cross-sectional asset pricing regressions, policy uncertainty betas explain cross-sectional variations in stock returns across portfolios that differ in their growth opportunities.

In Chapter 2, joint with Javier Gil-Bazo, we study the voluntary transmission of information and how it impacts financial markets. More specifically we investigate the use of Twitter by asset management firms using a novel database of 1.4 million Twitter posts between 2009 and 2017 combined with machine learning (ML) algorithms. We find that larger and younger fund families use Twitter more intensively. Investors do not react to the amount of social media activity, but to the tone of the information posted. This relation is economically significant; a one standard deviation increase in the positiveness of a fund family's tweets in a given month increases its assets under management by 15 basis points or 11 million USD in the following month. We provide evidence suggesting that asset managers use social media to persuade investors rather than to alleviate information asymmetries, since the positive tone of tweets do not predict higher subsequent fund performance.

Chapter 3 presents a model for the market of mutual funds in which investors learn about managerial ability from two types of signals: a standard noisy signal from which investors learn following Bayes rule; and a qualitative signal from which investors learn following a more general Pseudo-Bayesian rule. Using recent developments in the axiomatic decision making literature, I embed the learning process into a portfolio selection program and study how capital allocation is affected by the presence of both type of signals. The model predicts that i) flows are increasing on the tone of the qualitative signal, ii) reputation costs decrease the probability of investors verifying information, and iii) verification costs and risk aversion decrease the probability of funds manipulating information.

Chapter 1

Stroke of a Pen: Investment and Stock Returns under Energy Policy Uncertainty

This paper shows novel evidence that Energy policy uncertainty - as measured by uncertainty about the U.S. President signing an energy related executive order in the future - covaries positively with corporate investment and aggregate consumption growth, and its innovations carry a negative price of risk. I propose and test a q-theory explanation in which firms invest in energy-efficient capital when facing energy policy uncertainty. This uncertainty amplifies differences in investment between growth and value companies as the benefits of substituting energy for capital increase with growth opportunities. As investment grows, aggregate current consumption decreases relative to future consumption, creating time varying expected variation in aggregate market returns and consumption growth. Without an investment factor, uncertainty betas explain cross-sectional variation in stock returns across portfolios that differ in their growth opportunities. However, since investment reacts to uncertainty endogenously, an asset pricing model that accounts for an investment factor absorbs the cross-sectional differences in expected returns explained by this policy uncertainty. My findings suggest that uncertainty about future energy policies in the last four decades can explain firms' adoption of energy-efficient capital.

1.1 Introduction

The impact of policy uncertainty on the real economy has been the subject of debate among academics and market participants in the last decade (Bloom 2009; Bloom et al. 2018; Baker et al. 2016). In April 2020 the level of policy uncertainty had a fivefold increase compared to 20 years ago, triggered by a global pandemic, U.S. political and demographic tensions, and a global trade war.[1] As policy decisions become harder to anticipate, firms' investment generally dampens and financial markets become more volatile (Bloom 2009; Kelly et al. 2016; Gulen and Ion 2015). While there is a growing body of literature on the impact of policy uncertainty into firms' total factor productivity (Pástor and Veronesi 2012, 2013), much less is known about how policy uncertainty affects the demand of non-capital factors of production such as energy.[2] Since the impact of policy uncertainty on energy demand depends theoretically on the level of risk aversion and substitutability in the economy (Stewart 1978), how this uncertainty affects firms' decisions and asset prices is ultimately an empirical question. This study is the first to examine how uncertainty about future energy policies impacts investment and stock returns. More specifically I address these questions: How does investment react to the uncertainty about future energy policies? Do investors require compensation for holding equity from firms exposed to this uncertainty? Does this uncertainty capture patterns in consumption and aggregate returns?[3]

In recent decades there has been an increase in the supply of brown energy sources (e.g. oil or coal), due to technological changes and OPEC countries failing to control an increasing supply of oil (Gilje et al. 2016; Dou et al. 2020b), while simultaneously experiencing worldwide environmental concerns that have increased the demand for cleaner energy sources and environmentally friendly companies (Pástor et al. 2019). Given the importance of energy in the economy, it is not uncommon for governments to

[1]The Economic Policy Uncertainty index of Baker et al. (2016), a standard measure of policy uncertainty shows that in April 2020 its level more than quintupled from 80 to 423 over two decades. `https://www.policyuncertainty.com/`

[2]See, for example, Riem (2016); Snowberg et al. (2007); Colak et al. (2017); Mattozzi (2008); Brogaard and Detzel (2015). The empirical evidence exploring the relation between uncertainty and investment is ambiguous, and its sign depends on the source of the uncertainty e.g. productivity vs growth-opportunities quality (Dou 2017). Uncertainty shocks to productivity might have a different effect as firms temporarily pause investment and hiring (Bloom 2009; Bai et al. 2011; Bloom et al. 2018)

[3]The idea that factor uncertainty can trigger capital investment goes back to Stewart (1978). Risk averse managers exploit the substitution between capital and non-capital factors, to dial up investment when the price of a non-capital factor is uncertain.

intervene when energy-related events jeopardize the economy.[4] Moreover, energy has become such an important point in the political agenda, that energy and environmental policies are as cyclical as the economic policies between political parties.[5]

In this study I use energy-related U.S. executive orders to measure energy policy uncertainty, U.S. executive orders are difficult to anticipate, making them a suitable tool to study the impact of policy uncertainty on firms' behavior and financial markets. This contrasts with public laws for example, which can be highly anticipated given the long process required for their approval and media coverage. Moreover, executive orders capture the way that the incumbent U.S. President manages the country on a daily basis. Also, executive orders provide the President with a tool to make unilateral policy decisions with minimal interference from either Congress or the courts just with the "stroke of a pen" (Palmer 2002).[6]

Formally, I define energy policy uncertainty as the conditional volatility in a rolling probability model that forecasts, from the viewpoint of an economic agent, the occurrence of an energy related executive order in the future. Based on a topic analysis, I model the economic agent's information set as consisting of oil, business cycle, and political information. However, for robustness, I show that the main results do not depend on how the information set is modelled. Using an almost complete information set as in Jurado et al. (2015) yields qualitative similar results in all

[4]In 2017 U.S. President Donald J. Trump signed an Executive Order to increase Arctic drilling by 2022. Despite a judge in Alaska ruling that the Executive Order was unconstitutional, by August 2020 Trump's administration finalized the plan to open oil drilling in the Arctic `https://www.nytimes.com/2020/08/17/climate/alaska-oil-drilling-anwr.html`

[5]As an example, in 2010 U.S. President Barack Obama signed Executive Order 13543 creating the National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling suggesting new regulations to mitigate the impact of offshore drilling. On the other hand, U.S. President Donald Trump signed in 2019 Executive Order 13868 promoting the Energy Infrastructure and Economic Growth by facilitating Oil and Gas pipelines. Time varying energy and/or environmental policies can be the result of a political cycle induced by time-varying risk aversion (Pástor and Veronesi 2017) as well as an environmental concern in the spirit of Pástor et al. (2019) similar to the inequality aversion modelled in Pástor and Veronesi (2018).

[6]Palmer (2002) states that the phrase stroke of a pen is defined by Safire's Political Dictionary as "by executive order; action that can be taken by a Chief Executive without legislative action.". Its use has been traced to a nineteenth-century poem *Wanted - A Man* by the American poet Edmund Clarence. "Give us a man of God's own mold, Born to marshal his fellow-men; One whose fame is not bought and sold At the stroke of a politician's pen; Give us the man of thousands ten, Fit to do as well as to plan; Give us a rallying-cry, and then, Abraham Lincoln, give us a Man!"

empirical specifications.

To rationalize these empirical findings, I extend a production-based asset pricing model to consider firms that require energy and capital to produce a good. In the model, to hedge against higher energy costs in bad times, managers substitute energy for energy-efficient capital. This behavior is larger across firms with higher marginal q (marginal benefit of investing), and therefore amplifies cross-sectional differences in investment between growth and value companies. Since investment correlates negatively with expected returns in the cross-section, uncertainty amplifies valuations between growth and value companies. Finally, given that incentives to invest increase with uncertainty, under reasonable preference assumptions, households substitute current for future consumption as expected returns decrease. I therefore formulate the following hypotheses.

Hypothesis 1: Cross-sectional differences in investment explained by firms' growth opportunities are amplified when uncertainty is high. To test this hypothesis, I run cross-sectional investment-Q regressions (Gala et al. 2019) in which I interact energy policy uncertainty with variables proxying for growth opportunities. Consistent with the hypothesis, across U.S. public firms, differences in investment between small and large companies, and companies with high and low average Q are amplified when uncertainty increases. In particular, a one standard deviation increase in the unconditional level of uncertainty covaries with an increase in the regression coefficient between investment and average Q of 15 percent, and 20 percent between investment and size. Moreover, this increment in uncertainty covaries with a 1.2 percent increase in quarterly corporate investment or 480 million USD for the average firm.

Hypothesis 2: If firms invest more when uncertainty is high, under reasonable assumptions on households' intertemporal preferences, current consumption is substituted with current investment. If firms' growth opportunities decrease over time, future expected consumption increases. This predictable pattern in marginal utility can be tested by forecasting aggregate returns and consumption (Papanikolaou 2011; Kogan and Papanikolaou 2014). Predictability regressions of the U.S. monthly compounded value weighted CRSP return on energy policy uncertainty, and control variables documented to capture expected return variation show a negative and significant relationship between energy policy uncertainty and aggregate expected returns for horizons up to one year. A one standard deviation increase in energy policy uncertainty from its unconditional mean, covariates with a one percent decrease in the monthly aggregate expected return. Moreover, this finding is not explained by time-varying risk aversion across the political cycle (Pástor and Veronesi 2017), nor the dynamics of oil prices (Jones and Kaul 1996). Similarly, I also investigate if energy policy uncertainty predicts consumption growth. Given the endogeneity between aggregate market returns and consumption growth in a consumption based asset pricing model (Lucas 1978; Rubinstein 1976), I follow a similar methodology to Harvey (1988) and simultaneously predict the aggregate market return as well as consumption growth. For horizons up to six years, energy policy uncertainty positively predicts consumption growth. A one standard deviation increase in the unconditional level of energy policy uncertainty captures an increase in consumption growth between 17 and 24 percent per year for horizons of up to 6 years.

Hypothesis 3: Given that investment differences are amplified with energy policy uncertainty, expected returns should vary across firms with different uncertainty betas. However, controlling for investment, differences in uncertainty exposure should not help explain differences in expected returns since *ceteris paribus*, investment negatively correlates with returns in the cross-section. To test this hypothesis, I run cross-sectional linear asset pricing regressions in which one of the factors consists of the innovations to energy policy uncertainty. Following Maio and Santa-Clara (2012) in an intertemporal capital asset pricing model (ICAPM) (Merton 1973) framework, I extend common asset pricing models with innovations to energy policy uncertainty. Since the cross-sectional differences in investment are captured across firms' growth opportunities, I use the 25 size and book-to-market testing portfolios in Fama and French (1992, 1993). Asset pricing models that do not consider an investment factor yield negative and significant prices of risk for the uncertainty. However, in the presence of an investment factor the magnitude of the price of risk decreases and even becomes insignificant. Fama and French's five factor model (Fama and French 2015) extended with the innovations to energy policy uncertainty significantly reduces the magnitude of the price of risk, while using the q^4 and q^5 model of Hou et al. (2014) and Hou et al. (2020) completely absorbs the uncertainty price of risk.

I perform an extensive robustness analysis to ensure that the methodology used to estimate energy policy uncertainty does not drive my main results. In particular, I use a quasi-natural experiment to examine how energy policy uncertainty betas change between energy and non-energy sensitive companies: the OPEC announcement in November 2014 to not cut oil supply despite the increasing supply of oil from non-OPEC countries as studied in Dou et al. (2020b). The difference-in-differences estimator of the uncertainty beta between oil and non-oil related companies, shows that after the announcement, the energy policy uncertainty beta of oil related companies increased by 100%. Additionally, using lobbying data available since the 1990s, I show that firms in energy related sectors with higher lobby expenditures have lower energy policy uncertainty betas. This provides indirect evidence on the risk management benefits of lobbying by energy-policy sensitive companies. Finally, to ensure my results are not driven by my choice of the information set, I follow Jurado et al. (2015) and use a data-rich methodology to estimate energy policy uncertainty to show that the main results of the paper are robust to a more general specification of the economic agent's information set.

My study provides an investment-based explanation for a series of recent examined empirical patterns regarding climate risk and financial markets. Bolton and Kacperczyk (2021) find that stocks from companies with higher CO2 emissions earn higher expected returns that are not explained by their exposure to common factors, suggesting that investors are already considering compensation for carbon emissions risk. In my framework, companies exposed to CO2 risk are those with a lower degree of substitution between energy and capital. As these companies invest less to hedge against future volatility on energy costs, all things equal, they generate higher expected returns. Pástor et al. (2019) develop a demand-side model in which environmentally friendly stocks under-perform brown stocks. Equivalently, since in my framework these companies invest more relative to brown companies in order to hedge energy risks in the future, they earn lower expected returns. Finally, this study provides an indirect and involuntary mechanism in which government decisions impact the adoption and accumulation of energy-efficient capital.

The remainder of this paper is organized as follows, section (1.2) reviews the literature. Section (1.3) presents a stylised model of corporate investment used to develop the main testable hypotheses. The data sources and variable construction used herein and through the robustness tests are described in section (2.2). Section (1.5) builds the main measure of energy policy uncertainty. Section (1.6) empirically investigates how energy policy uncertainty amplifies cross-sectional differences in investment. Section (1.7) studies empirically the predictability power of energy policy uncertainty into aggregate market returns and consumption growth. The market reaction to energy policy uncertainty is studied in section (1.8) while section (1.9) presents all robustness tests. Finally, concluding remarks are provided in section (1.10).

1.2 Literature Review

I contribute to the literature on the relation between political (policy) uncertainty and asset prices (Pástor and Veronesi 2012, 2013; Kelly et al. 2016; Füss and Bechtel 2008; Mattozzi 2008; Bialkowski et al. 2008; Brogaard and Detzel 2015; Döpke and Pierdzioch 2006).[7] My paper is closest to Brogaard and Detzel (2015) who show that innovations to the News Based Economic Policy Uncertainty index of Baker et al. (2016) earn a negative price of risk. Different to theirs, I study the mechanism driving these results by studying how firms react to uncertainty. Moreover, the negative price of risk found by the authors in the cross-section of expected stock returns is not consistent with the negative impact that policy uncertainty has on corporate investment (e.g. Gulen and Ion 2015).

I also contribute to the literature by proposing a new proxy for policy uncertainty. There is a growing literature studying the relation among policy uncertainty, financial markets, and investment. However, policy uncertainty is unobservable to the researcher. Because of this, it has been studied indirectly by either looking at periods that are known to have high policy uncertainty (Kelly et al. 2016), or by using more general measures of uncertainty that indirectly capture policy uncertainty (Baker et al. 2016). To the best of my knowledge extant measures of policy uncertainty do not directly exploit political data which is nowadays widely available. Studies exploiting events such as elections to study how political uncertainty affect financial markets and corporate decisions include Kelly et al. (2016), Bialkowski et al. (2008), Colak et al. (2017), Füss and Bechtel (2008), Goodell and Vahamaa (2013), and Li and Born (2006). These studies have documented the pervasive effect that this uncertainty has on investment as well on making financial markets more volatile. However, the low frequency of these events only captures a small source of policy uncertainty, focusing exclusively on the uncertainty regarding structural changes that come after a change in the political party in power.[8]

Other studies rely on proxies available at higher frequencies allowing the study of financial markets and firms' behaviour as policy uncertainty evolves. Among these measures The News-Based Economic Policy Uncertainty Index (EPU) of Baker et al. (2016) has been highly used in academic research. The energy policy uncertainty index developed in this paper

[7] Other studies have focused on the reaction of firms to political (policy) uncertainty in the form of lobbying (Grotteria 2019)

[8] Other authors such as Mattozzi (2008) study the performance of portfolios expected to perform different depending on the result of the Bush vs Gore election in 2000 showing that a fraction of the political uncertainty during that period could have been hedged.

complements the EPU index of Baker et al. (2016) as it focus only on energy related matters, and its shown that impacts firms in a different way. In fact, from Figure (1.2) we see that both measures complement each other with a correlation of 0.18. Given that the information set used to compute the uncertainty depends on oil and political variables, my measure of uncertainty covariates with EPU only in moments of time in which these two variables are relevant, the gulf war, the financial crisis, and the recent increase in the supply of oil by non OPEC members.

I also contribute to the literature studying the relation between uncertainty and investment. Among the most important studies are (Pástor and Veronesi 2006, 2009; Bloom 2009; Bai et al. 2011; Bloom et al. 2018; Bachmann and Bayer 2014; Dou 2017; Dou et al. 2020a).[9] Similar to my findings, Dou (2017) studies the impact of two types of idiosyncratic uncertainties that affect assets in place and growth opportunities separately. He develops a general equilibrium model in which under poor risk sharing conditions that avoid the idiosyncratic volatility of the quality of growth options to be diversified, higher uncertainty increases the valuation of growth companies and increases investment in equilibrium. The mechanism studied in this paper is similar, with the main difference being that energy policy uncertainty is not diversifiable.

Since my paper studies how corporate investment and market valuations react to energy policy uncertainty, I also contribute to the literature in energy economics that studies the relation between the energy sector and energy-efficient investment in firms. A non exhaustive list of papers in this literature include Reuter et al. (2012); Hassett and Metcalf (1993); Barradale (2010); Chassot et al. (2014); Margolis and Kammen (1999). This literature studies regulation that encourages firms to invest in energy-efficient capital either by imposing carbon taxes, feed-in-tarifs, or tax incentives for energy related R&D. My main finding suggests that uncertainty regarding future energy policies has a similar impact on investment for firms, regardless the industry where they operate. Contrary to most studies in this literature who focus on firms in the utilities sector.

Additionally, I contribute to the literature that relates policy uncertainty with the macro-economy. Some of the most relevant studies include Karnizova and Li (2014); Demir et al. (2018); Li and Zhong (2020); Klößner and Sekkel (2014); Gulen and Ion (2015); Liu and Zhang (2015). Finally I contribute to the literature studying energy and environmental concerns such as climate risk into financial markets and firm's decisions. Papers in

[9]Other studies include Bai et al. (2011); Christiano et al. (2010, 2014); Basu and Bundick (2012)

this branch include Gilje et al. (2016); Jin and Jorion (2006); Chiang et al. (2015); Pástor et al. (2019); Dou et al. (2020b); Hong et al. (2019); Bolton and Kacperczyk (2021); Brøgger and Kronies (2020). Bolton and Kacperczyk (2021) show that companies with higher CO2 emissions earn higher expected returns, which they interpret as investors requiring a compensation for holding Climate Risk. Pástor et al. (2019) develop a general equilibrium model in which green companies, companies with a higher ESG score, earn lower returns in equilibrium. My framework provides an alternative supply-side interpretation of these results. Companies with higher CO2 emissions, or lower ESG scores, are companies that do not invest in energy-efficient capital. As this companies invest less in equilibrium, they earn higher expected returns.

1.3 A Stylised Model of Investment and Stock Returns under Factor Uncertainty

In this section I present a stylised dynamic model to study how corporate investment and expected returns vary in the presence of uncertainty about energy prices (factor uncertainty). The model builds on the Investment CAPM presented in Zhang (2017) extended to consider two inputs to the firms' production technology: capital and energy. The model preserves the classical characteristics of the neoclassical paradigm as it contains rational expectations, absence of market frictions, and firms maximize their equity value. The model is in partial equilibrium, firms take the pricing kernel as given, and the government acts exogenously by randomly setting energy prices. As a result, uncertainty regarding future energy prices, amplifies cross-sectional differences in investment and expected returns captured by the q theory of investment (Kaldor 1966; Tobin 1969; Hayashi 1982; Cochrane 1991; Liu et al. 2009; Zhang 2017).[10] The appendix provides all mathematical details.

Consider a two dates economy, t and $t+1$ with a continuum of heterogeneous firms indexed by $i \in [0,1]$. Firms produce an homogeneous good that requires capital K (e.g. Property, Plant and Equipment - PPE) and energy E (e.g. electricity) using a Cobb-Douglas technology $Y = K^{\alpha}E^{\beta}$ where $\alpha > 0, \beta > 0$, and $\alpha + \beta < 1$, after deciding optimally all other required inputs such as labor, intangible capital, or raw materials. This technology implies an inverse relation between capital and energy given

[10] This is contrary to models that explicitly model the government's optimization problem as in Pástor and Veronesi (2012, 2013)

output $E = (YK^{-\alpha})^{(1/\beta)}$, and equivalently a substitution between energy and capital $\partial E / \partial K < 0$.

This assumption is consistent with the evidence reported in the literature on energy economics. For instance, Tovar and Iglesias (2013), find that elasticity regressions between production costs and factors such as capital, energy, labor, and intermediate materials in the US yield negative estimates of cross-price elasticities between energy and capital, which suggests a systematic adoption of energy efficient technology in recent decades for U.S. firms.[11] The price of energy w_t, is randomly drawn from a stationary distribution with constant mean $\mathbb{E}[w_{t+1}] = \mu$ and variance $Var(w_{t+1}) = \sigma_e^2$. The volatility of energy price σ_e captures energy policy uncertainty in the model. Firm i starts period t with an amount of capital K_{it} and energy demand E_{it} to produce output Y_{it}. I assume that the firm's PPE configuration is not instantaneously adjustable, and since there are no shocks to the TFP of firms, capital and output are determined in advance. Firms face convex investment adjustment costs $a(I/K)K^2$ (e.g. Zhang 2005; Kogan and Papanikolaou 2012). where $a > 0$. All firms operate in both dates with a liquidation value of zero, and a depreciation rate of 100 %.

Firms take as given the stochastic discount factor (SDF) in the economy $M_{t,t+1}$. I assume that the price of energy w_{t+1} covaries positively with the SDF with a constant correlation such that $cov(M_{t,t+1}, w_{t+1}) = \rho\sigma_m\sigma_e > 0$, where $\sigma_m = \sqrt{Var(M_{t,t+1})}$. This assumption although restrictive, is consistent with empirical evidence: Edelstein and Kilian (2009) show that energy-price shocks have a negative impact on real consumption of unanticipated changes in discretionary income, shifts in precautionary savings, and changes in the operating cost of energy durables.[12]

Given current output Y_{it}, capital K_{it}, energy demand E_{it}, energy price per unit w_t, and the stochastic discount factor $M_{t,t+1}$, firm i chooses investment and future output to maximize shareholder's value which equals

[11] As agents dislike uncertainty regarding energy prices, induced innovation towards energy-efficient technology is more likely to occur (See Popp 2002). The discussion of whether capital and energy are substitutes or complements in firms' production functions is extensive with the literature on energy economics presenting mixed evidence. A common approach is to assume complementarity in the short run and substitution in the long run. See Haller and Hyland (2014) for a detailed discussion.

[12] Unreported monthly regressions of the natural logarithm of oil prices (1980m1-2019m12) and gas prices (1997m1-2019m12), report a positive correlation with the monthly NBER recession dummies, the probability of recession and the Sahm Rule estimated by the FRED at St Louis. This confirms the empirical evidence that energy prices increase in bad times.

current market price plus dividends

$$\begin{aligned} P_{it}+D_{it} = \max_{I_{it},Y_{i,t+1}} \Big\{ & Y_{it} - w_t(Y_{it}K_{it}^{-\alpha})^{(1/\beta)} - I_{it} - \frac{a}{2}\Big(\frac{I_{it}}{K_{it}}\Big)^2 K_{it} \\ & + \mathbb{E}\Big[M_{t+1}\Big(Y_{t+1} - w_{t+1}(Y_{t+1}K_{t+1}^{-\alpha})^{(1/\beta)}\Big)\Big]\Big\} \end{aligned} \tag{1.1}$$

The first order condition with respect to future output $Y_{i,t+1}$ is

$$I_{i,t} = Y_{i,t+1}^{\frac{1-\beta}{\alpha}}\Big(\frac{R_{ft}}{\beta}\mathbb{E}\Big[M_{t+1}w_{t+1}\Big]\Big)^{\frac{\beta}{\alpha}} \tag{1.2}$$

where $R_{ft} = \mathbb{E}[M_{t+1}]^{-1}$ is the gross risk free rate in the economy. Equivalently, the first order condition with respect to future capital $K_{i,t+1}$ is

$$1 + a\Big(\frac{I_{it}}{K_{it}}\Big) = \frac{\alpha}{\beta} I_{i,t}^{-\frac{\alpha+\beta}{\beta}} Y_{i,t+1}^{\frac{1}{\beta}}\mathbb{E}\Big[M_{t+1}w_{t+1}\Big] = q(Y_{i,t+1}, I_{it}) \tag{1.3}$$

Equation (1.2) shows that given optimal future output $Y_{i,t+1}$, investment increases with the present market value of the cost per energy $\mathbb{E}[M_{t,t+1}w_{t+1}]$, while Equation (1.3) states that in the optimum, firms invest up to the point in which the marginal investment cost equals marginal q.[13] Following Cochrane (1991) I can express the firm's first order conditions without the SDF, given the ex-dividend equity value $P_{it} = \mathbb{E}\Big[M_{t+1}\Big(Y_{t+1} - w_{t+1}(Y_{t+1}I_{i,t}^{-\alpha})^{\frac{1}{\beta}}\Big)\Big]$, and the gross return definition $R_{i,t+1} = \Big(P_{i,t+1} + D_{i,t+1})\Big)/P_{it}$ as follows

$$\mathbb{E}\Big[R_{i,t+1}\Big] = \frac{\alpha}{\beta}\frac{Y_{i,t+1}^{\frac{1-\beta}{\beta}} I_{i,t}^{-\frac{\alpha}{\beta}}\mu - 1}{\frac{I_{it}}{Y_{i,t+1}}\Big(1 + a\frac{I_{it}}{K_{it}}\Big) - \frac{\alpha}{\beta R_{ft}}} \tag{1.4}$$

Given equations (1.2), (1.3), and (1.4) I derive testable predictions to study how investment and stock returns relate with energy policy uncertainty. I start by deriving the standard results in any q theory model regarding investment and expected returns, and then show how these relations are amplified in the presence of factor uncertainty. The following proposition relates uncertainty with investment

[13] The assumption of a stochastic discount factor is required to study the asset pricing implications of uncertainty, but is not required to study investment. For instance Stewart (1978) shows that if risk averse managers receive utility for consuming a fraction of dividends, non-capital factor uncertainty increases investment when capital can substitute other factors in the production function as shown in the appendix.

Proposition 1. *Investment increases with uncertainty, and it increases more across firms with larger growth opportunities (firms that invest more)*

$$\frac{\partial I_{it}}{\partial \sigma_e} > 0 \text{ and } \frac{\partial^2 I_{it}}{\partial \sigma_e^2} > 0 \tag{1.5}$$

The predicted relation between investment and uncertainty can be observed in Figure (1.1). This relation is due to the fact that firms' marginal q is increasing with uncertainty, as the present value of energy costs $\mathbb{E}[M_{t,t+1}w_{t+1}]$ increases with σ_e since $\rho > 0$. Since the benefits of investment increase with energy price uncertainty, but marginal investment costs remain fixed, in equilibrium investment increases with uncertainty. Moreover, the relation between investment and uncertainty is convex. This implies that in the cross-section, growth firms, firms which invest to exploit growth opportunities, invest even more relative to value companies when uncertainty is high. This mechanism amplifies the cross-sectional differences explained by the q-theory components of the model.

The second hypothesis relates aggregate consumption and energy policy uncertainty. Even though my model is in partial equilibrium, and firms take the SDF as exogenous, I can derive predictions with respect to aggregate consumption in a setup in which a representative household owns the firms, and derives consumption out of the output of all firms. As in Papanikolaou (2011), I assume that households have preferences for later resolution of uncertainty.[14]

[14] Although not explicitly modeled, preference for early resolution of uncertainty is required in a general equilibrium set-up to relate an increase in investment with states of higher marginal utility. More specifically, in a continuous time economy with consumption and leisure in which households have Duffie and Epstein (1992) utility (or Epstein and Zin 1989 in discrete time). In this set-up households have preferences on consumption and leisure of the form $J_0 = \mathbb{E}_0 \int_0^\infty h(C_t, N_t, J_t)dt$ where $h(C, N, J) = \frac{\rho}{1-\theta^{-1}} \left\{ \frac{(C,N^\psi)^{1-\theta^{-1}}}{[(1-\psi)J]^{(\gamma-\theta^{-1})/(1-\gamma)}} - (1-\gamma)J \right\}$, ρ is a time preference parameter, γ is the coefficient of risk aversion, θ is the elasticity of intertemporal substitution, and ψ balances the relative shares of consumption and leisure. Under this parametrization, early resolution of uncertainty $\psi\theta < 1$ implies that leisure is a good $\psi(1-\theta^{-1}) < 0$. (Papanikolaou 2011)

Proposition 2. *Uncertainty on future energy prices σ_e positively predicts aggregate consumption growth*

$$\frac{\partial g_{t+1}}{\partial \sigma_e} > 0 \text{ where } g_{t+1} = \mathbb{E}[\frac{C_{t+1}}{C_t}] = \mathbb{E}[\frac{Y_{t+1}}{Y_t - I_t}] \tag{1.6}$$

where C, Y and I represent aggregate consumption, output and investment.

This proposition states that all things equal, an increase in uncertainty covaries with aggregate consumption growth. Moreover, this predictability on consumption growth gets translated into SDF predictability (Harvey 1988), and therefore firm returns are negatively predicted in the time-series.

Proposition 3. *Uncertainty negatively predicts returns in the time series*

$$\frac{\partial \mathbb{E}[R_{i,t+1}]}{\partial \sigma^e} = \frac{\partial \mathbb{E}[R_{i,t+1}]}{\partial I_{it}} \frac{\partial I_{it}}{\partial \sigma^e} < 0 \tag{1.7}$$

given the standard inverse relation between investment and expected returns in the q-theory of investment $\frac{\partial \mathbb{E}[R_{i,t+1}]}{\partial I_{it}} < 0$

Finally, I derive a prediction regarding the cross-section of expected returns similar to a CCAPM beta representation which relates firm characteristics to firm's beta

Proposition 4. *Uncertainty amplifies cross-sectional differences in expected returns captured by investment*

$$\mathbb{E}[R_{i,t}] - r_{ft} = \beta^S_{it} \lambda_t \tag{1.8}$$

where the firm's consumption beta is defined as

$$\beta^S_{it} = -\frac{\alpha}{\beta} \frac{\left(\frac{Y_{i,t+1}}{A}\right)^{\frac{1-\beta}{\beta}} I_{i,t}^{-\frac{\alpha}{\beta}}}{\frac{A I_{it}}{Y_{i,t+1}} \left(1 + a \frac{I_{it}}{K_{it}}\right) - \frac{\alpha}{\beta R_{ft}}} \rho \frac{\sigma^e}{\sigma^M} \tag{1.9}$$

and λ is the price of consumption risk $\lambda_t = R_{ft}(\sigma^M)^2$.

This proposition shows that differences in expected returns captured by the firm's consumption beta are amplified when energy policy uncertainty increases. Firms that invest more in equilibrium have a more negative betas, and earn lower expected returns. These differences in expected returns between companies with low and high investment, or equivalent growth and value companies, is amplified when energy policy uncertainty is higher. Finally, I show that since a firm's CCAPM beta is amplified by the same magnitude (σ_e), investment absorbs the cross-sectional variation across firms with different exposure to energy policy uncertainty

Proposition 5. *Given future output Y_{t+1}, investment explains the cross-sectional variation in expected returns as proposition (9) can be rewritten as*

$$\mathbb{E}[R_{i,t}] - r_{ft} = \beta^I_{it}\lambda^e_t$$

where the firm's investment beta is defined as

$$\beta^I_{it} = -\frac{\alpha}{\beta}\frac{\left(\frac{Y_{i,t+1}}{A}\right)^{\frac{1-\beta}{\beta}} I_{i,t}^{-\frac{\alpha}{\beta}}}{\frac{AI_{it}}{Y_{i,t+1}}\left(1 + a\frac{I_{it}}{K_{it}}\right) - \frac{\alpha}{\beta R_{ft}}}$$

and the price of risk takes into account both consumption risk and energy price uncertainty $\lambda^e_t = \rho\sigma_e\sigma_M R_{ft}$.

1.4 Data

1.4.1 Firm Accounting and Financial Data

Accounting and Financial data comes from CRSP and the Compustat Quarterly Database. From the quarterly Compustat database I download all firm-quarter observations up to 2019q4 (1,789,987). I keep only observations with ISO currency code in US dollars (curcdq), drop observations with missing assets (atq) or stockholders' equity (seqq) for a total of (1,235,343 observations). Accounting variables are defined as follows: Market equity is defined as the number of common shares outstanding (cshoq) times the calendar close price in the quarter (prccq). Size is the natural logarithm of market equity, book-to-market is the ratio of Stockholder's equity (seqq) to market equity. Book debt is defined as the sum of debt in current liabilities (dlcq) and long-term debt (dlttq). Profitability is the quarterly operating income after depreciation (oiadpq) over the sum of book debt and market equity. Leverage is defined as the book value of debt over the book value of debt plus market equity.

From the monthly CRSP database I download all observations from December 1961 up to December 2019 (4,230,439). I keep only companies trading in the NYSE AMEX or Nasdaq universe (exchcd 1, 2, and 3) with sharecodes (shrcd) equal to 10 or 11 for a total of 3,223,430 observations. I use the WRDS linking table between gvkey and permno identifiers to match the Compustat database with CRSP. I lag accounting variables by 2 months following Campbell et al. (2008) to correct for look-ahead bias. The merged database contains 2,355,017 firm-month observations. To compute market betas I download the Daily CRSP database between January 1965 and December 2019 (86,115,478 observations) and define market beta following Dimson (1979) computing intra-monthly regressions with respect to current, lagged, and led market returns as in Bali et al. (2019). Daily market excess return and daily risk free rates are obtained from Prof. Kenneth French's website.[15]

1.4.2 Financial and Macroeconomic Data

To construct the macro-finance data needed as a robustness test to model the investors' information set I rely on the methodology presented in Jurado et al. (2015) which I update until 2019. Jurado et al. (2015) present a Big Data methodology for uncertainty estimation using 147 macroeconomic and financial variables. Macroeconomic variables come from the FRED-MD database (McCracken and Ng 2015), financial variables are constructed following Jurado et al. (2015), and Ludvigson and Ng (2007, 2009). For details on the construction of the financial variables see the Appendix for Updates of Uncertainty Data available from Prof. Sydney Ludvigson website.[16] I replicate the construction of all the 147 time series except from the VXO index which is available from the FRED-MD database, and the Cochrane-Piazzesi factor (Cochrane and Piazzesi 2005) which I exclude given that it does not cover the same sample as the rest of the variables. Portfolio data for the construction of the variables comes from Prof. Kenneth French's website, dividend data comes from the monthly index CRSP database, and aggregate earnings data comes from Prof. Robert Shiller website.

The analysis in this paper relies heavily on oil and gas price data, as well as macroeconomic data commonly used in the predictability literature (Fama and French 1989, 1988; Stambaugh 1999; Campbell and Yogo 2006) such as the term and default spread, the aggregate dividend yield and the aggregate payout ratio. Oil prices correspond to the West Texas Intermediate standard price per barrel, gas prices correspond to the Henry

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

[16] https://www.sydneyludvigson.com/data-and-appendixes

Hub Natural Gas Spot price. Oil and gas information are obtained from the FRED at St. Louis.

The term spread is defined as the difference between the monthly average of the 10 year and 1 year risk free rate (DGS10-DGS1) and the default spread as the difference between the monthly average of the rate on BAA and AAA bonds obtained as well from the FRED at St. Louis. The dividend price ratio is the natural logarithm of the fraction of aggregate dividends inferred using the CRSP value weighted return with and without dividends (vwretd, vwretx) which is averaged across the last 12 months over an aggregate price index (See the Appendix of Jurado et al. 2015 for a detailed explanation). The aggregate dividend to earnings ratio is obtained from Prof. Robert Shiller's website.[17]

1.4.3 Political Data

The main political data used in this paper contains U.S. executive orders classified into 20 topics from the Comparative Agendas Project.[18] To ensure the consistency of the data I double check all executive orders in the database from public sources to ensure the database has no timing mistakes. First I check that the total number of executive orders available in the dataset from the Comparative Agendas project correspond to the total number of executive orders reported by official sources. To determine the true number of executive orders that were issued in a particular month I recollect data from the national archives and check for consistency.[19] Although prior to the first half of the 20th century presidents used executive orders in their mandates, these executive orders were not documented and archived until the 1940s. I am able to obtain 974 executive orders from 1937 to 2019, and count the number of executive orders issued each month which corresponds almost entirely with the dataset provided by the comparative agendas project.

The Comparative Agendas Project also contains data on public laws passed by the Senate and the House of Representatives which are used as control variables in the robustness test. The database contains a random subsample of all public laws starting in 1948. These public laws and executive orders are classified into 20 different policy topics based on the variable (pap_majortopic).

[17] http://www.econ.yale.edu/~shiller/

[18] https://www.comparativeagendas.net/

[19] https://www.archives.gov/federal-register/executive-orders/disposition

The dataset containing the subsample of public laws does not contain information about the month within each year in which the public law was issued before 1973. In order to determine the month in which the public law was issued I use two methodologies. First, I use the id of the public law provided in the database and web-scrape the information about the month from one of three sources. The library of congress contains information on all public laws issued since the period of George Washington, up to 1951.[20] Public laws from 1952 to 1973 are available from the Legis Work website.[21] Finally public laws from 1974 to nowadays are available from the US congress website.[22]

To determine the exact month in which the public law was issued the id of each law contains the number of the congress. Congresses have a number assigned since the first congress in 1789. Congresses from 1948 to 1951 correspond to numbers 80 to 81, congresses from 1952 to 1973 correspond to congresses 82 to 92 and congresses from 1974 to 2019 correspond to congresses 93 to 115. The Legis Work website organizes public laws into Volumes and not congress numbers. Congresses have mostly 2 volumes of laws which are normally split into half during the legislative mandate. Congresses from 1952 to 1973 correspond to volumes 65 to 86.

Once all public laws are downloaded from these websites the second step consists of assigning the law to the month and year when it was signed. From the library of congress this can be done by searching for the name of the month, year and date within the description of each law. For public laws from 1952 to 1973 obtained from Legis Work this is done by searching for sentences with words containing month names and obtain the year by looking at all words inside each sentence. Finally the congress website provides a more friendly format to obtain the date of each law.

The information of some of the public laws is not digitalized in these three sources. For these laws, I download the original text in image format. Using Optical Character Recognition (OCR) algorithms, I isolate the text of each law and using textual analysis isolate the part of the Public Law containing the year and month. I check that the year inferred from the OCR algorithm corresponds to the year provided by the original dataset for robustness of the OCR algorithm. I also collect information regarding the political party in power from the data-planet website. I collect data on the party of the president of the United States, as well as the political party

[20] `https://www.loc.gov/law/help/statutes-at-large`

[21] `http://legisworks.org/sal/`

[22] `https://www.congress.gov/public-laws/`

holding majority in the Senate and the House of Representatives.[23]

1.5 Measuring Energy Policy Uncertainty

In this section I construct a measure of energy policy uncertainty by fitting a probability model to estimate how likely it is to have an energy related executive order signed by a U.S. President in the future. Define by N_t the number of energy related executive orders signed in month t. The random variable defined as

$$Y_t = \begin{cases} 1 & \text{if } N_t > 0 \\ 0 & \text{if } N_t = 0 \end{cases} \tag{1.10}$$

has a conditional Bernoulli distribution with probability $Pr(Y_t = 1) = p_t$. The estimation of probability p_t is performed as follows: Given an information set I_{t-1} available for an economic agent at time $t-1$, and assuming the process Y_t is stationary, she estimates a probability model based on the history of realizations of $\{Y_s\}_{s=0}^{t}$.

$$\hat{p}_t = Prob(Y_t = 1|I_{t-1}) = f(I_{t-1}, \theta) \tag{1.11}$$

Where f is the functional form of a Probit model. The uncertainty about the value of variable Y_{t+1} before its realization can be seen as its conditional standard deviation

$$\mathcal{U}_t = \sqrt{Var(Y_{t+1}|I_t)} = \sqrt{\hat{p}_{t+1}(1-\hat{p}_{t+1})}. \tag{1.12}$$

This measure provides time varying uncertainty on $\mathcal{U}_t$ due to changes in the information set as well as new realizations of Y_t which updates the parameters in the underlying probability model $f(I_t, \theta)$. To model the information set required to fit the probability model I start with a base specification that includes the level and return of the West Texas Intermediate price per barrel, the aggregate dividend price ratio, and the Presidential Dummy of Santa-Clara and Valkanov (2003).

I show in robustness tests that the choice of the agents' information set does not qualitatively change the main results in the paper. However, there is evidence that the environmental and energy agenda of politicians differ between Republican and Democratic mandates (Gustafson et al. 2020), governments are more likely to change existing policies in bad times (Pástor and Veronesi 2012, 2013), oil and financial markets are strongly interdependent (Jin and Jorion 2006), and textual analysis of the text in executive

[23] https://data-planet.libguides.com/politicalpartycontrol

orders suggest oil behaviour triggers the occurrence of energy policies. As seen in Figure (1.3) during the 1970s and 1980s Oil is one of the topics most discussed within the text of executive orders given its importance and the consequences of the oil crisis.

Table (1.4) provides estimations of the probability model in which the left hand side variables is the probability of having at least one energy related executive order in the following 1, 3, 6, and 12 months. The return on oil, more than the level of the WTI predicts the occurrence of energy related executive orders for horizons of more than 3 months (Columns 2-4). The business cycle captured by the dividend price ratio predicts the occurrence for energy related executive orders for all specifications in an inverse u-shape depending on the horizon considered. Energy policies are more likely to occur after low market valuations, and finally they are more likely to occur under Democrat mandates.

The uncertainty used in the rest of the paper assumes a forecasting horizon of one month. In particular I fit a probit model as follows:

$$\hat{p}_t = \Phi(\hat{\beta}_0 + \hat{\beta}_1 \text{wti}_{t-1} + \hat{\beta}_2 R^{oil}_{t-1} + \hat{\beta}_3 (d-p)_{t-1} + \hat{\beta}_4 \text{Republican}_{t-1}) \quad (1.13)$$

where $\{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4\}$ are computed recursively using Maximum Likelihood Estimation based on information $\{Y_s, \text{wti}_s, R^{oil}_s, (d-p)_s, \text{Republican}_s\}_{s=0}^{t-1}$, wti is the West Texas Intermediate oil price per barrel relative to the price in 1970, R^{oil}_t is the return on wti_t between month $t-1$ and t, $(d-p)$ is the CRSP Value Weighted log dividend price ratio, and Republican is a dummy variable that takes the value of one if the U.S. President in power at month t has a Republican affiliation, and $\Phi(.)$ is the standard normal cdf.

The first estimation corresponds to January 1980 using information available since January 1970.[24] Each estimation is done recursively using all information available until month $t-1$. Figure (1.2) plots the evolution of energy policy uncertainty starting from January 1985 to December 2018 plotted against the EPU index of Baker et al. (2016). My index complements the aggregate EPU index by isolating uncertainty variation in energy related events. As a result my measure has a transitory spike during the Iraqi invasion of Kuwait in august 1990, then it increased in 1993 when the OPEC failed to agree to cut production decreasing consistently during the 2000s. During the financial crisis the energy policy uncertainty

[24]Oil prices before 1973 were highly regulated and do not exhibit significant time-series variation. I only use information since 1970 despite having executive orders starting in 1950 since most likely energy related decisions in the 50s and 60s were related to Nuclear Energy and Coal which probably not as relevant as they used to be.

increased with aggregate uncertainty, and declining in 2017 following a regularization in the Supply of oil by the OPEC. My measure of energy policy uncertainty strongly rejects the null of unit root with a z-statistic of -3.78 (p=0.003) under a standard Dickey and Fuller (1979) test, which decreases the likelihood of biasing the its coefficient in predictability and cross-sectional regressions due to correlation between regression residuals and innovations to energy policy uncertainty (Stambaugh 1999).

So far I have only described the time-series behaviour of energy policy uncertainty and its ability to capture underlying changes in the world energy supply. However, if this uncertainty is anticipated by the market, it should be incorporated into asset prices. In order to study the asset pricing consequences of this uncertainty I use the unexpected component of the uncertainty estimation in cross-sectional asset pricing regressions later on. I fit an AR(1) process into the conditional variance of random variable Y_t as follows

$$p_t(1 - p_t) = \phi_0 + \phi_1 p_{t-1}(1 - p_{t-1}) + u_t \tag{1.14}$$

after estimating ϕ_0 and ϕ_1 via OLS, and define $\hat{u}_t$ is the unexpected component of uncertainty. As in Dou et al. (2020b) and Bansal and Yaron (2004) I assume that the variances follow an AR(1) process. The OLS estimator of ϕ_1 is 0.96 significant at the 1 percent level, and rejects the null of unit root with a z-statistic of -22.8. The first sub-table in Table (1.1) presents summary statistics of the number of energy related executive orders per month N_t, the indicator variable Y_t, as well as summary statistics of the energy political uncertainty index $\mathcal{U}_t$ and its innovations u_t. In average, between January 1980 and December 2018 there were 0.1 executive orders signed per month, with executive orders occurring in average in 7 percent of the months in the sample.

1.6 Cross-sectional differences in investment under uncertainty

In this section I study how cross-sectional differences in investment across firms are amplified when energy policy uncertainty is high. The main result in the q-theory of investment states that marginal q is a sufficient statistic of firms' asset growth, as it captures firms' investment opportunities (Hayashi 1982). The first hypothesis developed in section (1.3) shows that energy policy uncertainty should amplify differences in investment explained by firms' marginal q. Substitution between energy and capital, or equivalently investment in energy-efficient technology should be profitable for all firms when energy policy uncertainty is high, but more profitable for firms with higher investment opportunities.

I begin by extending an otherwise standard investment-Q regression in which growth opportunities are proxied by size (Gala et al. 2019) and Tobin's average Q (Tobin 1969; Hayashi 1982; Cochrane 1991), and growth opportunities and profitability are sufficient statistics for investment.[25] In particular I estimate the following panel regression for public U.S. firms

$$\text{Inv}_{it} = a + b \times \text{Size}_{it} + c \times Q_{it} + d \times \text{Profitability}_{it} + \gamma \times \mathcal{U}_t + \epsilon_{it} \quad (1.15)$$

where investment (Inv) is defined as capital expenditures (the quarterly change of reported annual compustat item capxy for quarters 2, 3, and 4) over total assets. I use capital expenditures as proxying for energy-efficient capital investment following Brinkerink et al. (2019) who suggest that energy-efficiency is improved after capital expenditures and not by changes in fixed assets. I also estimate the above specification interacting growth opportunities (Size and Average Q) with my measure of energy policy uncertainty. If energy policy uncertainty triggers investment in capital, and this behaviour is more prominent for growth firms we should expect $b < 0$, $c > 0$, and the interaction with uncertainty amplifying the impact in the same direction.

Table (1.2) presents the results of estimating the above equation with and without industry fixed effects at the four SIC digits level to account for the fact that the impact that energy policy uncertainty can have on a firm's growth opportunities depends on the industry it operates. Columns 1 and 2 present the baseline result with industry fixed effects in which differences in investment are captured by profitability and growth opportunities and the level of energy political uncertainty. The baseline specification is consistent with Gala et al. (2019). Smaller firms and firms with higher average Q invest more. Both significant at the 1% level. Other things equal, a one standard deviation increase in firm's size translates into a decrease in average investment from 1.9 to 1.56 percent of total assets, for the average firm, or a 22% reduction in investment. Equivalently, a one standard deviation increase in average Q, translates into an increase in monthly investment for the average firm between 1.9 to 2.26 percent of total assets or an increase of 18 percent for the average firm. Moreover, this column confirms that aggregate investment increases when energy policy

[25]Extending Q regressions with ad-hoc variables has been widely used in the literature mainly to study the role of financial frictions in driving differences between marginal and average Q (Gomes 2001; Cooper and Ejarque 2003; Abel and Eberly 1994; Barnett and Sakellaris 1998; Bolton et al. 2011)

uncertainty is higher. A one standard deviation increase in the level of energy policy uncertainty from its unconditional mean covaries with a 1.2 percent increase in quarterly investment. For an average quarterly capex over assets of 1.9 percent and the average firm in the sample having 10.4 USD Billions in total assets, this increase in uncertainty results in quarterly investments of 480 million USD for the average firm.

Columns 2 and 3 present the estimation results of interacting energy policy uncertainty with the first measure of firms' growth opportunities, average Q. A one standard deviation increase in energy policy uncertainty from its unconditional mean between March 1980 and October 2018, increases the marginal relation between average Q and investment across all firms by 15 % (from 0.19 to 0.22) a magnitude that is significant at the 1% percent level. This amplification in the relation between average Q and investment also occurs across firms within the same industry at the 1 % level, for which the marginal relation increases by 13.5% (from 0.22 to 0.23).

In columns 4 and 5, I interact the second measure of firm's growth opportunities, firm size, with energy policy uncertainty in the cross-sectional investment regressions. Across all firms in the sample, an increase of a one standard deviation in energy policy uncertainty above its unconditional mean between March 1980 and October 2018, amplifies the marginal relation between size and investment by 21 % (from -0.23 to -0.28) both across all firms and within the same industry. Both results are significant at the 1 % level. These results confirm the hypothesis that an increase in the level of energy policy uncertainty, increases the incentives to invest by those firms with larger growth-opportunities - firms with larger average Q and smaller firms.

1.7 Energy Policy Uncertainty, Consumption, and Aggregate Returns

In this section I study the relation between the level of energy policy uncertainty, aggregate market returns and household consumption. If the incentives of growth firms to invest in capital are sufficiently high when energy policy uncertainty increases, one can expect an overall increase in aggregate investment as confirmed in Table (1.2). This increase in aggregate investment should decrease *ceteris paribus* aggregate consumption in the current period as more output is used for investment rather than consumption. If this impact is transitory, and consumption and investment patterns are expected to reverse in the future, this creates a forecastable pattern in consumption, and expected returns (Lucas 1978; Rubinstein 1976).

The idea that incentives to growth firms to invest in capital creates forecastable patterns in expected returns and consumption has been explored extensively in a recent literature (e.g. Papanikolaou 2011; Kogan and Papanikolaou 2013, 2014; Dou 2017). These papers suggest that these patterns in investment are caused by investment-specific shocks that decrease the per-unit cost of capital investment, as well as improving the quality of growth opportunities. Although, the mechanism that I test has a similar impact on aggregate investment, in my setup, firms invest more in capital when energy policy uncertainty is high, not because capital goods are cheaper, nor because managers expect higher returns to investment, but rather because the market value of the the cost incurred if not investing in energy-efficient capital is larger.[26]

To test this hypothesis I follow the literature on return predictability (e.g. Campbell and Yogo 2006; Stambaugh 1999; Fama and French 1988, 1989; Cochrane 2008) and fit the following time-series model

$$R_{t\to t+k} = a + \delta \mathcal{U}_t + \gamma X_t + \epsilon_{t\to t+k} \tag{1.16}$$

where $R_{t\to t+k}$ is the log cumulative return between month t and $t+k$ of the CRSP Value Weighted portfolio including dividends, $\mathcal{U}_t$ is the level of energy policy uncertainty at time t regarding the possibility of an energy executive order at time $t+1$, and X_t is a vector of variables documented to capture expected return variation such as the log dividend yield

[26] In unreported tests I find that energy policy uncertainty does not help explain cross-sectional differences in expected investment growth, which shades light on the transitory impact of the uncertainty on investment. Interacting energy policy uncertainty with average Q, operating cash flows, and changes in return on equity in similar investment growth regressions of Hou et al. (2020) yield no significant results.

(Cochrane 2008; Fama and French 1989), the term and default spreads (Fama and French 1988), and the return on oil prices (Jones and Kaul 1996; Ready 2017). I also control for time varying risk aversion proxied by the political party in power (Pástor and Veronesi 2017, 2018) by including a dummy variable for those months in which the U.S. president is a republican (Santa-Clara and Valkanov 2003).

According to one of the hypotheses presented in Section (1.3), if households have a preference for late resolution of uncertainty, current aggregate consumption decreases at time t as investment becomes more atractive. Since investment translates into future output, agents expect consumption to grow and therefore lower marginal utility in the future, which captures lower expected returns ($\delta < 0$) as in the CCAPM. Table (1.3) presents OLS estimates with Newey and West (1987) standard errors with k lags to account for residual autocorrelation due to overlapping returns. For horizons of one month, one quarter, and one year, energy policy uncertainty negatively predicts market expected returns, above and beyond the variability captured by the dividend yield, the term and default spreads, the presidential dummy and oil returns.

As a common finding in predictability regressions, the magnitude of the coefficient of most variables and R^2s in the regression increases with the horizon (Cochrane 2008). However the coefficient of energy policy uncertainty has a different behaviour and becomes not significant for horizons larger than one year. For horizons of up to one year, energy policy uncertainty negatively captures expected return variation with a significance of 1% accounting for residual autocorrelation. The magnitude of this predictability is also economically significant. A one standard deviation increase in the level of energy policy uncertainty translates into a decrease in expected returns of 1.3 % in one month, 0.9 % monthly within one quarter and 1% per month within one year.

The dividend price ratio remains a significant predictor of expected returns for all horizons consistent with Fama and French (1988) and Cochrane (2008). My results are also robust to the political party in power in the U.S. Results of regressions with horizons up to one year, show that the coefficient of the dummy capturing a Republican mandate, are negative and significant at the 1% level, which replicates the finding of Santa-Clara and Valkanov (2003) in our sample, democrat presidencies tend to have higher expected returns.[27]

[27] Pástor and Veronesi (2017) propose a rational explanation for the presidential puzzle (Santa-Clara and Valkanov 2003). The fact that under democrat U.S. presidential mandates stock returns are higher than under republican mandates, is a consequence that

Next I test if energy policy uncertainty captures changes in expected consumption growth. This is a more direct way to test the hypothesis which states that when energy policy uncertainty is high, aggregate investment increases at the expense of consumption. Testing for predictability on consumption growth is not trivial. In the consumption CAPM (Rubinstein 1976; Breeden and Litzenberger 1978; Lucas 1978; Breeden 1979), expected consumption growth is a function of expected aggregate returns, which are simultaneously a function of the model primitives. To overcome this problem I estimate simultaneously the following two equations via GMM as in Harvey (1988)

$$\begin{aligned} \ln\left(\frac{c_{t+k}}{c_t}\right) &= \gamma_0 + \gamma_1\mathcal{U}_t + \gamma_2 R_{t\to t+k} + \gamma_3\text{term}_t + v_{t\to t+k} \\ R_{t\to t+k} &= \delta_0 + \delta_1(d-p)_t + \delta_2\text{term}_t + \delta_3\text{def}_t + \epsilon_{t\to t+k} \end{aligned} \tag{1.17}$$

where c_t are the US monthly personal consumption expenses. I estimate the system of equations with similar moment conditions as for an OLS estimation, for parameters $\theta = \{\gamma_0, \gamma_1, \gamma_2, \delta_0, \delta_1, \delta_2, \delta_3, \delta_4\}$

$$g(\theta) = \frac{1}{T}\sum_t \begin{Bmatrix} \ln\left(\frac{c_{t+k}}{c_t}\right) - \gamma_0 - \gamma_1\mathcal{U}_t - \gamma_2 R_{t\to t+k} - \gamma_3\text{term}_t \\ R_{t\to t+k} - \delta_0 - \delta_1(d-p)_t - \delta_2\text{term}_t - \delta_3\text{def}_t \end{Bmatrix} \times Z_t = 0 \tag{1.18}$$

with instruments $Z_t = \{\text{Constant}, \mathcal{U}_t, R_{t\to t+k}, \text{term}_t, \text{def}_t, (d-p)_t\}$, and an initial identity matrix in the first stage of the GMM estimation.[28] Table (1.5) presents the results of the two step GMM estimation, for horizons of one, three, five and six years. I choose yearly horizons for two reasons: First it eliminates the seasonal component of consumption, and second it allows the regressions to have higher volatility in the left hand side since it is known that consumption patterns are smooth and consumption growth is not volatile enough to meet the Hansen and Jagannathan (1991) bound.[29]

As observed in columns 1 to 4, the coefficient γ_1 capturing the predictability power of energy political uncertainty on expected consumption growth is positive and significant for all horizons. Point estimates increase with the prediction horizon and is significant at the 1% level except for the

democrats are more likely to be elected when risk aversion is higher. In particular, if high energy uncertainty coincides with republican mandates where risk aversion is lower, and as a consequence expected returns.

[28]When using returns in GMM regressions it is common to obtain numerically singular covariance matrices in the first step of the GMM estimation, which can be solved by giving equal importance to every single moment condition (Cochrane 2009).

[29]See Bulusu and Gómez Biscarri (2012) and references inside for a discussion on the difficulty of using consumption data in testing the CCAPM.

one year horizon for which is significant at the 5% level. The magnitude of this finding is also economically significant. A one standard deviation increase in the level of energy policy uncertainty translates into an expected increases in yearly consumption growth of 17, 43, 23 and 36 percent for horizons of 1, 3, 5, and 6 years.

All together the predictability pattern in consumption and aggregate market returns supports the hypothesis that in periods of time with high energy policy uncertainty, aggregate consumption decreases relative to future consumption which gives room for a forecastable pattern in expected marginal utility that translates into return and consumption growth predictability. As shown in robustness tests, this finding survives a more complete specification of the information set when computing energy policy uncertainty.

1.8 Energy Policy Uncertainty and the Cross-section of Expected Stock Returns

In this section I show that innovations to energy policy uncertainty are priced in the cross-section of portfolio returns sorted on growth opportunities. Moreover, I show empirically that since investment reacts to the level of energy policy uncertainty, differences in investment should capture the cross-sectional variation explained by energy policy uncertainty betas. If energy policy uncertainty anticipates future states of low marginal utility, assets that appreciate in relative terms following an unanticipated shock should earn lower expected returns. Given that market valuations of growth companies appreciate with unexpected news about energy policy uncertainty (positive betas), the price of risk of this innovations should be negative. To test this hypothesis I follow a standard asset pricing approach in which I extend a linear asset pricing model with the innovations to energy policy uncertainty.

Given that my interest is to relate expected returns among firms with different investment opportunities across size (Gala et al. 2019) and average Q (Hou et al. 2014), I use as testing portfolios the 25 portfolios sorted on size and book-to-market (Fama and French 1992, 1993, 1995) which proxy for differences in investment opportunities. I follow the intertemporal capital asset pricing model (ICAPM) of Merton (1973) to relate energy policy uncertainty with state variables capturing changes in the investment opportunity set and perform GMM estimations of the price of risk following Maio and Santa-Clara (2012).[30] In particular, given a linear asset pricing model with factors f_{it}, I estimate the price of risk of energy policy uncertainty from the following expected return-covariance formulation

$$\mathbb{E}[R_j^e] = \gamma Cov(R_i, R_m) + \sum_i \gamma_i^f Cov(R_i, f_i) + \gamma_u Cov(R_i, u) \tag{1.19}$$

where R_j^e corresponds to the expected excess return of testing portfolio j, R_m corresponds to the market factor, u corresponds to the innovations on energy policy uncertainty, and γ_i^f is the covariance price of risk of factor i. As a robustness test to check if my results are consistent with an ICAPM explanation, I show that my results satisfy the restrictions in Maio and Santa-Clara (2012): The ICAPM predicts that the covariance price of market risk γ should be a feasible estimate of the representative agent's risk aversion, and since energy policy uncertainty negatively predicts expected returns it must follow that $\gamma_u < 0$.[31]

In the empirical analysis I extend some standard asset pricing models with innovations to energy policy uncertainty. I use the CAPM model of Sharpe (1964), Lintner (1965), and Mossin (1966), the 3 factor model (FF3F) of Fama and French (Fama and French 1992, 1993, 1996), the 5 factor model (FF5F) of Fama and French (2015) which includes both an investment and profitability factor, the q^4 model of Hou et al. (2014) which also relate expected returns with profitability and investment as FF5F but the factors construction differs, and finally, the q^5 model in Hou et al. (2014) and Hou et al. (2020).

[30] To avoid a fishing-license (Fama 1991) arising from using a data rich environment in asset pricing regressions, I follow Maio and Santa-Clara (2012) and test if energy political uncertainty satisfies 3 restrictions. First, for energy political uncertainty to capture changes in the investment opportunity set, it should predict the distribution of aggregate returns. In particular, the first or second moment of the return distribution. Second, innovations to energy political uncertainty must earn a significant price of risk in the cross-section of expected returns with the sign of the price of risk equal to the sign obtained in the predictability regressions. Third, the market price of covariance risk obtained from the cross-sectional regressions should be a reasonable estimate of the Relative Risk Aversion (RRA) of the representative agent in the economy. Finally, I check that the absolute value of the z-statistic in a GMM regression are close to the threshold in Harvey et al. (2016) in at least one of the specifications studied. Since the predictability power of energy political uncertainty was assured in the last section I focus on the estimation of the covariance price of risk as well as the estimates for the relative risk aversion coefficient.

[31] Maio and Santa-Clara (2012) point that in equilibrium the covariance price of market risk equals the Relative Risk Aversion coefficient of the representative agent $\gamma = -W\frac{V_{WW}}{V_W}$, where W is aggregate wealth, and V is the Value Function result of the representative investor's optimization problem.

I use the above mentioned factors to achieve two goals. Standard asset pricing models such as the CAPM and FF3F, allow me to test if investors price unexpected innovations to energy policy uncertainty and how they adjust their market valuations. This result is derived directly if investors consider energy policy uncertainty a relevant state variable as in the ICAPM. The second set of asset pricing models (FF5F, q^4, and q^5), which include factors related to investment and profitability, have a different use. They test if changes in the investment policy of firms after changes in energy policy uncertainty, is enough to explain cross-sectional differences across firms with lower and higher exposure to u. They allow me to test if differences in expected returns across companies with different u betas, can be explained because systematically some of these companies invest relatively more or less intensively. Naturally, if firms adjust their investment policy in the presence of uncertainty, magnitudes and z-statistics of the price of risk in u should be higher across the first set of asset pricing models. Across the CAPM, FF3F, FF5F asset pricing models, factors mktrf_t smb_t, hml_t, cma_t, rmw_t and mom_t which are factors related to the market, size, book-to-market, investment, and profitability, are obtained from Prof. Kenneth French website. On the other hand, across the q^4 and q^5 models, the R_{me}, $R_{I/A}$, R_{roe}, R_{eg}, factors related to market equity, investment-over-assets, return on equity and expected growth are obtained from Prof. Lu Zhang's website.

The above models can be estimated with the following $N + K$ moment conditions following Maio and Santa-Clara (2012)

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \begin{Bmatrix} (R_{it} - R_{ft}) - \sum_{k=1}^{K} \gamma_k (R_{it} - R_{ft})(f_{kt} - \mu_k) \\ (f_{kt} - \mu_k) \end{Bmatrix} = 0$$

Where N is the number of testing portfolios, K is the number of factors in the model, f_k corresponds to a factor in each specification, and μ_k is the unconditional average of the factor. The vector of parameters $\theta = (\{\gamma_k\}_{k=1}^{K}, \{\mu_k\}_{k=1}^{K})$ is then estimated using a one step GMM procedure (Hansen 1982) using an identity matrix as optimal weighting matrix. Following the original methodology by Maio and Santa-Clara, I add to the 25 testing portfolios the market return to merge the cross-sectional component of the ICAPM with the time-series aggregate risk-return trade-off.

Using equally weighted moments is equivalent to running an ordinary least squares (OLS) cross-sectional regression of average excess returns on factor covariances (right-hand side variables), however the GMM estimator accounts for residual correlation among testing assets. Moreover, this methodology allows me to account for estimation error in the factor means as in Cochrane (2009)[Chapter 13] and Yogo (2006).[32] [33]

Table (1.6) provides estimates of GMM cross-sectional regressions. Column 1 to 5 present the expected value and price of risk of energy policy uncertainty together with the factor means after the GMM estimation. Columns 1 to 2 present the base specifications extending the CAPM and FF3F models. The covariance price of risk of energy policy uncertainty is significant and negative across all five specifications, which suggest that portfolio with assets that perform better when energy policy uncertainty is unexpectedly high, are preferred by investors, and therefore earn lower expected returns. Finally, as shown in Columns 4 and 5, the price of risk earned by innovations to energy policy uncertainty is completely captured by the portfolio exposure to the q factors, as studied in Section (1.3) of the paper. The investment policy of firms reacts to energy policy uncertainty, and its a sufficient measure to explain the cross-sectional differences in expected returns. Finally, I report the Mean Absolute Error (MAE) in the cross-sectional regressions which equals the average of the absolute value of pricing errors. The q^4 and q^5 models yield the lowest MAE of all five specifications, with MAEs of 0.11 and 0.09 percent respectively. [34]

A potential concern that can arise is that z statistics in the cross-section analysis are not high enough to overcome potential biases caused by data-snooping and publication biases (Harvey et al. 2016, 2019). I tackle this potential concern in two ways. First I show in the robustness test, that this is caused by the conservative selection of variables in the information set used to estimate energy policy uncertainty. By considering a large battery of variables into this information set, some z statistics in the first set of asset pricing models overcome the threshold of 3 in Harvey et al. (2016) or remain slightly below while remaining significant. Second, the objective of the paper is to understand how energy policy uncertainty shapes firms' investment decisions, and how these decisions can be translated into differences in expected returns.

To ensure that portfolios of firms with larger growth opportunities are in fact earning higher expected returns when energy policy uncertainty

[32] As pointed by Maio and Santa-Clara (2012) this procedure is more convenient when estimating an asset pricing in expected return-covariance form instead of expected return-beta form Brennan et al. (2004)

[33] Recall that factor premia must be estimated jointly when the factor considered is not a portfolio, since the factor does not price itself in the cross-section, see (Cochrane 2009[Chapter 13])

[34] In unreported tests I extend the model of Stambaugh and Yuan (2016), which contains two anomaly factors, and find that the factor also digests the price of risk of energy political uncertainty. Given the fact that the object of interest in the paper are the investment opportunities of firms, I exclude it from the analysis.

is un expectedly large, I compute the portfolio betas using the model in column 1 of Table (1.6) but presented in a expected return beta form $R^e_{it} = a + b\text{mktrf}_t + \beta u_t + \epsilon_{it}$, and report them in Figure (1.4).[35] As expected, portfolios of companies with more growth opportunities, small companies and companies with lower book-to-market ratios, tend to have a larger beta than value companies. Portfolios mainly of small companies, and companies with low book to market ratios have larger and positive u betas.

1.9 Robustness Analysis

In this section I provide robustness analysis supporting the main findings in the paper. First, I provide more evidence that the measure of energy policy uncertainty developed, is in fact capturing uncertainty about future energy policies. Second, I show that qualitatively the main results of the paper, are robust to a more complete modelling of the investors' information set to compute energy policy uncertainty.

1.9.1 Robustness on the measure of Energy Policy Uncertainty

I study if companies that operate in businesses that are more sensitive to energy policies, such as companies whose cash-flows are energy-price sensitive, are also exposed to my measure of energy policy uncertainty. I define energy exposure following Jin and Jorion (2006) as the sensitivity of stock returns with respect to oil and gas returns controlling for aggregate market returns. In particular I use a 60 month rolling window for each company to fit regressions

$$R_{it} = a + bR_{mt} + \beta^{oil} R^{oil}_t + \epsilon_{it} \tag{1.20}$$

and

$$R_{it} = a + bR_{mt} + \beta^{gas} R^{gas}_t + \epsilon_{it} \tag{1.21}$$

together with an energy policy uncertainty beta.

$$R_{it} = a + bR_{mt} + \beta^{energy} u_t + \epsilon_{it} \tag{1.22}$$

and study if energy sensitive companies coincide with companies whose stocks are sensitive to uncertainty regarding energy policies. I run cross

[35] Results are similar when computing the betas based on any specification in which the price of risk of innovations to energy policy uncertainty are significant.

sectional regressions between oil and gas betas on β^{energy} and market leverage to account for equity risk as follows

$$\beta_{it}^{oil,gas} = \delta_0 + \delta_1 \beta_{it}^{energy} + \delta_2 \text{Leverage} + \delta_3 \beta_{it} + \epsilon_{it} \tag{1.23}$$

Where $\beta_{it}^{oil,gas}$ represents either the oil or gas beta, leverage is the market debt ratio of the firm, and β_{it} is the company equity's market beta. Table (1.7) presents results of estimating the above regression allowing for fixed variation at the time, industry, and firm level. Clustered standard errors at the month level are presented in parenthesis. Consistent with the nature of the uncertainty, companies sensitive to energy policy uncertainty coincide mostly with oil-sensitive companies while the relation to gas sensitive companies is weak, although coefficients in columns 4 and 5 are significant, they are economically insignificant. As seen in columns 2, and 3, this is not due to these companies operating in energy sensitive sectors. Results including the industry fixed effect suggest that within industry, companies more exposed to oil fluctuations are in fact companies more exposed to innovations on energy policy uncertainty. Additionally, the fact that the coefficient δ_1 survives the inclusion of firm fixed effects suggest that this relation is not firm specific.

1.9.2 A quasi-natural experiment, The 2014 OPEC Announcement

Oil and gas betas are noisy estimates of the reaction of firms' valuations to oil and gas prices. To better study the energy policy uncertainty betas of oil-sensitive companies I use a quasi-natural experiment recently used by (Dou et al. 2020b): the 2014 OPEC announcement to not cut the supply of oil. In November 2014 in the 166 OPEC Meeting leaded by Saudi-Arabia, the OPEC decided to not cut oil production despite the increasing supply from non-OPEC countries, which lead to a decrease of 10 percent in oil prices in one day, and persistent high volatility for the next years.[36] Given the persistent increase in the volatility of oil prices, energy-sensitive companies should experience an increase in their energy policy uncertainty beta. To estimate this impact, I run the following difference in differences

[36]I thank Winston Dou for referring me to his work with Leonid Kogan and Wei Wu, which allowed me to implement the quasi-natural experiment in my analysis.

regression

$$\beta_{it}^{energy} = a + b \times \text{Oil related dummy}_{it} + c \times \text{After OPEC announcement dummy}_t \\ + d \times \text{Oil related dummy}_{it} \times \text{After OPEC announcement dummy}_t + \epsilon_{it} \tag{1.24}$$

where Oil related dummy$_{it}$ is equal to one if the SIC code provided by Compustat equals to 1311, 1381, 1389, 2911, or 5172 as in (Chiang et al. 2015).
After OPEC announcement dummy$_t$ equals one if the current month $t \geqslant 2014m11$. The difference-in-differences estimator d captures how the energy policy uncertainty beta of oil-related companies changed after the announcement. Table (1.8) provides OLS estimates with double clustered standard errors at the year-month and firm (gvkey) level. I keep a symmetric estimation sample of 4 years before and after the announcement, and provide in column 1 the standard specification while in column 2 I include time fixed effects and remove the after OPEC announcement dummy. Estimator d is economically and statistically significant with a value of 0.84. Figure (1.10) plots the average β_{it}^{energy} for oil related companies in solid black line, and for non oil related companies in dashed line. The average energy policy uncertainty beta of oil related companies before the announcement is 0.66 and becomes 1.32 for the four years after the announcement. As expected, non oil related companies did not suffer a change in their energy policy uncertainty betas after their announcement.

1.9.3 Does lobbying decrease the exposure to energy policy uncertainty?

If firms have the ability to create political connections and lobby, energy-sensitive betas of companies that actively incur in lobbying should experience a systematic reduction in their exposure to energy policy uncertainty. In order to test this hypothesis. I use a conditional beta model similar to Jin and Jorion (2006), and model political uncertainty betas as a function of lobby expenditures as follows.

$$R_{it} = a_i + b_i R_{mt} + \left(\beta_i^{energy} + \gamma_i \frac{L_{it}}{A_{i,t-1}}\right) u_t + \epsilon_{it} \tag{1.25}$$

where L_{it} corresponds to lobby expenditures, and $A_{i,t-1}$ corresponds to firm's total assets in the period before. Lobby expenditures are queried via the LobbyView API using the Compustat gvkey of the firms in my sample, and firms with missing lobby expenditures are treated as zero. Lobbying expenses are winsorized at the one percent level.[37] If lobby expenditures of ex-ante exposed companies reduce the exposure to political uncertainty we would expect $\gamma_i < 0$ for energy sensitive companies. I estimate the above equation using the entire sample for which lobby expenditures are available between 1997 and 2018. I aggregate coefficients γ_i using a simple average scaling for the fraction of lobby expenditures per total assets within companies in the same industry

$$\gamma_j \bar{l}_j = \frac{1}{|\mathcal{I}_j|} \sum_{i \in \mathcal{I}_j} \gamma_i \bar{l}_i \tag{1.26}$$

where $\bar{l}_i$ corresponds to the average lobby expenditures over assets of firm i, and $\mathcal{I}_j$ is the set containing all firms in industry j. I aggregate using the 12 industries definitions in Prof. Kenneth French's website. Results of the impact of lobbying into systematic exposure to energy policy uncertainty as well as the zero lobby betas, the betas for companies with no lobby expenses, are presented in Figure (1.5). As expected, firms in sectors such as energy, durables, manufacture and health benefit from lobbying to decrease their exposure to energy policy uncertainty, with energy being the sector with the largest reduction in exposure to uncertainty given their lobby expenditures. Not surprisingly, the exposure of zero lobbying energy firms to uncertainty is the largest across these sectors. This does not only contribute to ensure my measure of energy policy uncertainty is in fact robust, but it also provides evidence that lobby is an effective risk management tool in the presence of policy uncertainty.

1.9.4 Robustness tests on the Information Set

I perform robustness analysis to ensure that the main results in the paper are not driven by the specification of the information set presented in Section (1.5). I model the information set following Jurado et al. (2015) and perform a data-rich forecasting exercise in which I forecast the existence of at least one energy related executive order in the future. The details of the forecasting procedure are presented in the appendix. I refer to this measure of energy policy uncertainty as the "complete measure".

I repeat the main three econometric specifications with this measure to test the hypotheses developed in Section (1.3). First I repeat aggregate

[37] I thank Marco Grotteria for sharing his code to perform the API requests from the LobyView website using the Compustat gvkeys.

return and consumption growth predictability regressions. Table (A.1) presents results over forecasting horizons of one month, one quarter, one year, and three years. For this specification I modify the control variables in two ways. First, I exclude the level of the West Texas Intermediate oil price, given that the construction of the complete measure uses data available since the 1950s, and oil prices were strongly regulated until de mid 1970s. Second, following Maio and Santa-Clara (2012) I include two more control variables, corresponding to the state variables whose innovations correspond to the factors smb and hml of Fama and French (1993) (see Maio and Santa-Clara 2012 for variable construction), since they have been documented to capture expected return variation for horizons starting in the 1960s, and finally Republican dummies over the majority in the Senate and the House of Representatives to account for differences in the political agenda in the legislative branch of power not considered before.

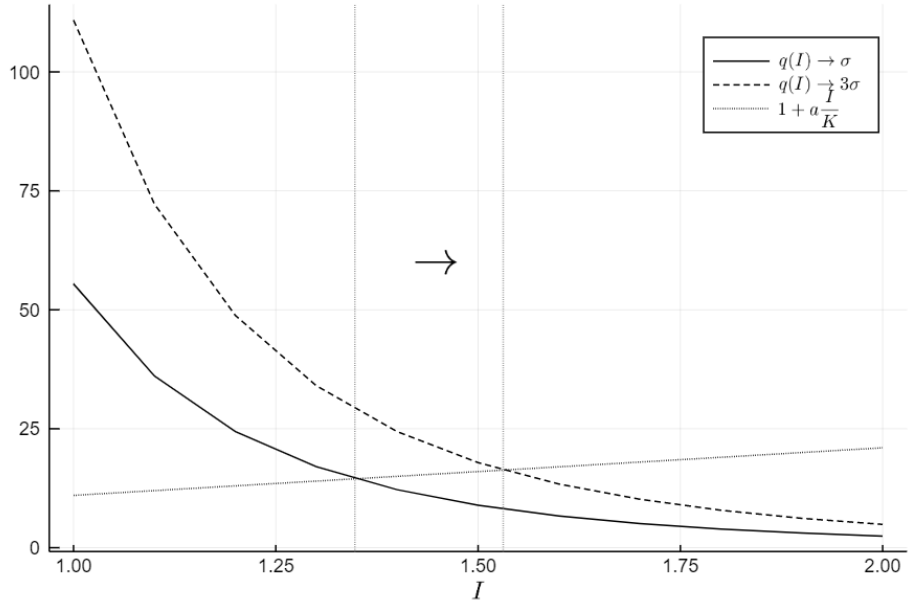
The complete measure of energy policy uncertainty negatively captures expected return variation for horizons between one quarter and three years. This result is stronger than the one presented in the main paper where predictability was only documented for horizons of up to one year.

Second, I study if this measure of uncertainty captures variation in expected consumption growth. Repeating GMM regressions presented in Section (1.7), Table (A.2) shows that the complete measure of energy policy uncertainty positively predicts expected consumption growth for horizons between one and six years. Thirdly, I study if its innovations are priced in the cross-section of expected returns following Section (1.8). Extending the five asset pricing models considered with innovations to the complete measure of energy policy uncertainty yield negative and significant prices of risk. Moreover, the price of risk for the first set of asset pricing models that exclude the investment factor, yield z statistics larger than 3 which decrease the likelihood of any data-snooping concerns in my analysis (Harvey et al. 2016). Finally, I study if differences in investment captured by a firm's growth opportunities are amplified when using this measure of uncertainty. Table (A.4) presents results of interacting investment-Q regressions with the firms average Q. I show that within each industry, the cross-sectional differences in investment are amplified when the complete measure of energy policy uncertainty is higher. These robustness tests ensure that my measure of energy policy uncertainty is capturing relevant state variables for energy-sensitive companies, and that results do not depend on the specification of the information set used to construct the uncertainty.

1.10 Conclusion

In this paper I show empirically that energy policy uncertainty measured as the blurriness in anticipating a U.S. President signing an energy-related executive order covaries positively with corporate investment, aggregate consumption growth, and its innovations carry a negative price of risk. I develop and test a q-theory explanation in which firm's invest in energy-efficient capital in anticipation of larger energy costs in bad times. This uncertainty amplifies cross-sectional differences in investment as the benefits of substituting energy for capital increase with growth opportunities. My results suggest that contrary to the pervasive consequences of policy uncertainty as a shock to the TFP of firms, energy policy uncertainty as it impacts the demand of a non-capital factor, has a positive impact on investment and asset prices as firms dial-up investment to hedge against future energy costs.

Figure 1.1: Theoretical relation between Energy Policy Uncertainty and Investment



Note: This graph shows theoretically how ceteris paribus investment increases when uncertainty σ_e increases. The model is solved assuming parameter values $R_f = 1.01$, $a = 10$, $K_{i,t} = 1$, $Y_{i,t+1} = 2$, $Y = 2K^{\alpha}E^{\beta}$ $\alpha = 0.7$, $\beta = 0.2$, $\mathbb{E}[w_{t+1}] = 0.5$, $\mathbb{E}[M_{t,t+1}] = 1/R_f$, $Var(M_{t,t+1}) = \mathbb{E}[M_{t,t+1}]$, $\sigma^2 = Var(w_{t+1}) = \mathbb{E}[M_{t,t+1}]$. Optimal investment corresponds to 1.35 and 1.53 respectively for σ and 3σ.

Figure 1.2: Energy policy uncertainty between 1985m1-2018m12, compared with the EPU index of Baker et al. (2016)

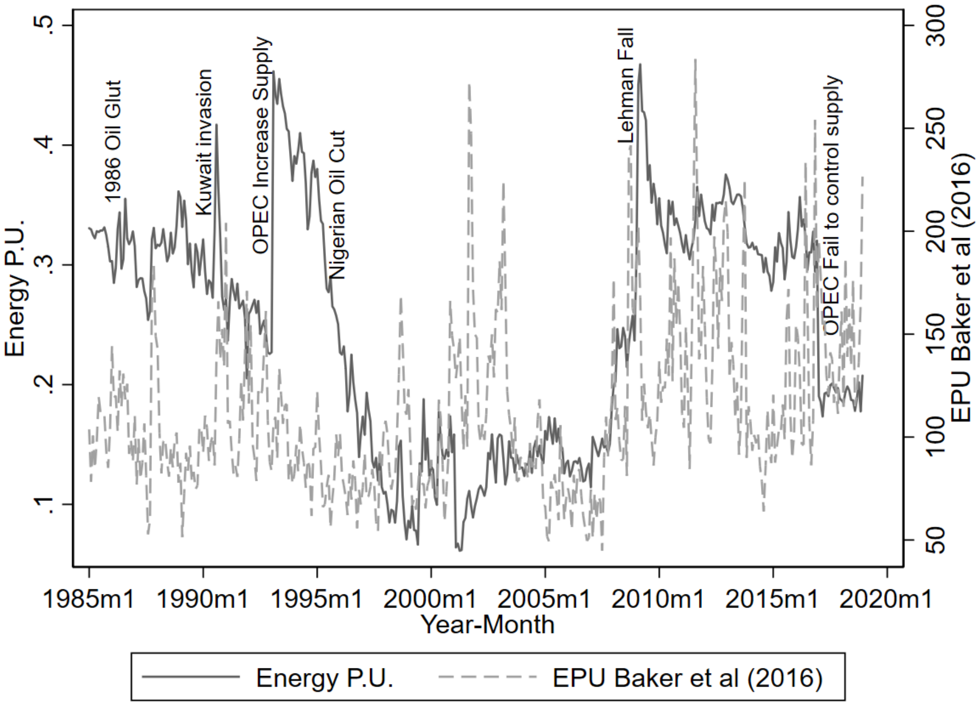


Figure 1.3: Number of Energy related U.S. Executive Order signed per year together with the most common topic inferred from its text

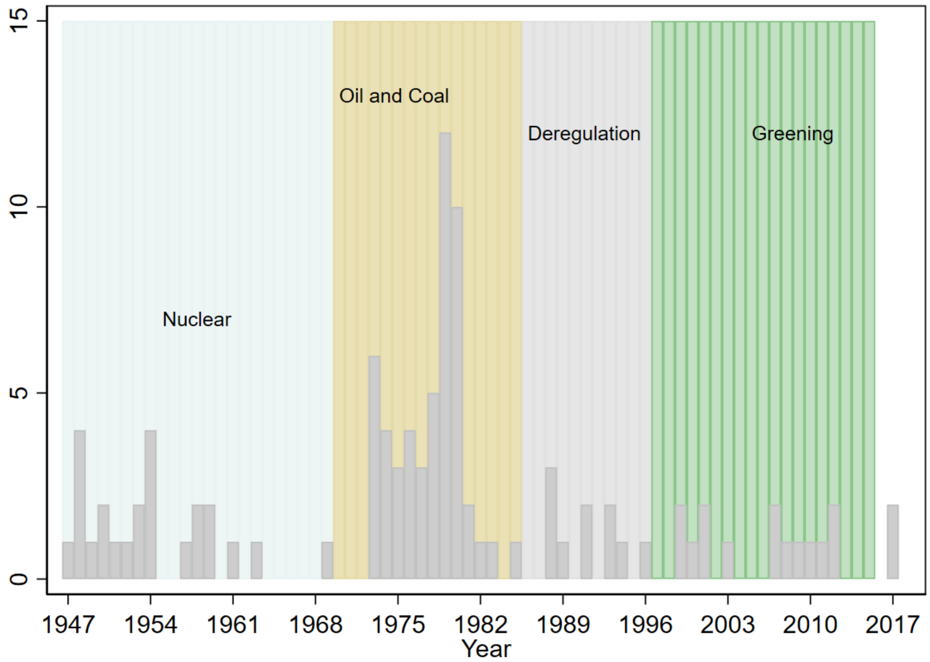


Table 1.1: Descriptive Statistics

	(1)				
	μ	σ	p1	p50	p99
N_t	0.1	0.3	0.0	0.0	2.0
Y_t	0.07	0.26	0.00	0.00	1.00
$\mathcal{U}_t$	0.36	0.05	0.26	0.35	0.50
u_t	0.00	0.02	-0.05	0.00	0.05
Observations	468				

Note: This table presents summary statistics regarding the total number of energy related executive orders signed by a US president N_t (pap_majortopic 8 in the Comparative Agendas database), and dummy variable Y_t that takes the value of 1 if $N_t > 0$ and 0 otherwise. The conditional volatility of variable Y_t captured by variable $\mathcal{U}_t$, and its innovations u_t defined as the residual of AR(1) process $\mathcal{U}_{t+1}^2 = \phi_0 + \phi_1 \mathcal{U}_t^2 + u_{t+1}$. Descriptive statistics are computed from January 1980 to December 2018.

Table 1.2: Investment Cross-sectional Regressions

$\text{Inv}_{it} = a + b \times \text{Profitability}_{it} + c \times Q_{it} + d \times \text{Size}_{it} + e \times \mathcal{U}_t + \epsilon_{it}$

	(1) Inv$_{it}$	(2) Inv$_{it}$	(3) Inv$_{it}$	(4) Inv$_{it}$	(5) Inv$_{it}$	(6) Inv$_{it}$
Profitability$_{it}$	4.90***	5.56***	5.03***	5.69***	4.76***	5.43***
	(0.77)	(0.72)	(0.77)	(0.72)	(0.77)	(0.72)
Q_{it}	0.18***	0.21***	0.12***	0.15***	0.18***	0.21***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Size$_{it}$	-0.23***	-0.24***	-0.23***	-0.24***	-0.11***	-0.11***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
$\mathcal{U}_t$	0.24*	0.29**	-0.34**	-0.32**	3.87***	4.14***
	(0.13)	(0.12)	(0.17)	(0.16)	(0.62)	(0.57)
$\mathcal{U}_t \times Q_{it}$			0.28***	0.30***		
			(0.08)	(0.07)		
$\mathcal{U}_t \times$ Size$_{it}$					-0.50***	-0.53***
					(0.08)	(0.07)
Constant	3.13***	3.10***	3.26***	3.22***	2.23***	2.14***
	(0.11)	(0.10)	(0.11)	(0.10)	(0.18)	(0.16)
Observations	575935	575935	575935	575935	575935	575935
Adjusted R^2 (%)	4.15	16.27	4.22	16.33	4.23	16.35
Industry F.E.	No	Yes	No	Yes	No	Yes
From	1981m1	1981m1	1981m1	1981m1	1981m1	1981m1
To	2018m10	2018m10	2018m10	2018m10	2018m10	2018m10

Note: This table presents results from monthly cross-sectional investment regressions. Quarterly accounting variables are merged with pricing data with a two month lag to account for look ahead bias. Investment is defined as the difference between the cumulative quarterly capital expenditures (capxy) between quarters n and $n-1$ for $n > 1$ divided over total assets. Profitability is measured as operating income after depreciation (oiadpq) over the sum of book debt (dlcq+dlttq) and market equity (prccq $\times$ cshoq). Average Q is computed as the book value of debt plus equity (dlcq+dlttq+prc$\times$ cshoq) divided by total assets (atq), size is the natural logarithm of market equity (prc $\times$ cshoq), $\mathcal{U}_t$ corresponds to energy policy uncertainty. Standard errors clustered by gvkey reported in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 1.3: Return predictability regressions
$R_{t\to t+k} = a + \delta\mathcal{U}_t + \gamma X_t + \epsilon_{t\to t+k}$

	(1) $R_{t\to t+1}$	(2) $R_{t\to t+3}$	(3) $R_{t\to t+12}$	(4) $R_{t\to t+36}$
$\mathcal{U}_t$	-0.12***	-0.27***	-1.09***	-0.69
	(0.04)	(0.06)	(0.32)	(0.62)
$(d-p)_t$	0.05***	0.11***	0.45***	0.68***
	(0.01)	(0.02)	(0.11)	(0.20)
term_t	0.00	0.00	0.03***	0.09***
	(0.00)	(0.00)	(0.01)	(0.03)
def_t	-0.00	-0.01	-0.01	-0.06
	(0.01)	(0.01)	(0.04)	(0.06)
Republican President_t	-0.02***	-0.05***	-0.18***	-0.23**
	(0.01)	(0.01)	(0.06)	(0.11)
R_t^{oil}	0.02	0.09**	0.06	0.05
	(0.03)	(0.05)	(0.11)	(0.15)
Constant	0.25***	0.53***	2.14***	3.10***
	(0.06)	(0.10)	(0.52)	(0.97)
Observations	468	468	468	444
Adjusted R^2 (%)	1.37	3.12	16.02	44.96

Note: This table presents results from the return predictability monthly regressions using energy policy uncertainty. $(d-p)_t$ is the natural logarithm of the aggregate VWCRSP dividend to price ratio, term_t is the term structure of interest rates defined as the difference between the 10 year and the 1 year risk free rate, def_t is the spread between BAA and AAA corporate bonds, R_t^{oil} is the monthly return of the West Texas Intermediate price per barrel. Republican President_t is a dummy variable that takes the value of 1 if the U.S. President at time t is Republican. Newey West standard errors for k lags reported in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Estimation sample is 1980m1 to 2018m12.

Table 1.4: Probability Model regressions

	(1) $\sum_{s=t}^{t+1} Y_s > 0$	(2) $\sum_{s=t}^{t+3} Y_s > 0$	(3) $\sum_{s=t}^{t+6} Y_s > 0$	(4) $\sum_{s=t}^{t+12} Y_s > 0$
wti_t	-0.00	-0.01	-0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.03)
R_t^{oil}	0.53	1.39*	1.27*	1.59**
	(0.87)	(0.72)	(0.67)	(0.69)
$(d-p)_t$	0.98***	1.00***	0.90***	0.66***
	(0.19)	(0.16)	(0.15)	(0.15)
Republican President_t	-0.39***	-0.40***	-0.42***	-0.34***
	(0.15)	(0.12)	(0.11)	(0.11)
Constant	2.53***	3.25***	3.32***	2.86***
	(0.77)	(0.63)	(0.60)	(0.60)
Observations	588	588	588	588
Pseudo R^2 (%)	8.08	7.63	6.11	3.96
Sample	1970m1-2018m12	1970m1-2018m12	1970m1-2018m12	1970m1-2018m12

Note: This table presents results from fitting probability models on the existence of one energy related executive order in the next month, quarter, semester and year. Probability estimation is performed using maximum-likelihood on a Probit model. wti_t and R_t^{oil} correspond to the level and return of the West Texas Intermediate oil price. $(d-p)_t$ is the log dividend price ratio, and Republican President_t is a dummy variable that takes the value of one if the U.S. President at month t has a Republican affiliation, and zero otherwise. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Estimation sample of the predictability regression is 1970m1 to 2018m12.

Table 1.5: Consumption growth regressions

$$g(\theta) = \frac{1}{T}\sum_t \begin{Bmatrix} \ln\left(\frac{c_{t+k}}{c_t}\right) - \gamma_0 - \gamma_1 \mathcal{U}_{t\to t+1} - \gamma_2 R_{t\to t+k} - \gamma_3 \text{term}_t \\ R_{t\to t+k} - \delta_0 - \delta_1 (d-p)_t - \delta_2 \text{term}_t - \delta_3 \text{def}_t \end{Bmatrix} \times \text{Instruments}_t = 0$$

	(1) $\ln\left(\frac{c_{t\to t+12}}{c_t}\right), R_{t\to t+12}$	(2) $\ln\left(\frac{c_{t\to t+36}}{c_t}\right), R_{t\to t+36}$	(3) $\ln\left(\frac{c_{t\to t+60}}{c_t}\right), R_{t\to t+60}$	(4) $\ln\left(\frac{c_{t\to t+72}}{c_t}\right), R_{t\to t+72}$
γ_0	4.73***	12.17***	20.86***	22.96***
	(0.21)	(0.60)	(0.99)	(1.14)
γ_1	1.78**	12.91***	11.95***	21.68***
	(0.79)	(2.16)	(4.56)	(5.87)
γ_2	-0.42***	-1.66***	-2.48***	-2.47***
	(0.07)	(0.16)	(0.23)	(0.26)
γ_3	0.07***	0.08***	0.11***	0.10***
	(0.01)	(0.01)	(0.01)	(0.02)
δ_0	229.14***	371.79***	423.46***	441.67***
	(17.62)	(18.95)	(20.33)	(16.19)
δ_1	47.18***	84.94***	95.50***	97.42***
	(3.58)	(4.11)	(4.52)	(3.66)
δ_2	-0.85	10.30***	10.03***	8.34***
	(0.82)	(0.91)	(0.85)	(0.84)
δ_3	-37.07***	-33.83***	-25.64***	-26.08***
	(5.37)	(4.09)	(3.63)	(2.93)
Observations	456	432	408	396
Sample	1980m1-2017m12	1980m1-2015m12	1980m1-2013m12	1980m1-2012m12

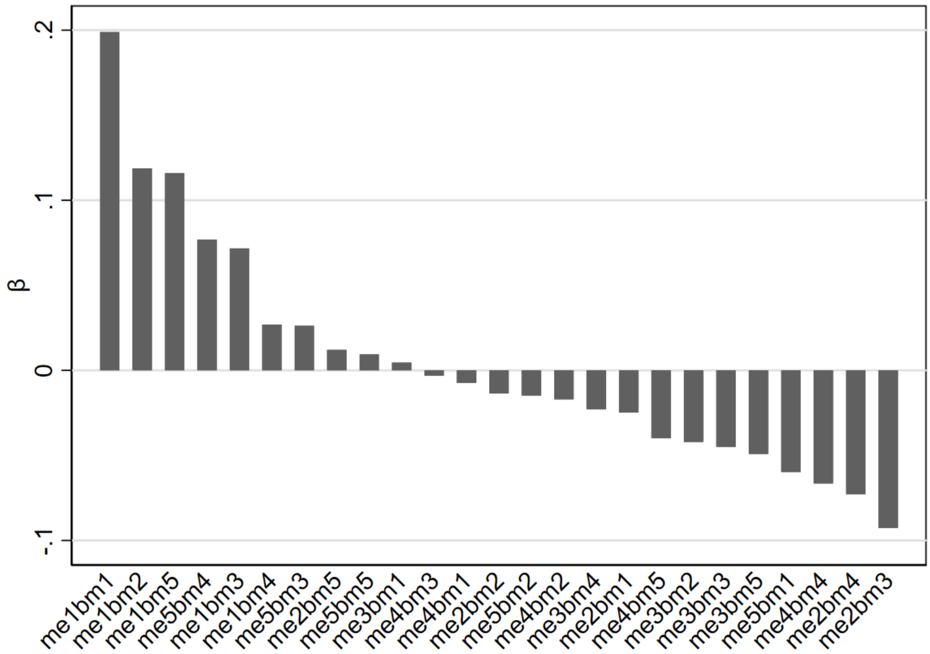
Note: Parameters $\theta = \{\gamma_0, \gamma_1, \gamma_2, \gamma_3, \delta_0, \delta_1, \delta_2, \delta_3\}$ minimize $g(\theta)' I g(\theta)$ and Instruments: Constant, $\mathcal{U}_t$, term, $d - p$, def. And, I is the identity matrix. This table presents results from the return and consumption growth predictability regressions on energy policy uncertainty. Parameters are estimated via GMM. $R_{t \to t+k}$ is the cumulative log return of the VWCRSP portfolio between month t and $t + k$. $\mathcal{U}_t$ is the level of energy policy uncertainty. Controls include the log dividend to price ration $(d - p)_t$, the term structure (term$_t$), the default spread (def$_t$). $\ln(c_{t+k}/c_t)$ is the growth on consumption between month t and $t + k$ measured as aggregate personal consumption expenses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Estimation sample of the predictability regression is 1980m1 to 2018m12.

Table 1.6: Cross-sectional return regressions

	Size and Book-to-Market				
	(1)	(2)	(3)	(4)	(5)
μ_{mktrf}	0.004*	0.005**	0.006***	0.006***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
γ_{mktrf}	-0.070	2.495	5.559**	5.903***	9.944***
	(2.267)	(2.056)	(2.216)	(2.046)	(2.737)
μ_{smb}		0.000	0.001		
		(0.001)	(0.001)		
γ_{smb}		1.211	2.149		
		(1.707)	(2.455)		
μ_{hml}		0.003**	0.002*		
		(0.001)	(0.001)		
γ_{hml}		5.589**	-4.225		
		(2.259)	(5.076)		
μ_{cma}			0.249***		
			(0.089)		
γ_{cma}			0.212**		
			(0.099)		
μ_{rmw}			0.353***		
			(0.103)		
γ_{rmw}			0.089**		
			(0.042)		
μ_{me}				0.002	0.002
				(0.001)	(0.001)
γ_{me}				5.805***	11.044***
				(2.202)	(2.854)
μ_{ia}				0.003***	0.003***
				(0.001)	(0.001)
γ_{ia}				17.788***	10.998*
				(4.669)	(5.843)
μ_{roe}				0.005***	0.005***
				(0.001)	(0.001)
γ_{roe}				13.657***	-4.229
				(3.883)	(6.520)
μ_{eg}					0.008***
					(0.001)
γ_{eg}					48.353***
					(14.931)
μ_u	0.001	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
γ_u	-57.653***	-34.558**	-27.699*	-17.717	-19.528
	(21.357)	(16.859)	(16.295)	(12.821)	(15.393)
Observations	468	468	468	468	468
MAE %	.24	.29	.17	.17	.18

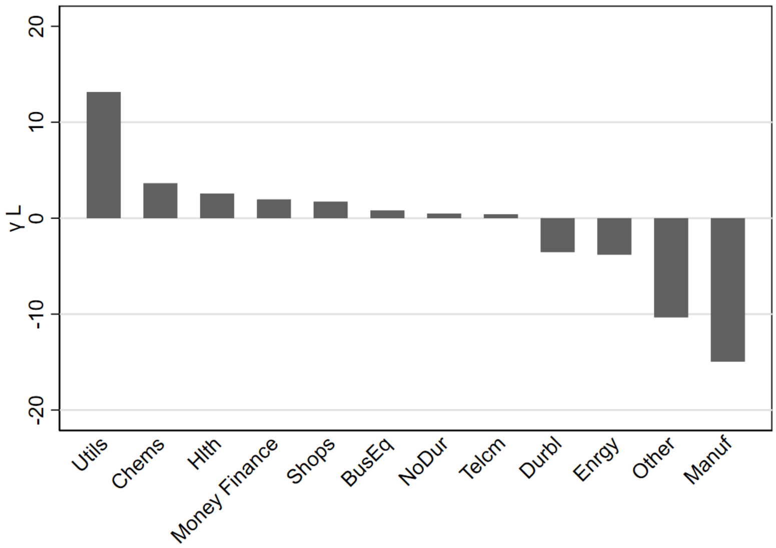
Note: This table presents results from estimating the price of energy policy uncertainty in expected return - covariance form by extending the CAPM model, Fama and French three and five factor models, and the q^4 and q^5 models with its innovations (u_t). Estimations are performed via GMM in which factor loadings (covariances) and covariance prices of risk are estimated jointly. Factors smb, hml, cma, rmw, correspond to the Fama and French factors related to size, book-to-market, investment and profitability. Factors, me, ia, roe, and eg, correspond to factors related to size, investment, profitability, and expected investment growth. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Estimation sample of the cross-sectional regression is 1980m1 to 2018m12.

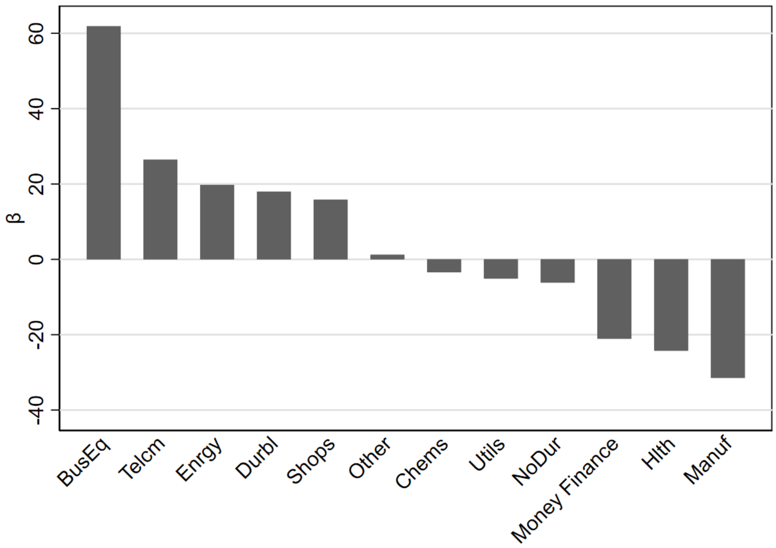
Figure 1.4: Differences in energy policy uncertainty betas across portfolios sorted on size and book-to-market



Note: This figure provides estimates for each portfolio sorted on size and book-to-market of running the following time series regression $R_{it} = a + b \times (R_{mt} - r_{ft}) + \beta u_t + \epsilon_t$, where $R_{mt} - r_{ft}$ is the excess return of the CRSP value weighted portfolio over the one month risk free rate. Portfolio returns R_{it} for each quintile in the double sorting of size and book to market firms are obtained from Prof. Kenneth French's website, me1 to me5 correspond to quintiles 1 to 5 on size, and bm1 to bm5 correspond to quintiles1 to 5 on book to market. Estimation sample is 1980m1 to 2018m12.

Figure 1.5: Average impact of lobby on policy uncertainty exposure by industry





Note: These figures present the equally weighted average by industry of the zero-lobbying exposure to energy policy uncertainty β_i^{energy} and the risk reduction of lobbying $\gamma_i \bar{l}_i$ from estimating equation $R_{it} = a_i + b_i R_{mt} + (\beta_i^{energy} + \gamma_i \bar{l_{it}}) u_t + \epsilon_{it}$ for each firm in the sample where $\bar{l}_{it}$ is the average lobby expenses over lagged total assets of firm i. Estimation sample is 1997m1 2018m10.

Table 1.7: Regressions of oil and gas betas on energy political betas

	(1) β^{oil}_{it}	(2) β^{oil}_{it}	(3) β^{oil}_{it}	(4) β^{gas}_{it}	(5) β^{gas}_{it}	(6) β^{gas}_{it}
β^{energy}_{it}	4.72***	4.42***	3.90***	0.07	-0.04	-0.06***
	(0.11)	(0.10)	(0.08)	(0.05)	(0.03)	(0.02)
$\beta_{it}/100$	5.12	10.26*	-1.79	-4.79	-2.24	-3.21
	(7.85)	(5.69)	(4.25)	(6.27)	(3.80)	(2.89)
Leverage$_{it}$	0.96*	-0.13	1.54***	0.69***	-1.91***	-3.17***
	(0.50)	(0.40)	(0.50)	(0.25)	(0.17)	(0.34)
Constant	-1.90***	-1.69***	-2.06***	0.33***	0.88***	1.17***
	(0.15)	(0.11)	(0.13)	(0.09)	(0.05)	(0.08)
Observations	404520	404520	404470	197236	197236	197219
Adjusted R^2 (%)	18.92	28.26	55.72	1.41	21.72	56.86
Time F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	No	Yes	Yes	No	Yes	Yes
Firm F.E.	No	No	Yes	No	No	Yes
From	1974m6	1974m6	1974m6	2002m1	2002m1	2002m1
To	2018m10	2018m10	2018m10	2018m10	2018m10	2018m10

Note: This table presents results from regressing firm level gas and oil betas on market beta, energy policy uncertainty beta and leverage. The beta from energy is computed using a 60 month rolling window of running firm's returns on the market return and innovations on the energy policy uncertainty measure. $R_{it} = a + bR_{mt} + \beta^{energy} u_t + \epsilon_{it}$ where R_{mt} is the return on the CRSP Value Weighted Market Portfolio, and u_t is the innovation on energy policy uncertainty. Oil beta β^{oil} is defined as the slope of regressing firm returns on the market return and the West Texas Intermediary monthly return using a 60 month rolling window. $R_{it} = a + bR_{mt} + \beta^{oil} R^{oil}_t + \epsilon_{it}$. Gas betas are computed using the return on the monthly Henry Hub Natural Gas Spot price $R_{it} = a + bR_{mt} + \beta^{oil} R^{gas}_t + \epsilon_{it}$, β is computed using daily returns within each month and is defined as the sum of coefficients $\beta = b_1 + b_2 + b_3$ from estimating the following regression month by month: $R^e_{is} = a + b_1 Rm^e_{i,s+1} + b_2 Rm^e_{is} + b_3 Rm^e_{i,s-1} + \epsilon_{is}$ for all days s within month t and R^e_m is the daily market excess return over the daily risk free rate. Leverage is computed as total debt = Compustat Quarterly items (dlcq+dlttq) over total debt plus market equity (prc $\times$ cshoq). Oil and gas prices come from the Federal Reserve Economic Data at St. Louis. Clustered standard errors at the month level reported in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

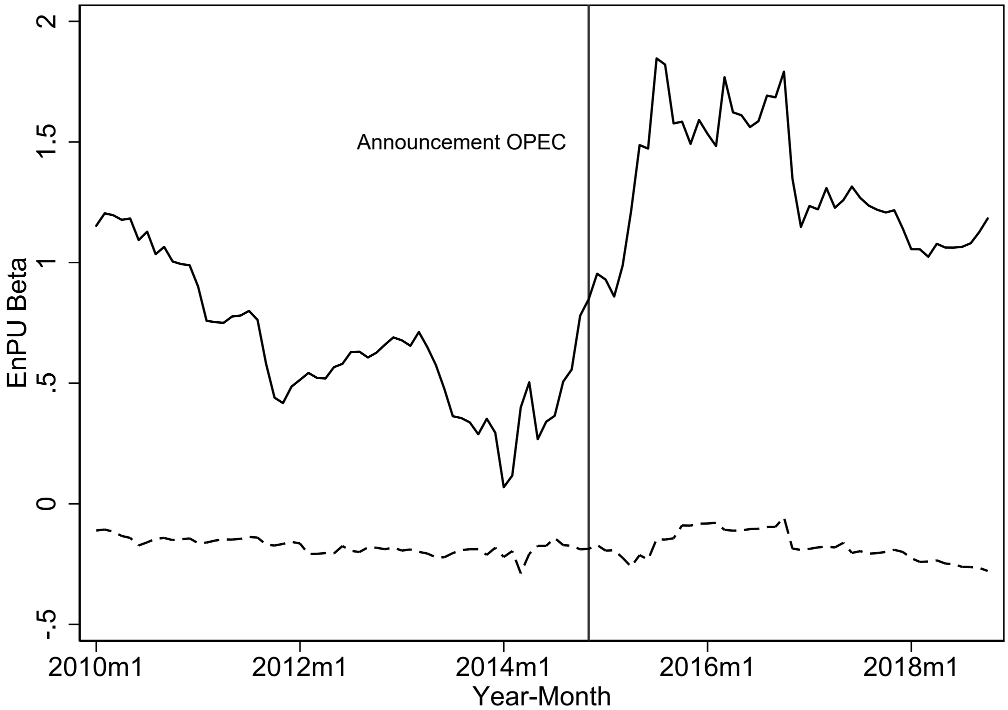


Figure 1.6: Differences in the average energy policy uncertainty beta between oil and non oil related firms

Table 1.8: Difference-in-Differences regressions of energy policy uncertainty betas on the 2014 OPEC announcement

	(1) β^{energy}	(2) β^{energy}
Oil related dummy$_{it}$	0.21***	0.21***
	(0.02)	(0.02)
After OPEC announcement dummy$_t$	0.07***	
	(0.01)	
After OPEC announcement dummy$_t$ $\times$ Oil related dummy$_{it}$	0.84***	0.84***
	(0.04)	(0.04)
Constant	-0.14***	-0.11***
	(0.00)	(0.00)
Observations	103021	103021
Adjusted R^2 (%)	2.18	2.75
Time F.E.	No	Yes
From	2010m1	2010m1
To	2018m10	2018m10

Note: This table presents Clustered standard errors at the month, firm, and month-firm level reported in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Chapter 2

Tweeting for money: Social media and mutual fund flows

Joint with Javier Gil-Bazo

In contrast to mandatory information disclosures, social media offer companies the opportunity to communicate with investors with few constraints on frequency, content, and format. To investigate the use of social media by asset management firms, we collect a database of 1.4 million Twitter posts by mutual fund families offering equity funds in the US from 2009 to 2017 and analyze their content using machine learning algorithms. We find that larger and younger families use Twitter more intensively. Investors do not respond to the amount of social media activity of a fund family but to the tone of its posts. A one standard deviation increase in the positiveness of a family's tweets in a given month increases its assets under management by 15 basis points, or USD 11 million, in the following month. However, tweets with a positive tone do not predict higher subsequent fund performance. These results suggest that asset managers use social media to persuade investors rather than to alleviate information asymmetries by either lowering search costs or disclosing privately observed information. Consistently with this explanation, families facing more difficulties in raising assets benefit the most from positive posts on Twitter.

2.1 Introduction

As of April 2021, more than 4 billion people in the world, 70% of the population aged 13 years and older, were using social media to communicate with others, entertain themselves, and stay informed.[1] The growing popularity of social media has raised concerns about their potential to misinform the public and manipulate individuals' opinions and behavior (e.g., Abramowitz, 2017; Aral and Eckles, 2019). In the context of financial markets, the online activities of some high-profile individuals have prompted investigations by the Securities and Exchange Commission (SEC).[2] If social media can be used to influence investors' perceptions, then companies issuing securities to the public have strong incentives to become active participants. In this paper we focus on the market for mutual funds and study whether asset managers use social media to attract money from investors.

As argued by Mullainathan et al. (2008), persuasion in finance involves exploiting investors' biases to change their perceptions. In this sense, social media is an ideal tool for persuasion for asset management firms, as it allows them to communicate with current and prospective investors without the strict constraints imposed by mandatory information disclosures on the timing, content, and framing of information.[3] For instance, firms may choose to communicate only positive information. They may also time their communications to maximize the impact on investors' decisions. And they may frame information in the most favorable way possible.

Asset managers have strong incentives to use social media in order to

[1] Data from datareportal.com (Global Social Media Stats).

[2] Mohamed (2021),"Big Short' investor Michael Burry says he'll stop tweeting after SEC regulators paid him a visit," Businessinsider.com, (https://markets.businessinsider.com/currencies/news/big-short-investor-michael-burry-stop-tweets-sec-regulators-visit-2021-3-1030222890); SEC (2018),"Elon Musk Charged With Securities Fraud for Misleading Tweets," (https://www.sec.gov/news/press-release/2018-219); ; Spichak (2021), "Elon Musk Hopes SEC Will Investigate Him over Dogecoin Tweets: 'It Would Be Awesome'," Newsweek, (https://www.newsweek.com/elon-musk-sec-investigation-dogecoin-bitcoin-cryptocurrency-tweets-1572290).

[3] Note, however, that advertisement and retail investor communication by asset management companies must comply with SEC rule 482 and FINRA rule 2210. In 2003, SEC rule 482 modified the Securities Act of 1933-Section 5 that stated that all fund advertisement must have information that is contained in the statutory prospectus. With rule 482, investment companies are allowed to include information not included in the statutory prospectus. This allows investment companies to include up-to-date information in rule-482 advertisements, such as information about current economic conditions that are not commonly included in a fund's prospectus. FINRA Rule 2210 governs communications with the public including communications with retail and institutional investors. The rule provides standards for the content, approval, recordkeeping and filing of communications with FINRA. The rule prohibits false, exaggerated, unwarranted, and misleading information communications, as well as projections of future performance.

influence investors' perceptions and increase their assets under management. However, it is unclear whether they will succeed in this endeavor. The mutual fund market is highly regulated and abundant hard information is already available through mandatory disclosures, such as fund prospectuses and statements of additional information. Also, if investors understand the ability of asset management firms to strategically select and frame information, any attempts to influence investors could be self-defeating.

The mutual fund industry is an ideal laboratory to study the role of social media communication in financial markets. First, thousands of actively managed mutual funds compete for investors' money. Second, there is asymmetric information about managerial ability and other determinants of fund performance. While asset management companies can closely monitor portfolio managers' decisions and influence their performance through the allocation of resources within the firm, investors can only learn about funds' future expected performance from public information such as past returns and infrequently disclosed portfolio holdings. Another important advantage of the mutual fund setting is that open-end mutual fund shares trade at their net asset value, which makes it possible for researchers to observe directly investors' response to firms' actions by looking at flows of money into and out of mutual funds.

Twitter is also particularly appropriate for our purposes given its rising popularity among investors. Indeed, a number of studies have shown evidence that Twitter activity can predict prices of stocks and other asset classes (Bollen et al., 2011; Ranco et al., 2015; You et al., 2017; Gholampour and van Wincoop, 2017; Gu and Kurov, 2020). Also, the presence of asset management firms in Twitter has grown at a very fast pace in the last years. In our sample, the number of posts on Twitter (tweets) by all mutual fund families went from almost zero prior to 2009 to over 20,000 tweets per month in 2017.

To investigate whether asset management companies influence investors' decisions through social media communications, we build a database of Twitter posts by mutual fund families managing domestic diversified equity funds in the US between January 2009 and October 2017. We then employ machine learning algorithms to classify tweets into positive or negative and compute the positiveness of the tone of asset management firms' tweets in a given month. Finally, we merge these data with the CRSP Survivor-Bias-Free US Mutual Fund database, which contains information on fund, manager and family characteristics.

Our results can be summarized as follows. First, 241 of 785 firms managing US diversified equity funds have a Twitter account and post at least one tweet during our sample period. Families that use Twitter tend to manage more assets, more funds, and funds in more investment categories than families that do not use Twitter, which suggests that economies of scale play a role in the decision to implement a social media strategy. Among those firms that use Twitter, more frequent users tend to be younger and to manage more assets and funds with higher past performance, higher expenses, higher loads, and lower volatility.

We find no association between the number of tweets by a family in a given month and flows of money to funds in that family in the following month, controlling for fund performance, observable fund and fund family characteristics that have been documented to predict fund flows, and time-invariant fund and family characteristics.

However, we find that a more positive tone in a family's tweets in a given month predicts significantly higher flows to the family's funds in the following month. The increase in flows to the family's funds following tweets with a positive tone is economically significant. A one standard deviation increase in the tone of tweets is associated with an increase in assets under management of 15 basis points (bp) in the following month, or 11 USD million for the average family. This result is robust to different ways of modelling the flow-performance relationship, to controlling for previously documented determinants of mutual fund flows, and to the inclusion of time, fund, and fund family fixed effects. To rule out the possibility that fund families tweet about events that are public knowledge and may be trigger fund flows, we repeat the analysis controlling for known events, such as manager turnover, social media mentions by third parties, and the fraction of funds with very recent stellar performance in the family. In all cases, the association between recent tweets and fund flows survives.

To further investigate the mechanism through which Twitter activity influences fund flows, we obtain data on share purchases and share redemptions from SEC filings, and run separate regressions for inflows and outflows. We find that positive Twitter posts both increase inflows and decrease outflows.

Our results are consistent with asset management firms using social media to persuade investors, consistent with the theory of Mullainathan et al. (2008). However, we consider two alternative explanations for the results documented in this paper. First, building on the work of Sirri and Tufano (1998), Hortaçsu and Syverson (2004), and Huang et al. (2007), asset management companies could use social media to reduce search costs for investors. This can be achieved by directing investors to information about fund offerings, fees, or past performance, that is already available but difficult to locate for investors. Under this hypothesis, we would expect the *amount* of social media activity to increase flows, since a reduction in search costs increases the number of new investors who are willing to pay the cost to learn about a mutual fund and become a potential buyer.[4] Our finding that the number of tweets does not predict flows does not support this prediction.

Second, we explore whether asset management companies use social media to convey to investors information that is not available to the public. More specifically, the model of Dumitrescu and Gil-Bazo (2016) of strategic communication by asset managers predicts that asset management companies will communicate information that is favorable for future fund performance and which is not already publicly available. Since this information is both new and truthful, favorable communications have a positive impact on flows of new money. But the model also implies that asset manager communications have predictive power with respect to future performance, controlling for publicly available information. To test this prediction, we investigate whether more positive tweets predict superior fund performance controlling for well-documented predictors of performance, including possible diseconomies of scale (Berk and Green, 2004; Chen et al., 2004; Pástor et al., 2015; Zhu, 2018). We find that the positiveness of an asset management company's tweets does not predict future fund performance. This evidence contradicts the information hypothesis.

In sum, the empirical evidence documented in this paper does not support the notion that social media communications of asset management firms alleviate information asymmetries between mutual fund companies and investors by either reducing search costs or conveying new information to investors.

If the purpose of social media communications is to persuade investors, we would expect social media activity to benefit more those asset managers that experience more difficulties in attracting investors' money. Consistently with this prediction, we find that the link between positive tweets and asset growth is stronger for fund families with fewer assets under management, managing fewer funds, and with lower recent flows into the family. We also find that flows to funds that cater to retail investors are three times more sensitive to the tone of asset managers' posts, which gives further credence to the persuasion hypothesis.

By unveiling the role of social media communications on mutual fund

[4]Of course, many investors who learn about a mutual fund may not find the fund a desirable investment. However, since open-end mutual funds cannot be shorted, the impact of lower search costs on net purchases is necessarily positive.

investors' decisions, our paper contributes to a large literature on the determinants of mutual fund flows (see Christoffersen et al. 2014, for a survey). More specifically, our paper is related to a number of studies that investigate the role of advertising in the mutual fund industry. Sirri and Tufano (1998) show that marketing effort, as proxied by fund fees, increases fund flows. Jain and Wu (2000) study a sample of 294 funds that are advertised either in Barron's or in Money magazine and find that even though the pre-advertisement performance of these funds is better than the performance of their benchmark, there is no superior performance in the post-advertisement period. Cronqvist (2006) investigates the content of mutual fund advertisements in Sweden and finds that most fund ads are not informative about fund quality. Nevertheless, fund ads influence individuals' portfolio decisions, steering them towards high-fee funds, locally concentrated portfolios, and funds investing in sectors with high recent performance. Gallaher et al. (2015) show that mutual fund families' advertising expenditures attract flows to the family's funds as well as to other funds in the industry, reduce redemptions, and increase the convexity of the flow-performance relationship. We contribute to this literature by studying a new and increasingly important means of communication which, unlike traditional media advertising, allows firms to interact directly and in real time with investors. Moreover, we use textual analysis and machine learning to measure the tone of communications.

Our paper is most closely related to the study of Hillert et al. (2016). These authors use textual analysis to determine the tone of shareholders' letters from asset management companies and their impact on fund flows. The authors find a positive association between tone and subsequent flows. Using daily information about flows for a subsample of funds, they also provide evidence that the reaction to shareholder letters appears right after shareholder letters are sent to investors, and reverts around five days after shareholders receive the letter. The authors also find evidence that funds that address their shareholders in a more personal manner, have better subsequent performance on average. Although shareholder letters provide some freedom to managers to communicate to shareholders, they are part of the shareholder's report (Form N-CSR and N-CSRS filings) and highly regulated by the SEC in terms of their frequency, format and content. Like Hillert et al. (2016), we find that fund families that post more positive information receive higher net flows and experience fewer redemptions. However, we find that the relation between positiveness of tweets and flows does not revert after a few days. Moreover, we find that these communications are not informative with respect to future fund performance.

Our paper also contributes more generally to the literature on non-mandatory corporate disclosures (Kim and Verrecchia, 1991; Dye and Sridhar, 2004; Dye and Sridhar, 2004; Cornelli et al., 2013; Bertomeu and Marinovic, 2016), and to the recent literature of textual analysis in Finance and Accounting (see Loughran and McDonald, 2016 for a survey of the literature). The study of Blankespoor et al. (2014) is particularly relevant to our paper. These authors show that when public firms use Twitter to disseminate firm-initiated news, information asymmetries decline as evidenced by narrower bid-ask spreads. In contrast, our results suggest that Twitter does not help alleviate information asymmetries in the mutual fund industry.

2.2 Data

In this section, we present the data used in the analysis. We draw on two datasets, the CRSP Survivor-Bias-Free US Mutual Fund database and a database of tweets from January 2009 to October 2017 posted by mutual fund families. From the former, we obtain information on mutual fund returns, assets under management, investment category, and expenses. Even though our Twitter database starts in 2009, we collect mutual fund data from 2006 so we can use three years of prior historical data to estimate risk-adjusted returns.

To construct variables at the mutual fund level, we follow the same share aggregation procedure as in Gil-Bazo and Ruiz-Verdú (2009). We start with 4,124,178 observations at the share class level between January 2006 and October 2017. We keep only diversified domestic equity funds (1,914,233 observations remaining), and drop passively managed funds (1,733,624 observations remaining). Total Net Assets (TNA) of a fund are the sum of the TNA under each share class. Returns and expense ratios are TNA-weighted averages across all share classes in the fund. The age of the fund is the age of the oldest share class in the fund.

To create some of our variables, we aggregate data at the fund family level based on the CRSP identifier mgmt_code. TNA at the family level is the sum of the TNA of each fund in the family, the age of the fund family is the age of the oldest fund in the family, and expenses and returns are weighted averages across all funds in the family (based on the TNA of each fund in the family).

For a subsample of funds, we obtain data on inflows and outflows, as in Christoffersen et al. (2013) and Ha and Ko (2019). These data can be obtained from SEC's N-SAR form, Item 28, which includes cash-flow information on a monthly basis at the portfolio level.[5]

Given the findings of Barber et al. (2016) and Berk and van Binsbergen (2016) that investors appear to use the CAPM to evaluate mutual fund performance, throughout the paper we focus on CAPM alphas as a determinant of flows, although we test the robustness of our results to using the three-factor and four-factor models of Fama and French (1993) and Carhart (1997) to estimate performance.[6] We compute the risk-adjusted return, $\hat{\alpha}_{it}$, of fund i in month t as the intercept plus the residual of the CAPM model:

$$\hat{\alpha_{it}} = r^{e}_{it} - \hat{\beta_{it}} r^{e}_{mt}, \tag{2.1}$$

where r^{e}_{it} is the excess return of fund i at month t over the risk free rate and r^{e}_{mt} is the excess return of the market portfolio over the risk free rate. We obtain the monthly risk-free rate and the market portfolio return from Prof. Kenneth French's website and $\hat{\beta_{it}}$ is estimated for each fund and month t by running OLS rolling regressions of excess returns on market excesss return over the three-year period ending in month $t-1$. If less than three years of data are available in a given window, we require the fund to have at least 30 months of data and run the regressions with the data available.

To construct the Twitter database of mutual fund families we obtain the names of all asset management companies in the CRSP database managing US equity funds. Then, we perform a manual search through each one of the family names represented in the variable mgmt_name in the CRSP database and group similar names using the CRSP aggregation variable mgmt_code. Finally, we search for each family's Twitter account in the asset management company's website.

Once the list of Twitter accounts is collected, we web scrape all tweets from accounts that are active in 2017. It is important to notice that if a fund family that was active in the past decided to cancel its Twitter account we would not be able to get this information. The web-scraping procedure downloads tweets historically starting from the most recent tweet up to the first one. Web-scraping algorithms can get banned temporarily and the download procedure may stop prematurely. To ensure we download all information, we compare the last tweet obtained for each company with the true first tweet of the account provided by Twitter.[7] Our database con-

[5]We thank Yeonjeong Ha and Kwangsoo Ko for kindly sharing their data with us.

[6]Evans and Sun (2020) show that mutual fund flows have become more sensitive to three-factor abnormal returns since Morningstar changed its methodology to compute fund ratings to account for funds' investment style.

[7]The first tweet of any active account was found using the webpage

tains 1,433,926 tweets from 362 different usernames, from January 2009 to October 2017.

The procedure used to measure the positiveness of tweets is explained in detail in the Appendix and can be summarized as follows. We first classify tweets into two categories (financial and nonfinancial) and then determine the tone (positive or negative) of each tweet. To determine the tone, we use a training sample with previously manually classified tweets and a well-known training sample of tweets provided by the University of Michigan. Whether we use one training sample or the other depends on whether the tweet is classified as financial or nonfinancial. The distinction between financial and nonfinancial tweets is based on Loughran and McDonald (2011), who argue that financial text should not be classified using training samples from other social sciences since its particular context may lead to classify common jargon as negative. The Appendix provides examples of tweets that have been classified as financial, nonfinancial, positive and negative.

To avoid any subjectivity in choosing the machine learning algorithm to classify the tweets, we use six different algorithms and select for each tweet the most voted label among them. If three algorithms classify a tweet as positive and three as negative, we consider the tweet to have a neutral tone. Using this voting scheme, all tweets in our sample are classified as either positive or negative. The approach also provides us with as a measure of confidence in the classification. In particular, we define the confidence of classifying tweet k as c as:

$$w_k^c = \frac{\text{Number of algorithms that classify tweet } k \text{ with label } c}{\text{Total number of algorithms}} \tag{2.2}$$

We then define the positiveness of a family's tweets in month t as follows:

$$\text{Positiveness}_t = \ln\left(\frac{1+M_t^p}{1+M_t^n}\right), \tag{2.3}$$

where $M_t^p(M_t^n)$ is the weighted count of positive (negative) tweets of that family in one month:

$$M_t^p = \sum_{k \in \mathcal{D}(t)} w_k^p x_k^p,\ M_t^n = \sum_{k \in \mathcal{D}(t)} w_k^n x_k^n, \tag{2.4}$$

where $D(t)$ is a monthly time interval, x_k^p (x_k^n) is an indicator variable that takes the value of 1 if tweet k at time t is positive (negative), and w_k^p (w_k^n)

https://discover.twitter.com/first-tweet, which is no longer available, although other websites provide the same service.

is the confidence in the tweet's positive (negative) label given the level of agreement among all classifiers for a particular tweet as in equation (2.2). Our measure of Positiveness is closely related to that employed by Antweiler and Frank (2004), but is more appropriate for handling Twitter accounts with zero tweets.

Figure (2.1) displays the total number of tweets across all mutual fund families, as well as the weighted count of positive and negative tweets per month. The figure shows a sharp increase in Twitter usage by mutual fund families, with a peak in 2016. As expected, positive tweets predominate over tweets classified as negative.

Out of 785 fund families in the final CRSP sample, 241 fund families tweet at least once during the sample period. This is the subsample we use in most of our analysis. To understand how this subsample differs from the rest, Table (2.1) presents descriptive statistics of both fund and family characteristics for the Twitter subsample and the full sample. At the fund level, there are no clear differences between funds managed by fund families in the Twitter subsample an funds in the entire sample. However, at the family level differences between fund families in both samples become more evident. Fund families in the Twitter subsample are on average older, manage more assets, more funds, funds in more different investment categories, and more funds that charge loads.

2.3 Determinants of Twitter activity by mutual fund families

We start our analysis by investigating which families are more likely to use Twitter. Although social media communication has low explicit costs, the implicit costs are non-trivial. Managing a social media communication strategy requires that social media managers coordinate with the marketing department and senior management in the process of setting goals, creating contents, and engaging with the public. In addition, contents need to be created, the firm's social media presence must be promoted, technological support is required, and the whole process must be carefully monitored and evaluated. Since such costs are likely to have a fixed component, we expect large asset management firms to be more likely to have a social media presence and use it actively. To explore this conjecture, in addition to the amount of assets under management to proxy for size, we use the number of funds and the number of different categories (both in logs) in which families offer funds. We also study whether younger firms are more likely to use Twitter since younger families have more incentives to gain visibility among investors. Finally, we control for the characteristics of funds in the family: asset-weighted average CAPM alpha over the previous 12 months; asset-weighted average expense ratio; number of funds in the family that charge loads (in logs); and asset-weighted average volatility of fund returns in the previous 12 months.

We analyze both the extensive and the intensive margins of families' Twitter usage. More specifically, in our tests, we employ two different dependent variables. The first variable, Twitter, is an indicator that equals one if the fund family has a Twitter account and uses it at least once in our sample period. The second variable, Number of Tweets, is defined for each family and month and is computed as the natural logarithm of 1 plus the number of Tweets posted by the family in that month.

We start by estimating a cross-sectional linear probability model with the Twitter indicator as the dependent variable using the full sample. Fund-level explanatory variables are first computed for each family and month. All variables are then aggregated at the family level by computing their time-series means within each family. Estimation results are presented in Panel A of Table (2.2). Family age is not associated with having a Twitter account. However, all three proxies for family size are positively and significantly associated with the family's presence in Twitter. In other words, smaller management companies are less likely to consider Twitter as a way of communicating with investors. There is no significant association between fund characteristics and a Twitter account.

We then regress the variable Number of Tweets on the same set of explanatory variables as in the previous regression, but defined at the family-month level, and lagged one month with respect to the dependent variable. In this case, we naturally restrict the sample to families with Twitter$=1$. We include family and time fixed effects and compute robust standard errors clustered at the month level. Estimation results are presented in Table (2.2). Conditional on having a Twitter account, both younger and larger firms tend to tweet more frequently. As for fund characteristics, families with better-performing funds and families with more expensive funds also tweet more. Return volatility, on the other hand, is negatively associated with the intensity of Twitter activity.

These results suggest that economies of scale are a key determinant of social media usage by asset managers. Conditional on having presence on Twitter, its usage appears to respond not only to cost considerations but also to the potential benefits of social media for firms: gaining visibility for younger firms, publicizing good performance, and raising assets for high-fee funds.

2.4 Twitter activity and fund flows

In this section we study how flows of new money to mutual funds respond to posts of fund families on Twitter. Following the literature, we compute inflows to fund i between month t and month $t+1$ as the growth rate in total net assets net of the fund's return:

$$\text{Flows}_{i,t+1} = \frac{\text{TNA}_{i,t+1} - \text{TNA}_{it}(1 + r_{i,t+1})}{\text{TNA}_{it}}, \tag{2.5}$$

where TNA_{it} is the total net assets of fund i at the end of month t, and $r_{i,t+1}$ is the fund's monthly return. To minimize the impact of outliers - mostly small funds with large percentage of inflows or outflows - we follow the literature and winsorize flows at the 1% level.

We begin our analysis by studying how the number of tweets by an asset management firm in a given month is related to flows to funds in that family in the following month, controlling for fund performance and other well-documented flow determinants.[8] Like Sirri and Tufano (1998), we allow for a convex flow-performance relationship. To model dependence on performance, we employ two different approaches. First, we define the variable Rank_{it} as the ranking of fund i's CAPM alpha in the 12-month period ending in month t against all other funds in the same Lipper category, normalized to be between $1/N$ (lowest performing fund) and 1 (highest performing fund), where N denotes the number of funds in the corresponding category and month.

Second, we use objective-adjusted abnormal return (OAR) as an alternative to performance rank. As argued by Ha and Ko (2019), OAR accounts for the potentially large dispersion in the cross-section of fund performance and its impact on the flow-performance relationship. We compute OAR_{it} by standardizing the 12-month CAPM alpha to have zero mean and unit standard deviation across all funds in the same investment category.

For both $\text{Performance}_{it} \in \{\text{Rank}_{it}, \text{OAR}_{it}\}$, we compute the following variables:

$$\begin{aligned}
\text{Low Performance}_{it} &= \min(\text{Performance}_{it}, p20) \\
\text{Mid Performance}_{it} &= \min(\text{Performance}_{it} - \text{Low Performance}_{it}, p80 - p20) \\
\text{High Performance}_{it} &= \text{Performance}_{it} - \text{Mid Performance}_{it} - \text{Low Performance}_{it},
\end{aligned} \tag{2.6}$$

[8] Henceforth, we restrict the sample to funds in fund families that have tweeted at least once between 2009 and October 2017.

where $p20$, $p80$ denote the 20th and 80th percentiles, respectively, of either the cross-sectional distribution of performance rank or OAR.

We first analyze the link between the number of tweets posted by a fund family and subsequent flows. More specifically, we estimate the regression equation:

$$\begin{aligned} \text{Flows}_{i,t+1} = {} & \gamma_0 + \gamma_1 \times \text{Number of Tweets}_{it} \\ & + \gamma_2 \times \text{Low Performance}_{it} + \gamma_3 \times \text{Mid Performance}_{it} \\ & + \gamma_4 \times \text{High Performance}_{it} \\ & + \gamma_5 \times X_{it} + \delta_{t+1} + \lambda_i + \mu_{cat} + \theta_{fam} + \nu_{i,t+1}, \end{aligned} \tag{2.7}$$

where $\text{Flows}_{i,t+1}$ is in %. Number of Tweets_{it} is the natural logarithm of one plus the number of tweets posted by fund i's family in month t. Low, Mid, and High Performance are calculated using both Rank and OAR based on 12-month CAPM alphas as in Equation (2.6). Following the large literature on the determinants of fund flows, the vector of lagged controls, X_{it}, includes the natural logarithm of the fund's total net assets, the fund's expense ratio, the fund's age (log of months since inception), and flows into the fund. Controls also include flows to funds in the same investment category in month $t+1$ and the standard deviation of returns in the 12-month period from $t-11$ to t. Importantly, we control for family size (log of assets under management) and family age (age of the family's oldest fund), since we know from the previous section that these variables are associated with the family's decision to use Twitter. δ_{t+1}, λ_i, μ_{cat}, and θ_{fam} denote month, fund, investment category, and family fixed effects, respectively.[9] Finally, $\nu_{i,t+1}$ denotes the error term.

We also study the relationship between flows and the tone of tweets by replacing Number of Tweets_{it} in equation (2.7) with Positiveness_{it}, which is the value of Positiveness_t for fund i's family as defined in equation (2.3). That is, we estimate:

$$\begin{aligned} \text{Flows}_{i,t+1} = {} & \gamma_0 + \gamma_1 \times \text{Positiveness}_{it} \\ & + \gamma_2 \times \text{Low Performance}_{it} + \gamma_3 \times \text{Mid Performance}_{it} \\ & + \gamma_4 \times \text{High Performance}_{it} \\ & + \gamma_5 \times X_{it} + \delta_{t+1} + \lambda_i + \mu_{cat} + \theta_{fam} + \nu_{i,t+1}, \end{aligned} \tag{2.8}$$

We estimate equation (2.7) using pooled OLS and compute robust standard errors clustered at the month, fund family, and month-fund family levels.

[9] In our sample, some funds change investment categories through time.

Table (2.3) presents the results. To model dependence of flows on performance we use OAR in columns (1)-(3) and Rank in columns (4)-(6).

In column (1) we do not include Number of Tweets or Positiveness. As previously documented in the literature, we find a convex relation between flows and performance. Fund size, flows to funds in the same category, and volatility are all negatively associated with flows. Flows are persistent as evidenced by the positive and significant coefficient on lagged flows. Finally, younger funds and larger families capture more flows.

In column (2), we include Number of Tweets. The estimated coefficient on this variable is small and statistically insignificant. Therefore, the intensity of social media communications by asset management companies appears to be unrelated to future fund flows, controlling for observable characteristics as well as time-invariant fund and family characteristics.

In column (3), we replace Number of Tweets with Positiveness, as in equation (2.8). The coefficient on Positiveness is positive and statistically significant at the 5% level, which suggests that flows respond to a positive tone in asset management companies' tweets. Note that this association cannot be driven by fund or family time-invariant characteristics that determine both the tone of families tweets and fund flows. It is not driven either by larger or younger companies' tendency to tweet more.

In columns (4) to (6), we show estimation results when we use Rank instead of OAR. Although the relation between flows and performance appears to be more convex, consistent with the results of Ha and Ko (2019), the association between Number of Tweets and Flows and between Positiveness and Flows is robust to modelling the flow-performance relationship in this way. In particular, the estimated coefficient on Number of Tweets is close to zero and non-significant and the estimated coefficient on Positiveness is positive and significant at the 5% level.

In unreported results, we repeat the analysis using the three-factor and four-factor models of Fama and French (1993) and Carhart (1997), respectively, to compute both Rank and OAR. Our conclusions are qualitatively and quantitatively similar.

In terms of the economic magnitude of the association, using the estimated coefficient of columns (3) and (6), a one standard deviation increase in Positiveness (0.78 for the Twitter subsample) corresponds to an increase in subsequent flows of 0.04% ($= 0.05 \times 0.78$), which for the average fund in the Twitter sample corresponds to an increase of USD 319,492 per month ($= 0.04\% \times$ USD 798.73 million).

To gauge the economic impact of Twitter posts' tone for the average family, we need to estimate the marginal effect of positiveness on flows at the family level. In Table (2.4) we estimate a version of equation (2.8) where all variables are collapsed at the family-month level. In column (1) we use OAR and include time fixed effects but not family fixed effects.[10] The estimated coefficient on Positiveness_{it} is statistically significant at the 1% level. A one standard deviation increase in Positiveness_{it} is associated with an increase of 0.148% (=0.19% $\times$ 0.78) in family flows, which given the average assets under management per family of USD 7.46 billion represents an increase in assets of USD 11 million.[11] In column (2), we add family fixed effects and estimate an almost identical coefficient on Positiveness, that is also statistically significant at the 1% level. Therefore, the association between Positiveness and flows to the family's funds is not driven by some unobservable time-invariant family characteristic. Results in columns (3) and (4) are obtained using Rank to model the flow-performance relationship and suggest a slightly stronger association between Positiveness and Flows, both with and without family fixed effects.

The results of Table (2.3) and Table (2.4) are consistent with social media influencing investor behavior. However, there is an alternative explanation for the positive association between Positiveness and fund flows. It could be that positive posts by asset management companies simply disseminate important information that is already public knowledge and that determines flows of money to mutual funds. To investigate this possibility, we evaluate whether the positive link between Twitter post tone and fund flows survives the inclusion of some variables that potentially impact fund flows. More specifically, we repeat the analysis controlling for: i) a change in the previous month in the fund's management company and the fund's portfolio manager, which can be perceived by investors as a positive signal for future returns (Khorana, 2001); ii) the number and tone of tweets by third parties that mention the fund family in the previous month; and iii) the fraction of funds in the family with monthly CAPM alpha in the top 5% of their investment category in the previous month (Nanda et al., 2004). Results in Table (2.5) indicate that the association between Positiveness and fund flows is still positive, similar in magnitude, and statistically significant at the 10% level after controlling for those performance-relevant events. Therefore, the coefficient on Positiveness is not simply picking up the effect of those events on fund flows.

[10]Naturally, the regression equation does not include fund fixed effects or investment objective fixed effects.

[11]One reason why the estimated increase in percentage flows is larger for the average family than for the average individual fund is that families with fewer funds, which are underrepresented in fund-level regressions, benefit more from positive tweets. We explore this possibility in Section 7.

2.5 Analysis of inflows and outflows

In this section we investigate whether Positiveness influences net flows by encouraging purchases of fund shares, discouraging redemptions, or both.

More specifically, we define Inflows and Outflows for fund i in month $t+1$ as:

$$\text{Inflows}_{i,t+1} = \frac{\text{New Sales}_{i,t+1}}{\text{TNA}_{i,t}} \tag{2.9}$$

$$\text{Outflows}_{i,t+1} = \frac{\text{Redeemed Cash}_{i,t+1}}{\text{TNA}_{i,t}} \tag{2.10}$$

As argued by Ha and Ko (2019) inflows and outflows are simultaneously determined by investors' rebalancing strategies. This mutual dependence between inflows and flows is tackled by performing a two-stage least square estimation. We follow closely Ha and Ko (2019) and run the following OLS regressions for inflows and outflows separately:

$$\text{Inflows}_{i,t+1} = a + \sum_{s=0}^{11} b_s \text{Inflows}_{i,t-s} + cX_{it} + \nu_{i,t+1}, \tag{2.11}$$

$$\text{Outflows}_{i,t+1} = a + \sum_{s=0}^{11} b_s \text{Outflows}_{i,t-s} + cX_{it} + \nu_{i,t+1}, \tag{2.12}$$

where X_{it} contains the same controls used in the flow regressions.

We then use the fitted values of the dependent variables $\widehat{\text{Inflows}}_{t+1}$ and $\widehat{\text{Outflows}}_{t+1}$ to estimate residual inflows and outflows:

$$\text{Inflows}_{i,t+1} = a + b\widehat{\text{Outflows}}_{t+1} + \epsilon^{I}_{i,t+1} \tag{2.13}$$

$$\text{Outflows}_{i,t+1} = a + b\widehat{\text{Inflows}}_{t+1} + \epsilon^{O}_{i,t+1} \tag{2.14}$$

Finally, fitted residual inflows, $\hat{\epsilon}^{I}_{i,t+1}$, and outflows, $\hat{\epsilon}^{O}_{i,t+1}$, are regressed on Positiveness and performance (Rank and OAR):

$$\begin{aligned}
\hat{\epsilon}^{I}_{i,t+1} = {} & \gamma_0 + \gamma_1 \times \text{Positiveness}_t \\
& + \gamma_2 \times \text{Low Performance}_{it} + \gamma_3 \times \text{Mid Performance}_{it} \\
& + \gamma_4 \times \text{High Performance}_{it} + \nu_{i,t+1} \\
\hat{\epsilon}^{O}_{i,t+1} = {} & \gamma_0 + \gamma_1 \times \text{Positiveness}_t \\
& + \gamma_2 \times \text{Low Performance}_{it} + \gamma_3 \times \text{Mid Performance}_{it} \\
& + \gamma_4 \times \text{High Performance}_{it} + \nu_{i,t+1}
\end{aligned} \tag{2.15}$$

In Table (2.7), we show estimation results for inflows in columns (1) and (2). The estimated coefficients for both OAR and Rank confirm the existence of a convex relationship between inflows and performance, consistent with Christoffersen et al. (2013). The coefficient on Positiveness is positive and significant at the 1% level.

In columns (3) and (4) we show estimation results for outflows. Consistent with Christoffersen et al. (2013), outflows appear to be highly sensitive to poor performance. We also find a statistically significant and negative association between Positiveness and outflows. The association is similar in magnitude to that between Positiveness and fund inflows.

Therefore, social media communications appear to increase net flows not only by fostering purchases of new shares but also by deterring investors from redeeming old shares.

2.6 Alternative hypotheses

The results presented in the previous sections are consistent with the persuasion hypothesis of Mullainathan et al. (2008). In this section, we consider two alternative explanations. First, building on the work of Sirri and Tufano (1998), Hortaçsu and Syverson (2004), and Huang et al. (2007), asset management companies could use social media to reduce search costs for investors. This can be achieved by directing investors to information about fund offerings, fees, or past performance, that is already available but difficult to locate for investors. In the model of Huang et al. (2007), a reduction in search costs increases expected fund flows by increasing the number of new investors that are willing to pay the cost to learn about a mutual fund and become a potential buyer. Therefore, the search cost hypothesis predicts that asset management social media activity will on average result in higher flows. In contrast, the results of Table (2.3) indicate that the number of tweets does not predict flows. Moreover, our finding in Table (2.7) that positive tweets mitigate outflows is inconsistent with the search cost hypothesis, as investors face no search costs with respect to the funds they already hold.

Second, we explore the notion that asset management companies use social media to convey to investors relevant information that is not available to the public. More specifically, the model of Dumitrescu and Gil-Bazo (2016) of strategic communication by asset managers predicts that asset management companies will communicate information that is favorable for future fund performance and which is not already publicly available. Since in equilibrium communications are truthful and favorable, they should have a positive impact on flows of new money. But such communications should also possess predictive ability with respect to future fund performance. To test this prediction of the model, we regress one-month ahead performance on Positiveness while controlling for past performance and fund and family characteristics that have been documented in the literature to predict performance. We also allow for time, fund, investment category, and family fixed effects. One difficulty that arises with this test is the fact that net performance is partially determined by investors' reaction. In particular, if there are diseconomies of scale in asset management, fund performance will deteriorate as money flows to funds that are expected to outperform (Berk and Green, 2004). To account for that possibility, we control for lagged assets under management as well as recent fund flows.

Therefore, we estimate the regression equation:

$$\hat{\alpha}_{t+1} = \rho_0 + \rho_1 \alpha_{i,t-3:t} + \rho_2 \text{Positiveness}_{i,t} + \rho_3 X_{it} + \delta_{t+1} + \lambda_i + \mu_{cat} + \theta_{fam} + \nu_{i,t+1}, \tag{2.16}$$

where $\alpha_{i,t-3:t}$ is the fund's abnormal return in the previous three months. δ_{t+1}, λ_i, μ_{cat}, and θ_{fam} denote time, fund, investment category, and family fixed effects, respectively. X_{it} is a vector of control variables that includes: fund size; expense ratio; flows (in month t); portfolio turnover; 12-month return volatility; an indicator variable that equals one if the fund charges loads; fund age; family size and family age. Standard errors are clustered at the fund, month, and fund-month levels.

Table (2.9) shows the estimation results. In columns (1) and (2) we use CAPM alphas both as the dependent variable and as a control (in this case, measured over the previous three months), whereas in columns (3) and (4) we use Carhart's (1997) four-factor alphas. In columns (1) and (3) we do not include fund fixed effects, so we are asking whether positive tweets allow investors to identify funds that will outperform their peers in the following months.

Without fund fixed effects, the estimated coefficient on past recent performance is insignificant for CAPM alpha (column 1), but positive and significant for past four-factor alpha (column 3), which suggests that cross-sectional differences in four-factor alphas, but not in CAPM alphas, persist in the short term. When we include fund fixed effects (columns 2 and 4), the estimated coefficient on past performance is negative and significant for both measures of performance, which is consistent with mean reversion in fund performance (Carhart, 1997). Also, fund performance declines in with fund size both in the cross section and in the time series, which is consistent with prior studies (e.g., Chen et al., 2004). Expense ratios are negatively related to performance in the cross-section, but not in the time series. Finally, the family's assets under management are negatively related to fund performance.

In terms of our variable of interest, Positiveness, none of the coefficients on this variable is positive, which implies that funds in families that post more positive tweets do not outperform in the following month either in the cross-section or in the time-series. In fact, all coefficients are negative, although insignificant with the exception of column (2), CAPM alpha with fund fixed effects, where the coefficient is significant at the 10% level. The results therefore do not support the idea that positive tweets help investors select funds that will wither outperform their peers or deliver higher performance than in other periods. In other words, asset management firms do not seem to use social media to convey private information about future fund performance.

2.7 Further evidence of social media as a persuasion tool

If the main purpose of social media communications is not to reduce information asymmetries but to persuade investors to purchase mutual funds, then asset management firms that have more difficulties competing for investors' money in terms of objective signals of performance are the ones that can potentially benefit the most from using social media for persuasion purposes.

To test this conjecture, we estimate again our baseline flow regressions augmented with interactions of Positiveness with several variables that are intended to proxy for asset managers' incentives to use social media to attract flows: Category Share, defined as the total net assets of the fund over the total net assets of all funds in the same Lipper category; % of new funds, defined as the percentage increase in the number of funds competing in the same Lipper category; Family Size, defined as the natural logarithm of the size of the fund family; and Family Flows computed using equation (2.5) based on the family's total net assets and weighted average returns across all funds in the family;

To the extent that persuasion exploits investors' biases, it is natural to think that persuasion is more effective among retail investors than institutional investors. To test this hypothesis, we interact Positiveness with Retail, a dummy variable that equals 1 if all share classes in the fund are distributed among retail investors.

Table (2.6) shows the estimation results. The coefficient on the interaction between Positiveness and Category Share is negative and statistically significant at the 1% significance level. A one standard deviation decrease in Category Share below its mean triples the coefficient con positiveness from 0.052 to 0.14. The coefficient on the interaction term of Positiveness with the % of new funds is negative and statistically significant at the 10% level. The coefficient on the interaction term of Retail with Positiveness is positive and statistically significant at the 5% level. The magnitude of the estimated coefficient on this interaction term suggests that retail investors are three times more sensitive to Positiveness than funds sold to institutional investors (the estimated effect is 0.15 for retail funds and 0.05 for institutional funds). Finally, the coefficient on the interaction terms between Positiveness and Family Size and Family Flows are both negative and significant at the 1% level, consistent with the hypothesis that families that have a harder time competing for investors' money benefit the most from using social media.

In sum, the results of (2.6) are consistent with persuasion being more effective for funds that struggle to compete and those that target retail investors.

2.8 Conclusions

Social media provide asset managers with a powerful tool to circumvent constraints on traditional mandatory disclosures and persuade investors. We find that a positive tone of Twitter posts predicts subsequent an increase in flows to the fund not explained by performance or fund characteristics, that is statistically and economically significant.

In contrast, we do not find that the number of tweets increases investor flows, as we would expect if asset managers were using social media to help investors locate relevant information. We cannot find, either, any evidence that the tone of Twitter communications contains valuable information about future fund performance, as one would expect if social media were used to convey new information to investors.

Therefore, our results suggest that asset management companies use Twitter as a way to improve investors' perception of the quality of their asset management services. Consistently with this hypothesis, we show that the benefit of positive tweets concentrates in families managing funds that struggle to compete and those that cater to retail investors.

Clients of mutual fund management firms could benefit a great deal from the enhanced, more frequent, and easier-to-access information that social media can provide. Instead, the results of our paper suggest that incentives to influence investors' perceptions dominate incentives to alleviate information asymmetries.

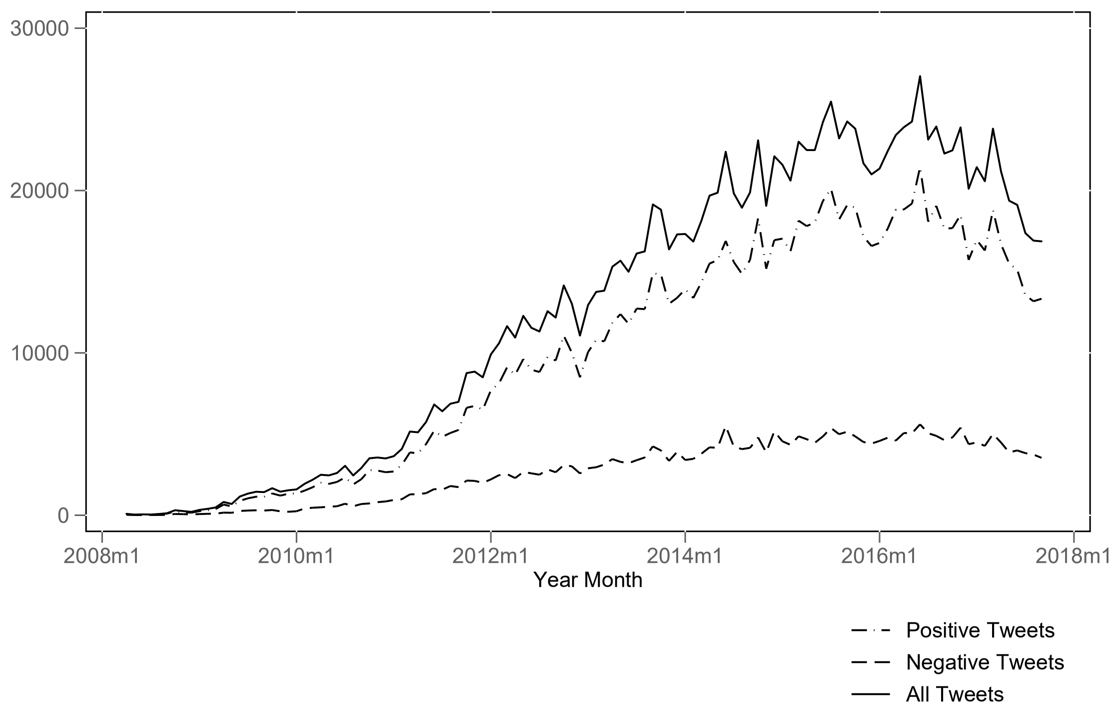


Figure 2.1: Evolution of tweets by fund families through time. The figure shows the number of tweets by fund families per month. The solid black line shows the total number of tweets obtained based on the fund family identifier mgmt_cd for the entire CRSP database. The dot-dashed line and the dashed line represent out of the entire sample the number of tweets classified as positive and negative, respectively.

Table 2.1: Descriptive Statistics, Fund Family characteristics

	Twitter Subsample						Full Sample					
Variable	mean	median	s.d.	1st pctl.	99th pctl.	Obs.	mean	median	s.d.	1st pctl.	99th pctl.	Obs.
Fund Characteristics												
Performance %	-0.13	-0.13	1.50	-4.13	4.14	352,852	-0.14	-0.14	1.63	-4.39	4.36	599,959
Age (Years)	11.52	9.75	9.93	0.50	57.58	396,572	11.40	9.83	9.30	0.50	50.75	716,867
Flows %	0.11	-0.46	6.30	-17.60	26.63	391,877	0.02	-0.50	6.42	-19.82	26.93	707,471
Expense ratio	0.01	0.01	0.01	0.00	0.02	396,572	0.01	0.01	0.01	0.00	0.02	716,867
Total Net Assets (USD Millions)	798.7	136.7	3,023.9	15.70	11,613.5	396,572	663.1	126.1	2,566.3	15.7	9,206	716,867
Front-end Load	0.05	0.06	0.01	0.03	0.06	76,755	0.05	0.06	0.01	0.03	0.06	104,281
Back-end Load	0.02	0.02	0.01	0.00	0.02	38,174	0.02	0.02	0.00	0.00	0.02	78,833
Turnover	0.78	0.48	5.01	0.00	5.24	396,572	0.65	0.38	3.80	0.00	4.96	716,867
Tenure (Months)	94.58	85.00	60.31	8.00	278.00	342,112	96.07	85.00	63.22	8.00	295.00	501,654
Flows Category	-0.00	-0.01	0.24	-0.14	0.19	396,535	-0.00	-0.01	0.81	-0.14	0.17	716,741
Family Characteristics												
Positiveness	0.58	0.00	0.78	-0.36	2.64	16,133						
Number of tweets per month	43.55	0.00	102.31	0.00	468	16,081						
Family Age (Years)	29.47	20.58	24.30	2.83	87.75	16,133	22.12	16.75	19.22	2.75	84.58	47,492
Family Size (USD Billions)	7.46	7.46	2.40	2.86	12.96	16,133	6.40	6.08	2.29	2.80	11.81	47,492
Number of Funds	21.82	7.00	33.95	1.00	155.00	16,133	10.70	2.00	22.86	1.00	110.00	47,492
Number of Investment Categories	5.44	3.00	5.11	1.00	23.00	16,133	3.38	2.00	3.82	1.00	19.00	47,492
Funds with Loads %	27.75	14.67	32.03	0.00	100.00	16,133	31.14	12.50	37.50	0.00	100.00	47,492
Unique fund families						241						785

Note: This table contains summary statistics of fund and fund family characteristics for two samples. The first sample, Twitter Subsample, consists of fund families managing US equity funds that have tweeted at least once from January 2009 to October 2017. The Full Sample includes all families managing US equity funds in the same period. The first set of rows show descriptive statistics for variables computed at the fund-month level. Performance is the monthly CAPM alpha computed from the asset-weighted average return of all share classes of the fund. Age is the number of years since the inception of the oldest share class in the fund. Flows is the fund's monthly growth in the fund's total net assets net of the fund's return. Expense ratio is the asset-weighted average across all share classes of the fund, expressed in decimal units. Total Net Assets is the sum of total net assets of all share classes of the fund. Front-end and back-end loads denote the asset-weighted average of the maximum loads across all share classes. Turnover denotes the annual turnover of the fund's portfolio. Tenure is the number of months since the manager took over the fund. Flows Category denotes the relative flows to all the funds in the same Lipper investment category as the fund in question. The second set of rows display family characteristics. Number of Tweets in the total number of posts on Twitter by the family in a given month. Positiveness is defined as in equation (2.3). Family Age is the age in years of the family's oldest fund. Family Size is the sum of Total Net Assets across all of the family's funds. Number of Funds is the total number of funds managed by the family. Number of Investment Categories is the number of different Lipper investment categories to which the family's funds belong. Funds with loads is the percentage of funds in the family that charge either front-end or back-end load. Expense Ratio is the Asset weighted average of the expense ratios of all funds within each fund family. Volatility is the asset weighted average of the rolling 12 month volatility of the returns of each fund within each family presented in percentage.

Table 2.2: Determinants of Twitter Activity

	(1) Twitter	(2) Twitter	(3) Twitter	(4) # of Tweets	(5) # of Tweets	(6) # of Tweets
Family Age	0.026	0.004	0.023	-0.489***	-0.450***	-0.465***
	(0.027)	(0.027)	(0.026)	(0.077)	(0.080)	(0.083)
Family Size	0.061***			0.110***		
	(0.010)			(0.017)		
Number of Funds		0.157***			0.059*	
		(0.022)			(0.033)	
Inv. Categories			0.231***			0.130**
			(0.038)			(0.050)
CAPM alpha	1.967**	2.053**	2.143**	1.338***	1.428***	1.429***
	(0.921)	(0.899)	(0.901)	(0.342)	(0.348)	(0.348)
Expense Ratio	5.659	1.949	2.717	71.226***	64.056***	63.864***
	(4.206)	(3.967)	(3.978)	(6.114)	(5.885)	(5.904)
Funds with loads	0.035	0.051	0.046	0.149*	0.156*	0.159**
	(0.043)	(0.043)	(0.043)	(0.081)	(0.082)	(0.080)
Volatility	-2.443**	-2.232*	-2.738**	-9.649***	-9.995***	-10.143***
	(1.188)	(1.179)	(1.189)	(1.320)	(1.391)	(1.389)
Observations	687	687	687	15613	15613	15613
Adjusted R^2 (%)	9.24	11.23	9.64	71.66	71.62	71.63
Time FE	No	No	No	Yes	Yes	Yes
Family FE	No	No	No	Yes	Yes	Yes
Estimation	Cross-sect.	Cross-sect.	Cross-sect.	Panel	Panel	Panel
Sample	Full	Full	Full	Twitter	Twitter	Twitter

Note: This table shows estimation results for regressions of Twitter activity on family characteristics. Columns (1) to (3) display results for cross-sectional regressions of a dummy variable that equals 1 if a fund family has tweeted at least once in our sample, on family characteristics. Columns (4) to (6) provide results of running a regression of the natural logarithm of one plus the number of tweets (# of Tweets) posted by a fund family in a given month on fund family characteristics. In columns (1) to (3) explanatory variables are averaged across time for each family. In columns (4) to (6) the unit of observation is family-month and family characteristics are lagged one month. Flows to the family are calculated using equation (2.5) with the Total Net Assets of the fund family and the asset-weighted average return of the family. Family Age is the natural logarithm of one plus the age of the oldest fund in the family in years. Family size is the natural logarithm of all the total net assets managed by the company in USD millions. Number of Funds is the log of the number of funds managed by the family. Inv. Categories denotes the log of the number of different Lipper investment categories across all funds in the family. The CAPM alpha is the asset-weighted average 12-month CAPM alpha across funds in the family. Expense Ratio is is the asset-weighted average of expense ratios across funds in the family. Funds with loads denotes the percentage of funds in the family that charge front-end or back-end loads. Volatility is the asset-weighted average of each fund's 12-month rolling volatility of returns. In columns (1) to (3) robust standard errors are presented in parentheses, while in columns (4) to (6) robust standard errors are clustered at the time level. *** $p<0.01$, ** $p<0.05$, *$p<0.1$.

Table 2.3: Flows and Twitter Activity

	(1)	(2)	(3)	(4)	(5)	(6)
Number of Tweets		0.00			0.00	
		(0.01)			(0.01)	
Positiveness			0.05**			0.05**
			(0.02)			(0.03)
Low OAR	0.47***	0.47***	0.47***			
	(0.04)	(0.04)	(0.04)			
Mid OAR	0.56***	0.56***	0.56***			
	(0.03)	(0.03)	(0.03)			
High OAR	0.96***	0.96***	0.96***			
	(0.05)	(0.05)	(0.05)			
Low Rank				3.29***	3.29***	3.29***
				(0.27)	(0.27)	(0.27)
Mid Rank				1.64***	1.64***	1.64***
				(0.07)	(0.07)	(0.07)
High Rank				8.76***	8.77***	8.78***
				(0.61)	(0.61)	(0.61)
Size	-0.79***	-0.79***	-0.79***	-0.81***	-0.81***	-0.80***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Flows to the same category	-0.24***	-0.24***	-0.25***	-0.04	-0.04	-0.04
	(0.09)	(0.09)	(0.09)	(0.03)	(0.03)	(0.03)
Volatility	-29.88***	-29.88***	-29.84***	-29.42***	-29.42***	-29.38***
	(2.47)	(2.47)	(2.47)	(2.48)	(2.48)	(2.48)
Expense ratio	5.95	5.81	5.27	14.36	14.30	13.71
	(18.43)	(18.43)	(18.44)	(18.54)	(18.54)	(18.56)
Age	-1.54***	-1.54***	-1.55***	-1.57***	-1.57***	-1.57***
	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
Lagged Flows	0.14***	0.14***	0.14***	0.14***	0.14***	0.14***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Family Size	0.39***	0.39***	0.38***	0.38***	0.38***	0.38***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Family Age	-0.10	-0.10	-0.10	-0.11	-0.11	-0.11
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Constant	9.07***	9.06***	9.04***	8.36***	8.36***	8.33***
	(1.17)	(1.17)	(1.17)	(1.17)	(1.18)	(1.17)
Observations	277509	277509	277509	277542	277542	277542
Adjusted R^2 (%)	14.82	14.82	14.82	14.68	14.68	14.68
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Inv. Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Fund FE	Yes	Yes	Yes	Yes	Yes	Yes
Fund Family FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table shows estimation results for regressions of mutual fund flows (in %) on Number of Tweets, Positiveness, and control variables. Flows are computed using equation (2.5). Number of Tweets is computed as the natural logarithm of one plus the number of tweets posted by the fund's fund family in the previous month. Positiveness is computed in the previous month as in equation (2.3). Low, Mid, and High Performance are computed using objective adjusted 12-month CAPM alphas (normalized to have zero average and unit standard deviation across all funds with the same Lipper class investment objective). Low, Mid, and High Ranks are computed using ranks based on the 12-month CAPM alpha of each fund under the same Lipper class investment objective. Size is the natural logarithm of the total net assets under management of a fund in the previous month. Flows to the same category are computed as the percentage of contemporaneous net flows to all funds with the same Lipper class investment objective. Volatility is the 12 month rolling volatility of returns. Expense Ratio is in decimal units. Age is the natural logarithm of the age of the fund in months. Lagged Flows denotes one-month lagged flows to the fund. Family size is the natural logarithm of the assets under management by the fund family in the previous month, and family age is the natural logarithm of one plus the age of the family in months. Robust standard errors, clustered at the month, fund family, and month-fund family levels, are shown in parentheses. *** $p<0.01$, ** $p<0.05$, *$p<0.1$.

Table 2.4: Flows and Positiveness (Fund family level)

	(1)	(2)	(3)	(4)
Positiveness	0.19***	0.18***	0.22***	0.21***
	(0.03)	(0.04)	(0.03)	(0.04)
Low OAR - Family	0.71***	0.80***		
	(0.05)	(0.06)		
Mid OAR - Family	0.38***	0.34***		
	(0.05)	(0.05)		
High OAR - Family	0.83***	0.89***		
	(0.07)	(0.07)		
Low Rank - Family			-2.29***	-1.67***
			(0.50)	(0.52)
Mid Rank - Family			2.38***	2.05***
			(0.16)	(0.17)
High Rank - Family			5.86***	7.91***
			(0.99)	(1.01)
Family Size	-0.00	-0.15***	0.00	-0.13**
	(0.01)	(0.06)	(0.01)	(0.06)
Volatility - Family	-4.29*	-25.66***	-8.53***	-22.87***
	(2.57)	(4.23)	(2.52)	(4.25)
Expense Ratio - Family	-22.35***	74.20***	-25.63***	86.31***
	(6.85)	(22.95)	(6.85)	(23.02)
Age - Family	-0.17***	-0.97***	-0.25***	-1.28***
	(0.03)	(0.15)	(0.03)	(0.15)
Flows Category - Family	-0.05	-0.10	-0.06	-0.10
	(0.09)	(0.10)	(0.09)	(0.10)
Lagged Flows - Family	0.28***	0.20***	0.30***	0.21***
	(0.01)	(0.01)	(0.01)	(0.01)
Constant	1.16***	6.12***	1.31***	6.76***
	(0.20)	(0.81)	(0.21)	(0.82)
Observations	45100	45095	45100	45095
Adjusted R^2 (%)	14.14	18.62	13.33	17.79
Time FE	Yes	Yes	Yes	Yes
Fund Family FE	No	Yes	No	Yes

Note: This table shows estimation results for regressions of mutual fund flows (in %) on Positiveness and control variables. Flows are computed using equation (2.5). Positiveness is computed in the previous month as in equation (2.3). Family-level aggregated variables are asset-weighted averages across funds in the family, except for Family Size, which is the natural logarithm of the sum of all the TNAs among the funds in the family, and Family Age, which is the natural logarithm of one plus the age of the oldest fund in months. Low, Mid, and High Performance are computed using objective adjusted 12-month CAPM alphas (normalized to have zero average and unit standard deviation across all funds with the same Lipper class investment objective). Low, Mid, and High Ranks are computed using ranks based on the 12-month CAPM alpha of each fund under the same Lipper class investment objective. Size is the natural logarithm of the total net assets under management of a fund in the previous month. Flows to the same category are computed as the percentage of contemporaneous net flows to all funds with the same Lipper class investment objective. Volatility is the 12 month rolling volatility of returns. Expense Ratio is in decimal units. Age is the natural logarithm of the age of the fund in months. Lagged Flows denotes one-month lagged flows to the fund. Family size is the natural logarithm of the assets under management by the fund family in the previous month, and family age is the natural logarithm of one plus the age of the family in months. Robust standard errors, clustered at the month, fund family, and month-fund family levels, are shown in parentheses. *** $p<0.01$, ** $p<0.05$, *$p<0.1$.

Table 2.5: Flows, Positiveness and Other Information

	(1)	(2)	(3)	(4)	(5)
Positiveness	0.05*	0.05*	0.05*	0.05*	0.05*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Change Family	-0.31				
	(0.32)				
Change Manager		-0.36			
		(0.65)			
Number of external tweets			0.03		
			(0.04)		
External Positiveness				0.01	
				(0.05)	
Fraction of stars in Family					0.56***
					(0.15)
Low Rank	3.29***	3.29***	3.29***	3.29***	3.28***
	(0.27)	(0.27)	(0.27)	(0.27)	(0.27)
Mid Rank	1.64***	1.64***	1.64***	1.64***	1.63***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
High Rank	8.75***	8.75***	8.75***	8.75***	8.66***
	(0.61)	(0.61)	(0.61)	(0.61)	(0.61)
Size	-0.81***	-0.81***	-0.81***	-0.81***	-0.81***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Flows to the same category	-0.04	-0.04	-0.04	-0.04	-0.04
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Volatility	-28.41***	-28.41***	-28.41***	-28.42***	-28.53***
	(2.48)	(2.48)	(2.48)	(2.48)	(2.48)
Expense ratio	17.37	17.39	17.58	17.36	17.32
	(18.36)	(18.36)	(18.37)	(18.36)	(18.36)
Age	-1.61***	-1.61***	-1.61***	-1.61***	-1.61***
	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
Lagged Flows	0.14***	0.14***	0.14***	0.14***	0.14***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Family Size	0.39***	0.39***	0.39***	0.39***	0.40***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Family Age	-0.14	-0.14	-0.14	-0.14	-0.14
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Observations	277542	277542	277542	277542	277542
Adjusted R^2 (%)	14.63	14.63	14.63	14.63	14.64
Time FE	Yes	Yes	Yes	Yes	Yes
Fund Family FE	Yes	Yes	Yes	Yes	Yes
Fund FE	Yes	Yes	Yes	Yes	Yes

Note: This table shows estimation results for regressions of mutual fund flows (in %) on Positiveness, variables that capture fundamental information about the fund, and control variables. Flows are computed using equation (2.5). Positiveness is computed in the previous month as in equation (2.3). Change Family is a dummy that equals 1 if the CRSP identifier mgmt_cd changes in the previous. Change Manager is a dummy that takes the value of one if the name of the manager provided by CRSP changes in the previous month. The number of external tweets is defined as the natural logarithm of one plus the number of third-party tweets mentioning the fund family in the previous month. External Positiveness is defined as in equation (2.3) using only third-party tweets that mention a fund family. Fraction of Star Funds is the number of funds inside the fund family that are in the top 5 % of one-month CAPM alpha within their investment objective category in the previous month, divided by the total number of funds under management. Low, Mid, and High Performance are computed using objective adjusted 12-month CAPM alphas (normalized to have zero average and unit standard deviation across all funds with the same Lipper class investment objective). Low, Mid, and High Ranks are computed using ranks based on the 12-month CAPM alpha of each fund under the same Lipper class investment objective. Size is the natural logarithm of the total net assets under management of a fund in the previous month. Flows to the same category are computed as the percentage of contemporaneous net flows to all funds with the same Lipper class investment objective. Volatility is the 12 month rolling volatility of returns. Expense Ratio is in decimal units. Age is the natural logarithm of the age of the fund in months. Lagged Flows denotes one-month lagged flows to the fund. Family size is the natural logarithm of the assets under management by the fund family in the previous month, and family age is the natural logarithm of one plus the age of the family in months. Robust standard errors, clustered at the month, fund family, and month-fund family levels, are shown in parentheses.

Table 2.6: Flows, Positiveness and Timing

	(1)	(2)	(3)	(4)	(5)
Positiveness	0.07***	0.07***	0.05	0.42***	0.04
	(0.03)	(0.03)	(0.04)	(0.13)	(0.03)
Low OAR	0.52***	0.52***	0.55***	0.52***	0.52***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Mid OAR	0.63***	0.63***	0.81***	0.63***	0.63***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
High OAR	1.07***	1.07***	1.08***	1.07***	1.08***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Size	-0.72***	-0.73***	0.01	-0.72***	-0.73***
	(0.04)	(0.04)	(0.01)	(0.04)	(0.04)
Flows to the same category	-0.14	-0.20**	-0.32***	-0.17**	-0.18**
	(0.09)	(0.09)	(0.07)	(0.09)	(0.09)
Volatility	-34.48***	-34.42***	-21.73***	-34.59***	-34.51***
	(2.52)	(2.51)	(2.13)	(2.52)	(2.51)
Expense ratio	11.75	12.39	-71.37***	11.91	11.00
	(19.06)	(19.05)	(3.74)	(19.06)	(19.07)
Age	-1.90***	-1.85***	-0.62***	-1.89***	-1.88***
	(0.16)	(0.16)	(0.03)	(0.16)	(0.16)
Family Size	0.38***	0.38***	-0.05	0.42***	0.38***
	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)
Family Age	-0.13	-0.13	-0.22**	-0.18	-0.12
	(0.11)	(0.11)	(0.11)	(0.12)	(0.11)
Last Quarter Growth	2.82***	2.82***	4.34***	2.82***	2.82***
	(0.67)	(0.67)	(0.98)	(0.67)	(0.67)
Category Share	0.72				
	(0.68)				
Positiveness × Category Share	-2.10***				
	(0.49)				
% of new funds		0.05***			
		(0.02)			
Positiveness × % of new funds		-0.02*			
		(0.01)			
Retail			-0.04		
			(0.05)		
Positiveness × Retail			0.10**		
			(0.04)		
Positiveness × Family Size				-0.04***	
				(0.01)	
Family Flows					4.93***
					(1.01)
Positiveness × Family Flows					-1.29**
					(0.57)
Constant	10.59***	10.34***	5.68***	10.54***	10.55***
	(1.18)	(1.18)	(0.80)	(1.18)	(1.18)
Observations	277509	277509	277542	277509	277467
Adjusted R^2 (%)	13.4	13.39	7.04	13.39	13.41
Time FE	Yes	Yes	Yes	Yes	Yes
Inv. Category FE	Yes	Yes	Yes	Yes	Yes
Fund FE	Yes	Yes	No	Yes	Yes
Fund Family FE	Yes	Yes	Yes	Yes	Yes

Note: This table shows estimation results for regressions of mutual fund flows (in %) on Positiveness, variables that capture fundamental information about the fund, control variables, and variables that capture the timing of . Flows are computed using equation (2.5). Positiveness is computed in the previous month as in equation (2.3). Change Family is a dummy that equals 1 if the CRSP identifier mgmt_cd changes in the previous. Change Manager is a dummy that takes the value of one if the name of the manager provided by CRSP changes in the previous month. The number of external tweets is defined as the natural logarithm of one plus the number of third-party tweets mentioning the fund family in the previous month. External Positiveness is defined as in equation (2.3) using only third-party tweets that mention a fund family. Fraction of Star Funds is the number of funds inside the fund family that are in the top 5 % of one-month CAPM alpha within their investment objective category in the previous month, divided by the total number of funds under management. Low, Mid, and High Performance are computed using objective adjusted 12-month CAPM alphas (normalized to have zero average and unit standard deviation across all funds with the same Lipper class investment objective). Low, Mid, and High Ranks are computed using ranks based on the 12-month CAPM alpha of each fund under the same Lipper class investment objective. Size is the natural logarithm of the total net assets under management of a fund in the previous month. Flows to the same category are computed as the percentage of contemporaneous net flows to all funds with the same Lipper class investment objective. Volatility is the 12 month rolling volatility of returns. Expense Ratio is in decimal units. Age is the natural logarithm of the age of the fund in months. Lagged Flows denotes one-month lagged flows to the fund. Family size is the natural logarithm of the assets under management by the fund family in the previous month, and family age is the natural logarithm of one plus the age of the family in months. Last Quarter growth is defined as the percentage increase of total net assets in the last three months. The market excess return is defined as the factor *mktrf* from Keneth French's website. Category Share is defined as the total net assets of the fund over the total net assets of all funds in the same lipper category. Family share is defined as the total net assets of the fund over the total net assets of the fund family. % of new funds is defined as the percentage increase in the number of funds of the same lipper category in the sample. Robust standard errors, clustered at the month, fund family, and month-fund family levels, are shown in parentheses.

Table 2.7: Inflows, Outflows, and Positiveness

	Inflows		Outflows	
	(1)	(2)	(3)	(4)
Positiveness	0.10***	0.08***	-0.08***	-0.07***
	(0.02)	(0.02)	(0.01)	(0.01)
Low OAR	0.27***		-0.54***	
	(0.03)		(0.04)	
Mid OAR	0.64***		-0.35***	
	(0.03)		(0.02)	
High OAR	0.98***		-0.05	
	(0.07)		(0.03)	
Low Rank		2.13***		-4.51***
		(0.26)		(0.28)
Mid Rank		2.05***		-0.77***
		(0.08)		(0.06)
High Rank		4.59***		-0.60
		(0.91)		(0.49)
Observations	44342	44351	44342	44351
Adjusted R^2 (%)	5.04	3.87	2.67	2.76

Note: This table shows estimation results for regressions of residual inflows and residual outflows (both in %) on Positiveness and fund performance. In a first stage (not reported) Inflows (Outflows) are regressed on 12 lags of the variable and controls: Size, the natural logarithm of the total net assets under management of a fund in the previous month; Flows to the same category, the percentage of contemporaneous net flows to all funds with the same Lipper class investment objective; Volatility, the 12 month rolling volatility of returns; Expense Ratio, in decimal units; Age, the natural logarithm of the age of the fund in months; one-month lagged flows to the fund; Family size, the natural logarithm of the assets under management by the fund family in the previous month; and Family age, the natural logarithm of one plus the age of the family in months. In a second stage (not reported) we regress Inflows (Outflows) on the fitted values of Ouflows (Inflows) estimated in the first stage. Fitted residuals from the regression are then regressed on Positiveness and the three performance variables. Positiveness is computed in the previous month as in equation (2.3). Low, Mid, and High Performance are computed using objective adjusted 12-month CAPM alphas (normalized to have zero average and unit standard deviation across all funds with the same Lipper class investment objective). Low, Mid, and High Ranks are computed using ranks based on the 12-month CAPM alpha of each fund under the same Lipper class investment objective. OLS robust standard errors in parentheses. *** p<0.01, ** p<0.05, *p<0.1.

Table 2.9: Predictive Regressions

	(1) $\alpha_{t\rightarrow t+12}$	(2) $\alpha_{t\rightarrow t+36}$	(3) $\alpha_{t\rightarrow t+48}$	(4) $\alpha_{t\rightarrow t+60}$
Positiveness$_t$	0.013	-0.148***	-0.258***	-0.323***
	(0.017)	(0.034)	(0.044)	(0.055)
$\alpha_{t-3\rightarrow t}$	-0.028***	0.070***	0.073***	0.105***
	(0.005)	(0.009)	(0.011)	(0.012)
Size	-0.099***	-0.347***	-0.438***	-0.559***
	(0.010)	(0.021)	(0.026)	(0.033)
Expense ratio	-0.412***	-1.295***	-1.677***	-2.011***
	(0.011)	(0.021)	(0.026)	(0.032)
Past Flows	0.017*	0.063***	0.059**	0.008
	(0.010)	(0.019)	(0.024)	(0.029)
Turnover	-0.064***	-0.123***	-0.166***	-0.086**
	(0.010)	(0.022)	(0.027)	(0.034)
Volatility	-37.321***	-170.662***	-252.267***	-317.325***
	(1.714)	(3.075)	(3.605)	(4.009)
Family Size	-0.112***	-0.063*	-0.143***	-0.116*
	(0.016)	(0.035)	(0.043)	(0.061)
Family Age	0.040	-0.010	0.068	-0.041
	(0.040)	(0.084)	(0.111)	(0.184)
Charges Loads	0.012	0.063***	0.103***	0.142***
	(0.009)	(0.018)	(0.022)	(0.027)
Fund Age	0.014	0.023	-0.045*	-0.092***
	(0.011)	(0.021)	(0.027)	(0.034)
Constant	-0.142*	2.490***	5.488***	9.227***
	(0.077)	(0.151)	(0.192)	(0.238)
Observations	254375	166729	130398	98268
Adjusted R^2 (%)	17.99	32.36	38.29	43.02
Time FE	Yes	Yes	Yes	Yes
Investment Category FE	Yes	Yes	Yes	Yes
Fund FE	No	No	No	No
Fund Family FE	Yes	Yes	Yes	Yes

Table 2.8: Predictive Regressions

	(1) α_{t+1}	(2) α_{t+1}	(3) α^{4F}_{t+1}	(4) α^{4F}_{t+1}
Positiveness_t	-0.008	-0.009*	-0.003	-0.005
	(0.005)	(0.005)	(0.004)	(0.004)
$\alpha_{t-3:t}$	-0.001	-0.014***		
	(0.002)	(0.002)		
$\alpha^{4F}_{t-3:t}$			0.012***	-0.005***
			(0.002)	(0.002)
Size	-0.011***	-0.176***	-0.005**	-0.114***
	(0.003)	(0.008)	(0.002)	(0.006)
Expense ratio	-0.035***	0.068***	-0.032***	0.013
	(0.003)	(0.016)	(0.002)	(0.012)
Lagged Flows	0.002	-0.009***	0.008***	-0.001
	(0.003)	(0.003)	(0.002)	(0.002)
Turnover	0.003	0.032***	-0.010***	0.013***
	(0.003)	(0.005)	(0.002)	(0.004)
Volatility	-2.507***	3.331***	-6.429***	-4.532***
	(0.560)	(0.670)	(0.418)	(0.488)
Charges Loads	0.002	0.011	0.000	-0.024**
	(0.003)	(0.016)	(0.002)	(0.011)
Fund Age	0.001	0.029**	0.002	0.029***
	(0.003)	(0.012)	(0.002)	(0.009)
Family Size	-0.103***	-0.100***	-0.055***	-0.053***
	(0.005)	(0.005)	(0.003)	(0.003)
Family Age	0.025**	0.021*	0.023**	0.022**
	(0.012)	(0.012)	(0.009)	(0.009)
Constant	-0.023	-0.270***	0.143***	0.061***
	(0.024)	(0.028)	(0.018)	(0.021)
Observations	301443	301406	301443	301406
Adjusted R^2 (%)	11.63	11.84	9.77	10.25
Time FE	Yes	Yes	Yes	Yes
Investment Category FE	Yes	Yes	Yes	Yes
Fund FE	No	Yes	No	Yes
Fund Family FE	Yes	Yes	Yes	Yes

Note: This table shows estimation results of regressions of fund's monthly alpha (in %) on Positiveness, past performance, and control variables. Positiveness is defined as in equation (2.3). Performance is defined in two ways: The CAPM abnormal return compounded over the last three months $\alpha(t-3:t)$ and the four-factor abnormal return compounded over the last three months $\alpha(t-3:t)$. Size is the natural logarithm of the total net assets of the fund. The expense ratio is in decimal units. Lagged Flows are the net flows to the fund in the previous month. Turnover is the fund's portfolio turnover. Volatility is the 12-month rolling volatility of returns. Charges Loads is a dummy that equals 1 if the fund charges eith front-end or back-end loads. Age is the natural logarithm of the age of the fund in months. Family size is the natural logarithm of the total net assets of the fund family. Family Age is the natural logarithm of the age of the family in months. Robust standard errors clustered at the family, month, and family-month levels are presented in parentheses. *** p<0.01, ** p<0.05, *p<0.1.

Table 2.10: Flows, Positiveness, and Family Characteristics

	(1)	(2)	(3)	(4)	(5)
Positiveness	0.07***	0.07***	0.05	0.42***	0.04
	(0.03)	(0.03)	(0.04)	(0.13)	(0.03)
Low OAR	0.52***	0.52***	0.55***	0.52***	0.52***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Mid OAR	0.63***	0.63***	0.81***	0.63***	0.63***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
High OAR	1.07***	1.07***	1.08***	1.07***	1.08***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Size	-0.72***	-0.73***	0.01	-0.72***	-0.73***
	(0.04)	(0.04)	(0.01)	(0.04)	(0.04)
Flows to the same category	-0.14	-0.20**	-0.32***	-0.17**	-0.18**
	(0.09)	(0.09)	(0.07)	(0.09)	(0.09)
Volatility	-34.48***	-34.42***	-21.73***	-34.59***	-34.51***
	(2.52)	(2.51)	(2.13)	(2.52)	(2.51)
Expense ratio	11.75	12.39	-71.37***	11.91	11.00
	(19.06)	(19.05)	(3.74)	(19.06)	(19.07)
Age	-1.90***	-1.85***	-0.62***	-1.89***	-1.88***
	(0.16)	(0.16)	(0.03)	(0.16)	(0.16)
Family Size	0.38***	0.38***	-0.05	0.42***	0.38***
	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)
Family Age	-0.13	-0.13	-0.22**	-0.18	-0.12
	(0.11)	(0.11)	(0.11)	(0.12)	(0.11)
Last Quarter Growth	2.82***	2.82***	4.34***	2.82***	2.82***
	(0.67)	(0.67)	(0.98)	(0.67)	(0.67)
Category Share	0.72				
	(0.68)				
Positiveness × Category Share	-2.10***				
	(0.49)				
% of new funds		0.05***			
		(0.02)			
Positiveness × % of new funds		-0.02*			
		(0.01)			
Retail			-0.04		
			(0.05)		
Positiveness × Retail			0.10**		
			(0.04)		
Positiveness × Family Size				-0.04***	
				(0.01)	
Family Flows					4.93***
					(1.01)
Positiveness × Family Flows					-1.29**
					(0.57)
Constant	10.59***	10.34***	5.68***	10.54***	10.55***
	(1.18)	(1.18)	(0.80)	(1.18)	(1.18)
Observations	277509	277509	277542	277509	277467
Adjusted R^2 (%)	13.4	13.39	7.04	13.39	13.41
Time FE	Yes	Yes	Yes	Yes	Yes
Inv. Category FE	Yes	Yes	Yes	Yes	Yes
Fund FE	Yes	Yes	No	Yes	Yes
Fund Family FE	Yes	Yes	Yes	Yes	Yes

Note: This table shows estimation results of regressions of fund flows (in %) on Positiveness, past performance, and controls. Flows are computed using equation (2.5). Positiveness is computed in the previous month as in equation (2.3). Low, Mid, and High Performance are computed using objective adjusted 12-month CAPM alphas (normalized to have zero average and unit standard deviation across all funds with the same Lipper class investment objective). Low, Mid, and High Ranks are computed using ranks based on the 12-month CAPM alpha of each fund under the same Lipper class investment objective. Size is the natural logarithm of the total net assets under management of a fund in the previous month. Flows to the same category are computed as the percentage of contemporaneous net flows to all funds with the same Lipper class investment objective. Volatility is the 12 month rolling volatility of returns. Expense Ratio is in decimal units. Age is the natural logarithm of the age of the fund in months. Lagged Flows denotes one-month lagged flows to the fund. Family size is the natural logarithm of the assets under management by the fund family in the previous month, and family age is the natural logarithm of one plus the age of the family in months. Number of Funds denotes the number of funds managed by the family in the previous month (in logs). Number of Investment Categories is the number of different investment categories to which the family's funds belong. All four family-level variables are cross-sectionally standardized. Robust standard errors, clustered at the month, fund family, and month-fund family levels, are shown in parentheses. *** $p<0.01$, ** $p<0.05$, *$p<0.1$.

Chapter 3

Learning from Quant (Qual)-itative Information

I develop a model in which mutual fund investors learn about managerial ability from past returns and a qualitative signal. When investors observe past returns, they update their beliefs following Bayes rule, but follow a more general Pseudo-Bayesian rule after observing the qualitative signal. I study how capital allocation is affected when investors learn about managerial ability from the qualitative signal and the strategic transmission of qualitative information by fund managers. My model predicts that: flows are increasing in the tone of the qualitative signal, reputation costs decrease the probability of investors verifying information, and verification costs and risk aversion decrease the probability of funds manipulating information.

3.1 Introduction

Rational investors use all relevant information to make an investment decision. This includes both quantitative and qualitative information. If qualitative information conveys relevant knowledge about the future, investors should include this information when forming expectations. Despite the theoretical relevance that qualitative information has, we do not fully understand its role in financial markets.[1] I contribute to this debate by presenting a rational model in which mutual fund investors learn from both a qualitative and a quantitative signal about managerial ability. Rational models about mutual fund investors come in different flavours, but

[1]Qualitative information such as: press, analyst reports, share holder letters, tweets, conference calls among many others are available for investors at a low cost.

most of them consider investors learning about the ability of managers to generate positive abnormal returns (managerial ability). Since the pioneer work of Berk and Green (2004) the theoretical literature on active management studies how capital is allocated among funds as the result of investors learning about managerial ability. In this literature, the learning process is modelled assuming bayesian-investors that update their beliefs after observing the realization of a quantitative signal (e.g. fund returns), and invest in those funds that provide a better risk adjusted expected return.

This paper will address the following questions: How do rational investors learn about managerial ability when qualitative information is available? What is the role of qualitative information in capital allocation? How do fund managers optimally transmit this information to investors? To the best of my knowledge my model is the first to explicitly include qualitative information in the theoretical debate of how mutual fund investors learn about managerial ability, and how fund managers strategically disclose this information.

To address these questions I borrow from the *axiomatic decision making* literature a framework called *pseudo-bayesian updating* (Zhao 2020). Under this paradigm, qualitative information is modelled as a statement about the likelihood of future events. Consider an investor reading news about a fund manager winning an award, her favourite fund earning an extra Mornigstar star, or an article about the growth of cheap passive investments. Under this framework, the investor translates any of these statements into a relation of the form "Event A is more likely than event B". For example: "It is more likely that the fund's abnormal expected return (α) is above the average of the cohort of managers than below it". If the statement can be expressed using events that have a probability assigned investors update their beliefs by choosing the closest probability distribution such that the *qualitative statement* is included in the new distribution.[2] My model only includes the role of qualitative information in a friction-less environment, therefore failing to capture the convex relation between past performance and net flows as documented in the literature (Sirri and Tufano 1998, Chevalier and Ellison 1997).

The model assumes investors observe first the return of the fund (quantitative signal) and after updating their beliefs following Bayes rule, they observe a qualitative signal and update their beliefs following the *pseudo-bayesian updating* rule. First I consider a variation of the friction-less model of Berk and Green (2004) in which capital flows inelastically to funds with

[2]Formally the events must be measurable w.r.t. the investor's posterior distribution.

positive expected abnormal returns, and there are dis-economies of scale in the fund's technology that decrease these expected returns toward zero. I depart from Berk and Green (2004) assuming investors observe a public qualitative signal about a fund, after observing its return. I find that fund flows react positively to the tone of the qualitative information, and the shape of the performance flow relation changes in the presence of qualitative information.

Next I study the investor's capital allocation decision in a model similar to Huang et al. (2007) in which investors allocate capital to a risk free asset and a mutual fund that provides a risky return. I study how the qualitative signal affects the posterior distribution of investors and therefore their optimal capital allocation. Even though it is not possible to obtain a closed form solution of the optimal portfolio choice I derive a system of equations that can be solved numerically to understand how capital allocation reacts to the qualitative signal. This set-up is then extended by explicitly modelling the strategic transmission of information by fund managers.

Since qualitative information is *soft* information, it is easy to manipulate by fund managers in an attempt to receive higher flows and therefore more fees. To model this situation I develop a game in which investors verify the qualitative signal issued by fund managers at a cost. If investors verify the qualitative signal and find that indeed fund managers have manipulated it, funds face a reputation cost. [3] However, if investors process a signal that has been manipulated at face-value, they maximize their expected utility as if the signal issued by managers was the *true* signal, leading to over-investment in the fund and a loss in expected utility. In the mixed Nash equilibrium, the probability of investors verifying the signal is decreasing in the fund's reputation costs and increasing in the opportunity cost, faced by managers, of not manipulating the signal. On the other hand, the probability of funds manipulating the qualitative signal increases with the costs faced by investors when verifying the signal, and decreases with the loss of expected utility faced by investors when the fund managers manipulate the signal and investors interpret it at face-value.

My model contributes to the theoretical literature on the determinants of mutual funds (see Christoffersen et al. (2014) for an empirical and theoretical review). The theoretical literature of investors learning about mutual fund ability dates back to the attempt of Berk and Green (2004) to

[3]The reputation cost models the fact that if investors notice that a fund is manipulating information they would invest less in that fund in the future.

reconcile two empirical observations: Flows are increasing on past performance, and performance is not persistent over time (Carhart 1997, Sirri and Tufano 1998, Chevalier and Ellison 1997, Ippolito 1992). The authors reconcile these observations by modelling bayesian-investors that learn about managerial ability from past returns, and dis-economies of scale. In this framework, funds with higher returns receive more flows since the perceived managerial ability increases, and performance is not persistent since dis-economies of scale cause extra flows to decrease the expected return to zero.

There is evidence of convexity in the performance-flow relation (Sirri and Tufano 1998, Chevalier and Ellison 1997). Money flows more to recent winners and it fails to escape from recent losers. This observation can arise in rational models with the presence of market frictions. Huang et al. (2007) show how this convexity can arise when there are participation costs to invest in mutual funds. In their model, new investors need to pay a cost in order to learn about the *prior* distribution of managerial ability -how the alpha of a fund is distributed- and will only pay the cost if the perceived benefit of investing is greater than the participation cost. This benefit is increasing on past performance, which leads to funds with recent good performance receiving disproportionately more flows. Lynch and Musto (2003) show that this convex relation can be the result of fund incentives. Funds discard those strategies that underperform. Bad performance is less informative about future returns when funds discard these strategies, so flows are less sensitive to performance when they are poor. Pastor and Stambaugh (2012) model investors that learn jointly about managerial ability and decreasing returns to scale. Their model extends the idea of Berk and Green (2004) to assume investors learn about how much returns decrease when the size of the active management industry increases. The authors reconcile the observation that the size of the active management industry remains large despite its performance remaining poor compared to passive strategies. Dumitrescu and Gil-Bazo (2017) show that by including market frictions in the model of Berk and Green (2004) differences in the performance of funds are likely to persist, and those funds targeted to less sophisticated investors exhibit higher dispersion in expected performance.

The model of Dumitrescu and Gil-Bazo (2016) is the closest to my model. In their paper, the authors model investors that learn about two different quantitative signals, the return of the fund, and a non-return signal that captures everything from which an investor can learn. The main difference between their model and mine, is that I model the qualitative signal explicitly, and rely on a *pseudo-bayesian* updating rule to incorporate this signal

into the investors posterior. The capital allocation results of both models are similar but, in my model, the presence of qualitative information creates some room for its manipulation. On the contrary, in Dumitrescu and Gil-Bazo (2016), fund managers can only decide whether or not to issue the signal not whether to manipulate it. My model also contributes to the literature on non-bayesian investors. While my model assumes rational investors, the literature on behavioural non-bayesian investors is wide, covering topics like investor sentiment (Barberis et al. 1998), confirmation bias (Rabin and Schrag 1999), bounded rationality (Mullainathan 2002), law of small numbers (Rabin 2002), coarsely thinking (Mullainathan et al. 2008) or local thinking (Gennaioli and Shleifer 2010) among others.
The rest of the paper is organized as follows: Section 3.2.1 explains how qualitative information can be included in the learning process by investors. Section 3.2.2 extends the model of Berk and Green (2004) to understand how qualitative information affects capital allocation in a setup in which capital supply is perfectly inelastic and there is perfect competition. Section 3.2.3 studies the role of qualitative information by modelling the portfolio decision of investors extending the model of Huang et al. (2007). Section 3.3 studies the strategic transmission of qualitative information by fund managers. Section 3.4 provides concluding remarks. All proofs are presented in the appendix.

3.2 The Model

3.2.1 Preliminaries

The literature on non-bayesian updating models consider decision makers (e.g. investors) that would normally use bayes rule, but deviate from it due to some bias like temptation or bounded rationality. In this paper I use the *pseudo-bayesian* updating rule proposed recently by Zhao (2020) which generalizes the notion of bayesian updating to consider more general information. The main difference between the model of Zhao (2020) and models in the non-bayesian literature is the assumption of rationality.Even with qualitative information investors are able to behave rationally just like if they were bayesian. I impose this assumption in my model to study the effect of qualitative information in the light of standard rational models of mutual funds.

The starting point of my model comes from the *axiomatic decision making* literature. I borrow both the definition of a qualitative signal, and the way investors rationally learn about it. I follow Zhao (2020) who proposes an axiomatic framework for belief revision when qualitative information is of the form "event A is more likely than event B". In his framework, the decision maker selects the *closest* posterior such that the probability of A is greater than the probability of B encoded in the tuple (A, B). Under the pseudo-bayesian framework, the investor updates his posterior by choosing the closest posterior distribution such that the probability of A is greater than the probability of B. Formally the investor solves the following program

$$\begin{aligned} &\min_{g \ll f} d(f||g) \\ &\text{s.t.} \\ &g(A) \geqslant g(B) \end{aligned} \tag{3.1}$$

Where d is a statistical distance[4] between probability distributions, f is the prior distribution, g is the posterior distribution, and $g \ll f$ means that g is absolutely continuous with respect to f. This framework can also accommodate statements of the form "event A has occurred", with the tuple $(\varnothing, A^c)$ where $\varnothing$ is the empty set and A^c is the complement of A. Zhao (2020, Theorem 4) shows that quantitative information can be included into this setup with a collection of qualitative statements.

The author imposes two axioms on the process of investor learning: 1) *Conservatism* (Investors modify their prior as little as possible to include the qualitative statement) and 2) *Orthogonality* (The order in which investors receive the signals does not affect the final posterior distribution). The author shows that the Kullback-Leibler divergence is the only distance that preserves these two axioms. Since in my model I impose explicitly the order in which the investor receives the signals, I only rely on axiom 1 (*Conservatism*) which can be attained by any well defined metric in C^2[5]. I consider a modification of the Bhattacharyya distance between probability functions, instead of the original Kullback-Leibler divergence. This distance makes some of the algebra less cumbersome when working with normal priors, and provides the same intuition. In the model the distance between two functions $f(x), g(x)$ is defined as the negative of the Bhattacharyya Index.

$$d(f, g) = -2 \int_{-\infty}^{\infty} \sqrt{f(x)g(x)}dx \tag{3.2}$$

[4] Zhao (2020) uses the Kullback-Leibler divergence or maximum entropy, however as it will be discussed later, other well defined distances lead to the same posterior distribution.

[5] The space of functions with continuous first and second derivatives.

I also depart from Zhao (2020) in the way the qualitative signal is received by investors. In my model investors receive a signal of the form "Event A is equally likely than event B". If A and B are measurable events we can find an event A' such that $Pr(A) \geqslant Pr(B) \rightarrow Pr(A') = Pr(B)$. In my model, given a prior p.d.f. f, the investor chooses the posterior g that minimizes (3.2) such that $G(A) = G(B)$ where for any measurable event E, $G(E) = \int_E g(x)dx$. Minimizing the distance between the prior and the posterior is consistent with Axiom 2 (*Conservatism*) in Zhao (2020), investors adjust their beliefs no more than necessary to include the new information available. For the rest of the paper I assume fund returns (before fees and transaction costs) satisfy the following stochastic process:

$$R_t = \alpha + \epsilon_t \tag{3.3}$$

Where α is the true managerial ability unknown by fund managers and investors, and ϵ_t is white noise that represents the idiosyncratic risk of the fund investments. R_t can be understood as the return over a benchmark or passive portfolio with the same systematic risk. Investors have a prior distribution on α and after receiving a qualitative statement they update their beliefs following the pseudo-bayesian updating rule discussed above. Given the prior distribution on $\alpha \sim N(\mu, \sigma^2)$ we define the qualitative statement investors receive as:

$$G(\alpha \leqslant \tilde{\alpha}) = G(\alpha > \tilde{\alpha}) \tag{3.4}$$

which translates to $G(\alpha \leqslant \tilde{\alpha}) = \frac{1}{2}$. The qualitative information can be seen as a signal that moves the investor's median value of α from μ towards $\tilde{\alpha}$. I model the qualitative signal in this way to measure good information (*tone*, *sentiment*) in terms of $\tilde{\alpha}$. If f is a normal p.d.f. with mean μ and variance σ^2, the program investors solve after receiving the qualitative signal is:

$$\begin{aligned} \min_{G \in C^2} -2 \int_{-\infty}^{\infty} \Big(\frac{\exp\{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\}}{\sqrt{2\pi}\sigma} \frac{dG}{dx} \Big)^{\frac{1}{2}} dx& \\ \text{s.t.}& \\ G(\tilde{\alpha}) = \frac{1}{2}& \\ \lim_{x \to \infty} G(x) = 1& \\ \lim_{x \to -\infty} G(x) = 0& \end{aligned} \tag{3.5}$$

The objective function is the negative of the Bhattacharyya Index between a normal prior and an unknown p.d.f. $\frac{dG}{dx}$. The first constrain says that the probability mass on the left of the qualitative signal parameter $\tilde{\alpha}$ has to be equal to the probability mass to the right. The last two constraints ensure that the resulting anti derivative G is a proper c.d.f.. By modelling the program in terms of the c.d.f. and not the p.d.f. I only impose continuity in G and not on its derivative. We will see that in this formulation the posterior p.d.f. will not be continuous to accommodate one half of the probability mass before and after $\tilde{\alpha}$. The following proposition states the solution to program (3.5).

Proposition 6. *The non trivial solution to program (3.5) is given by the c.d.f.*

$$G(x) = \begin{cases} \frac{\Phi(\frac{x-\mu}{\sigma})}{2\Phi(\frac{\tilde{\alpha}-\mu}{\sigma})} & \text{if } x \geqslant \tilde{\alpha} \\ \frac{\Phi(\frac{x-\mu}{\sigma})}{2(1-\Phi(\frac{\tilde{\alpha}-\mu}{\sigma}))} & \text{if } x < \tilde{\alpha} \end{cases} \tag{3.6}$$

and the p.d.f.

$$g(x) = \begin{cases} \frac{\phi(\frac{x-\mu}{\sigma})}{2\Phi(\frac{\tilde{\alpha}-\mu}{\sigma})} & \text{if } x \geqslant \tilde{\alpha} \\ \frac{\phi(\frac{x-\mu}{\sigma})}{2(1-\Phi(\frac{\tilde{\alpha}-\mu}{\sigma}))} & \text{if } x < \tilde{\alpha} \end{cases} \tag{3.7}$$

where ϕ and Φ are the p.d.f. and c.d.f. of a standard normal variable.

Proof. See Appendix. □

Figure (3.3) shows what happens to the prior distribution after receiving a good qualitative signal, and Figure (3.4) after receiving a bad qualitative signal. We can interpret a signal being good for the fund if $\mu < \tilde{\alpha}$, bad if $\mu > \tilde{\alpha}$ and neutral if $\tilde{\alpha} = \mu$. Note that in the case in which $\tilde{\alpha} = \mu$ the effect is neutral since the prior distribution remains unchanged[6].

[6]For bayesian-updaters, any realization of the quantitative signal increases the precision of their posterior distribution. In my model, neutral information does not affect the shape of the distribution.

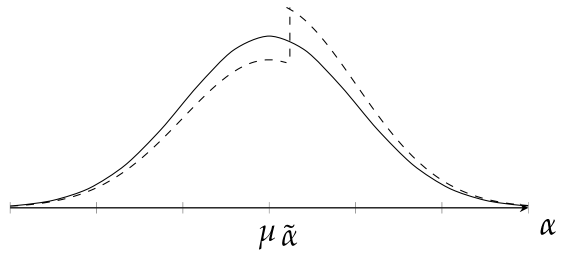


Figure 3.1: Prior (solid) and posterior (dashed) distributions after receiving a good signal about α, $\tilde{\alpha} > \mu$

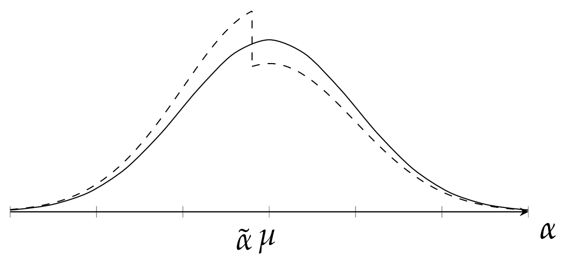


Figure 3.2: Prior (solid) and posterior (dashed) distributions after receiving a bad signal about α, $\tilde{\alpha} < \mu$

I derive the moment generating function m.g.f. and the first two moments to be used in Section (3.2.2). When working with exponential utilities having the m.g.f. helps finding a expression for the expected utility without the expectation operator (Dumitrescu and Gil-Bazo 2016,Huang et al. 2007, Huang et al. 2012). The following propositions describe the main characteristics of the investor posterior.

Proposition 7. *The Moment Generating Function m.g.f. of α under the posterior distribution (3.7) is:*

$$\Psi(t) = \frac{\exp\{\mu t + \frac{\sigma^2 t^2}{2}\}}{2} \left(\frac{\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})} + \frac{1-\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{(1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma}))} \right) \tag{3.8}$$

where Φ is the c.d.f. of a standard normal variable.

Proof. See Appendix □

Proposition 8. *The mean and variance of the posterior distribution are:*

$$\tilde{\mu} = \mu + \frac{\sigma\theta^{\phi}}{2} \tag{3.9}$$

$$\tilde{\sigma}^2 = \sigma^2\Big(1 + \frac{1}{2}\theta^{\phi}\big((\frac{\bar{\alpha}-\mu}{\sigma}) - \frac{1}{2}\theta^{\phi}\big)\Big) \tag{3.10}$$

where $\theta^{\phi} = \frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma})} - \frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}$, and where ϕ and Φ are the p.d.f. and c.d.f. of a standard normal variable.

Proof. See Appendix □

3.2.2 Fund flows under perfect competition and inelastic capital supply

In this section I extend the model of Berk and Green (2004) and explore the implications when investors learn from a qualitative signal. I start by discussing the main assumptions of the model and the main variations I propose to study the effects of qualitative information on capital allocation.

Consider an economy that lasts for three periods $t = 0, 1, 2$, all investors are symmetrically informed, and funds differ on their manager's ability to generate abnormal returns. In order for managers to obtain positive abnormal returns they must seek undervalued securities and trade, moving the price against them, decreasing the return available to pay to investors. To model this situation I assume funds have a monetary cost of managing an amount q_t of capital expressed by $C(q_t)$ such that $C(0) = 0, C'(q_t) \geqslant 0$, and $C''(q_t) > 0$. For simplicity I assume the parametrization $C(q) = aq^2$ as in Berk and Green (2004). Investors allocate capital to funds inelastically until the point in which the expected abnormal return is zero.

Let $R_t = \alpha + \epsilon_t$ be the return over the benchmark that a fund generates at time $t = 1, 2$ (for simplicity I suppress the fund index for most of the analysis that follows). ϵ_t is normally distributed with mean zero and variance $1/\omega$ where ω captures the precision of the quantitative signal. Investors begin with a prior on α which is normally distributed with mean α_0 and variance $1/\gamma$, where γ is the precision of the prior distribution. I consider the simplest case of the model in which managers receive a fixed fee f for each dollar under management. The total payout received by investors at time 1 is:

$$TP_{t+1} = q_t R_{t+1} - a q_t^2 - q_t f \tag{3.11}$$

Defining r_1 as the abnormal return that investors receive after costs and fees

$$r_{t+1} = R_{t+1} - a q_t - f \tag{3.12}$$

Where $a q_t + f = c(q_t) = \frac{C(q_t)}{q_t} + f$ is the unit cost. At time 0 investors have a prior on the true managerial ability of fund managers, and invest an amount q_0 on each fund. At time $t = 1$ investors observe a return r_1 and a qualitative signal of the form $Pr(\alpha \geqslant \tilde{\alpha}) = Pr(\alpha < \tilde{\alpha})$ and derive their posterior distribution first incorporating r_1 following Bayes rule, and then $\tilde{\alpha}$ following the *pseudo-bayesian* updating rule [7]. Denote as $\alpha_t = \mathbb{E}_t(R_t)$ the perceived managerial ability of a fund conditional on all information available up to time t. In every period, investors allocate capital on those funds with a positive expected abnormal return and withdraw capital from those with a negative expected abnormal return. In equilibrium the capital allocated to every fund makes its expected abnormal return equal to zero:

$$\mathbb{E}_0(r_1) = \mathbb{E}_1(r_2) = 0 \tag{3.13}$$

Which implies

$$\alpha_t = a q_t + f \tag{3.14}$$

As the perceived managerial ability α_t changes the amount of capital under management q_t changes to ensure equation (3.14) is satisfied at all points in time.

[7] I will talk undistinguishably between investors receiving the signal $Pr(\alpha \geqslant \tilde{\alpha}) = Pr(\alpha < \tilde{\alpha})$ and receiving the signal $\tilde{\alpha}$.

Proposition 9. *For any fund the evolution of α_1 and the amount of capital under management q_1 satisfies:*

$$\alpha_1 = (\alpha_0 + \frac{\omega}{\gamma+\omega} r_1) + \frac{1}{2\sqrt{\gamma+\omega}} \Big(\frac{\phi\big(J(\tilde{\alpha}, r_1, \alpha_0)\big)}{1 - \Phi\big(J(\tilde{\alpha}, r_1, \alpha_0)\big)} - \frac{\phi\big(J(\tilde{\alpha}, r_1, \alpha_0)\big)}{\Phi\big(J(\tilde{\alpha}, r_1, \alpha_0)\big)} \Big) \tag{3.15}$$

$$q_1 = \frac{\alpha_1 - f}{a} \tag{3.16}$$

Where $J(\tilde{\alpha}, r_1, \alpha_0) = \sqrt{\gamma+\omega}(\tilde{\alpha} - \alpha_0 - \frac{\omega}{\gamma+\omega} r_1)$.

Using proposition (9) I express the flow of new funds $F = \frac{q_1 - q_0(1+r_1)}{q_0}$ to a fund at time 1 as

$$F = \big(\frac{\omega}{a q_0 (\gamma+\omega)} - 1 \big) r_1 + \frac{\phi\big(J(\tilde{\alpha}, r_1, \alpha_0)\big)}{2 a q_0 \sqrt{\gamma+\omega}} \Big(\frac{1}{1 - \Phi\big(J(\tilde{\alpha}, r_1, \alpha_0)\big)} - \frac{1}{\Phi\big(J(\tilde{\alpha}, r_1, \alpha_0)\big)} \Big) \tag{3.17}$$

the following proposition tells the relation between the tone of the qualitative information and net flows

Proposition 10. *Fund flows are increasing in the tone of the qualitative signal*

$$\frac{\partial F}{\partial \tilde{\alpha}} > 0$$

Proof. See Appendix □

The tone of a signal $\tilde{\alpha}$ depends on the realization of r_1 which means their analysis has to be done jointly. r_1 modifies the median of the posterior distribution towards $\alpha_0 + \frac{\omega}{\gamma+\omega} r_1$ making $\tilde{\alpha}$ good news only if $\tilde{\alpha} > \alpha_0 + \frac{\omega}{\gamma+\omega} r_1$. In this way of modelling qualitative information, the *tone* of the signal is not absolute but depends on past quantitative performance. Figure (3.3) plots the performance-flow relation for a neutral, good and bad realization of the qualitative signal. Flows are increasing in past quantitative performance - apart from the region in which the constraint $q_1 \geqslant 0$ binds - and for a given realization of r_1 a good realization of the qualitative signal increases the level of flows.

3.2.3 Optimal Portfolio Choice in the presence of qualitative information

Consider a three period economy $t = 0, 1, 2$ in which investors allocate a budget between a risk free asset and a mutual fund. I normalize the return on the risk free asset to be zero and at time $t = 1, 2$ and the mutual fund produces a risky return r_t[8] following the process:

$$r_t = \alpha + \epsilon_t \tag{3.18}$$

The term α represents the managerial ability of the mutual fund manager to generate positive abnormal returns over a benchmark. This ability is constant over time, and is unknown both for investors and managers. The term $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ is independently and identically distributed (i.i.d.) over time. At time $t = 0$ investors have a prior distribution on managerial ability that is normally distributed:

$$\alpha \sim N(\alpha_0, \sigma_0^2) \tag{3.19}$$

At time 1 investors observe the realized return r_1 of the mutual fund and derive their posterior distribution following Bayes rule.

$$\alpha \sim N(\alpha_1, \sigma_1^2) \tag{3.20}$$

where

$$\alpha_1 = \frac{\sigma_\epsilon^2}{\sigma_0^2 + \sigma_\epsilon^2}\alpha_0 + \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2} r_1, \ \sigma_1^2 = \frac{\sigma_0^2 \sigma_\epsilon^2}{\sigma_0^2 + \sigma_\epsilon^2} \tag{3.21}$$

After observing the return r_1 investors receive a qualitative statement as defined in Section (1) of the form $Pr(\alpha > \tilde{\alpha}) = Pr(\alpha \leqslant \tilde{\alpha})$ and update their posterior distribution following the pseudo-bayesian updating rule described in Section (1). After updating their beliefs investors solve the following program

$$\begin{aligned}
&\max_{X_1 \geqslant 0} \mathbb{E}\big[- e^{-\gamma W_2} | r_1, \tilde{\alpha} \big] \\
&\text{s.t.} \\
&W_2 = W_1 + X_1 r_2
\end{aligned} \tag{3.22}$$

The following proposition describes the optimal holding at $t = 1$

Proposition 11. *The optimal holding at time* 1 X_1^* *after receiving the quantitative signal* r_1 *and the qualitative signal* $\tilde{\alpha}$ *is given by:*

$$X_1^* = \max\{X_1^u, 0\}$$

[8] In Huang et al. (2007) there are no dis-economies of scale and the return r_t is given completely to investors, that is why I choose to define the process using r and not R.

where X_1^u is the solution to the following system of equations which yields the optimal holding when short sales are allowed,

$$\phi(y)\Phi^{\alpha} + (y - \frac{\tilde{\alpha}}{\sigma_{\alpha}})(\Phi_2^{\alpha} + \Phi^{\alpha}\Phi(y)) = 0$$
$$y = z_{\alpha} + X_1^u \sigma_{\alpha}\gamma$$

and where $z_{\alpha} = \frac{\tilde{\alpha}-\alpha_1}{\sigma_{\alpha}}$, $\Phi^{\alpha} = \frac{1}{\Phi(z_{\alpha})} - \frac{1}{1-\Phi(z_{\alpha})}$, $\Phi_1^{\alpha} = \frac{1}{\Phi(z_{\alpha})}$, $\Phi_2^{\alpha} = \frac{1}{1-\Phi(z_{\alpha})}$ *and* $\sigma_{\alpha}^2 = (\sigma_1^2 + \sigma_{\epsilon}^2)$.

Proof. See Appendix □

Since I do not have a closed-form solution I can only infer the relation between net-flows and the signals received by investors graphically. Figure (3.4) plots the performance-flow relation for a neutral, good and bad realization of the qualitative signal. Flows are increasing in past quantitative performance - apart from the region in which the constraint $X_1 \geqslant 0$ binds - and for a given realization of r_1 a good realization of the qualitative signal increases the level of flows. The results are similar to the ones obtained in the more simpler framework of section 3.2.2, however the results in this section will let us understand the strategic transmission of information described in the next section.

3.3 Strategic transmission of qualitative information

It is possible to manipulate qualitative information to make it look more appealing for investors. Empirical research in textual analysis has shown that managers tend to disclose negative information by making it more complex to read -an issue concerning the *readability* of financial information- in an attempt to make the bad information look better (See e.g. Li (2008), Loughran and McDonald (2014)). In this section I analyze the strategic disclosure of qualitative information by the fund managers. Consider risk neutral managers that are willing to maximize fund flows into the fund. Define as $F(r_1, \tilde{\alpha})$ the flows to the fund after investors observe the return r_1 and the qualitative signal $\tilde{\alpha}$ at time 1. From subsection (3.2.2) and (3.2.3) we know that $F_{r_1} > 0$ and $F_{\tilde{\alpha}} > 0$. On the other hand consider an investor maximizing expected utility as in subsection (3.2.2), we can rewrite the maximum expected utility of program 3.22 as:

$$EU(W_2(X^*(r_1,\tilde{\alpha})), r_1, \tilde{\alpha}) = \mathbb{E}\Big[-\exp\{-\gamma W_2(X^*(r_1,\tilde{\alpha}))\}\Big| r_1, \tilde{\alpha}\Big] \tag{3.23}$$

Where $X^*(r_1, \tilde{\alpha})$ is the solution to program 3.22, and $W_2(X^*(r_1, \tilde{\alpha}))$ is the terminal wealth after investing $X^*(r_1, \tilde{\alpha})$ in the fund. Since $X^*(r_1, \tilde{\alpha})$ maximizes program 3.22 for any $\alpha \neq \tilde{\alpha}$

$$EU(W_2(X^*(r_1, \tilde{\alpha})), r_1, \tilde{\alpha}) > EU(W_2(X^*(r_1, \alpha)), r_1, \tilde{\alpha}) \tag{3.24}$$

In particular imagine the situation in which a fund, having the control on transmitting the information to investors, transmits a signal $\tilde{\alpha} + \Delta$ when the *true* signal is $\tilde{\alpha}$. Investors will maximize program (3.22) assuming the qualitative signal is the one issued by the fund managers. Figure (3.5) shows the loss in expected utility that investors face if take the signal that the managers issue at face value. In this case investors maximize a *fake* utility function, leading to a sub-optimal capital allocation. In the next section I study the case in which investors can decide to verify the signal that funds transmit at a cost, and a reputation cost for fund managers in case they are caught manipulating the signal.

3.3.1 The game

The quantitative signal r_1 is public knowledge among the investors and the fund managers while the qualitative signal $\tilde{\alpha}$ is first observed by the fund manager and then transmitted to investors. My game is similar to the one in Dumitrescu and Gil-Bazo (2016) but instead of the manager deciding whether or not to censor the qualitative signal, I assume that the fund can modify the signal to make it look more appealing to investors - this captures the idea of making bad information more complex for investors to understand. The fund manager transmits a signal $\tilde{\alpha} + \Delta$ where $\Delta > 0$ since $F_{\tilde{\alpha}}$. Investors can either take this signal at face-value and risk making an investment decision based on incorrect information, or verify the signal from some other source at a cost C [9]. I assume that the cost of verifying is low enough such that if the investor is sure that the manager is lying about the qualitative signal she will decide always to verify the information. If the manager is in fact lying and the investor verifies the information, the fund manager will suffer a reputation cost K that is assumed to be fixed. When the investor decides to trust the manager and not verify the information, if the manager manipulates the signal, the expected utility of the investor will be below optimal since the investor will behave as if $\tilde{\alpha} + \Delta$ was the true qualitative signal therefore investing too much in the fund

[9]Investors can verify the signal by looking at the financial press, analyst reports, ratings, etc

$(X(r_1, \tilde{\alpha}+\Delta) > X(r_1,\tilde{\alpha}))$. The following table presents the game and the best responses of the players:

		Investor: Believe	Investor: Suspect
Fund manager	$\tilde{\alpha}+\Delta$	$F(r_1,\tilde{\alpha}+\Delta)^*$ $EU(W_2(X(r_1,\tilde{\alpha}+\Delta)), r_1,\tilde{\alpha})$	$F(r_1,\tilde{\alpha})-K$ $EU(W_2(X(r_1,\tilde{\alpha}))-C, r_1,\tilde{\alpha})^*$
	$\tilde{\alpha}$	$F(r_1,\tilde{\alpha})$ $EU(W_2(X(r_1,\tilde{\alpha})), r_1,\tilde{\alpha})^*$	$F(r_1,\tilde{\alpha})^*$ $EU(W_2(X(r_1,\tilde{\alpha}))-C, r_1,\tilde{\alpha})$

Assuming that the cost of verifying C is small $C < \bar{C}$ where $\bar{C}$ is the solution to $EU(W_2(X(r_1,\tilde{\alpha}+\Delta)), r_1,\tilde{\alpha}) = EU(W_2(X(r_1,\tilde{\alpha}))-\bar{C}, r_1,\tilde{\alpha})$. Since there are no Nash equilibria in pure strategies I proceed to find it in mixed strategies.

Proposition 12. *Nash Equilibrium: The unique Nash equilibrium of the above game is given by the mixed strategies:*

$$\begin{cases} \tilde{\alpha}+\Delta & \textit{with probability } p \\ \tilde{\alpha} & \textit{with probability 1-p} \end{cases}$$

where

$$p = \frac{EU(W_2(X(r_1,\tilde{\alpha})), r_1,\tilde{\alpha}) - EU(W_2(X(r_1,\tilde{\alpha}))-C, r_1,\tilde{\alpha})}{EU(W_2(X(r_1,\tilde{\alpha})), r_1,\tilde{\alpha}) - EU(W_2(X(r_1,\tilde{\alpha}+\Delta)), r_1,\tilde{\alpha})}$$

for the fund manager, and

$$\begin{cases} Believe & \textit{with probability } q \\ Suspect & \textit{with probability } 1-q \end{cases}$$

for the investor, where

$$q = \frac{K}{F(r_1,\tilde{\alpha}+\Delta) - F(r,\tilde{\alpha}) + K}$$

Proof. See Appendix □

The mixed strategy gives us an idea of the distribution of funds that will lie in equilibrium, to understand how this proportion changes with respect of the parameters of the model we can derive the following comparative statistics.

Proposition 13. *The first order comparative statistics of the Nash equilibrium are:*

- *The fraction of investors that believe on what fund managers transmit increases with the reputation costs that fund managers have if caught manipulating information*
$$\frac{\partial q}{\partial K} > 0$$
- *The fraction of funds that manipulate information increases with the verification cost by investors:*
$$\frac{\partial p}{\partial C} > 0$$
- *If flows are strictly convex on the qualitative signal* $\frac{\partial^2 F}{\partial \tilde{\alpha}^2} > 0$ *the higher the qualitative signal the more investors will suspect of the signal and verify.*
$$\frac{\partial q}{\partial \tilde{\alpha}} < 0$$
- *The more fund managers can manipulate the signal, the more investors will suspect and verify it.*
$$\frac{\partial q}{\partial \Delta} < 0$$

Figure (3.6) shows how the probabilities of a fund manipulating information p and of an investor verifying the information q change with the performance of fund r_1. First we observe that there exists a threshold value $\bar{r}$ such that if the return of the fund is larger, the fraction of investors that verify the signal increases dramatically. This is due to the fact that after this threshold the benefit of investors verifying the information is greater than the cost of verifying. My model predicts that if past performance is good enough (above threshold $\bar{r}$) every investor will verify the signal since managers will have an extra incentive to manipulate the information. We can observe from the plots of (p) for different levels of the qualitative signal, that when past performance is good enough, good information is more likely to be manipulated than bad information. My model suggests that the behaviour described in Li (2008) and Loughran and McDonald (2014) in which firms manipulate textual information, should be stronger among those funds that perform better instead of the worst performing ones.

3.4 Conclusion

I develop a theoretical model of mutual fund investors that explicitly models qualitative information in the optimal portfolio decision of investors.

Investors react positively to good qualitative information since it is incorporated into their posterior distribution. Since qualitative information can be easily manipulated, I study the strategic behaviour of fund managers and investors when issuing and receiving a qualitative signal. "I find that in equilibrium, the fraction of funds that manipulate information increases with the cost of verifying the information. The fraction of investors that verify the information decreases with the reputation costs faced by funds if caught manipulating information.

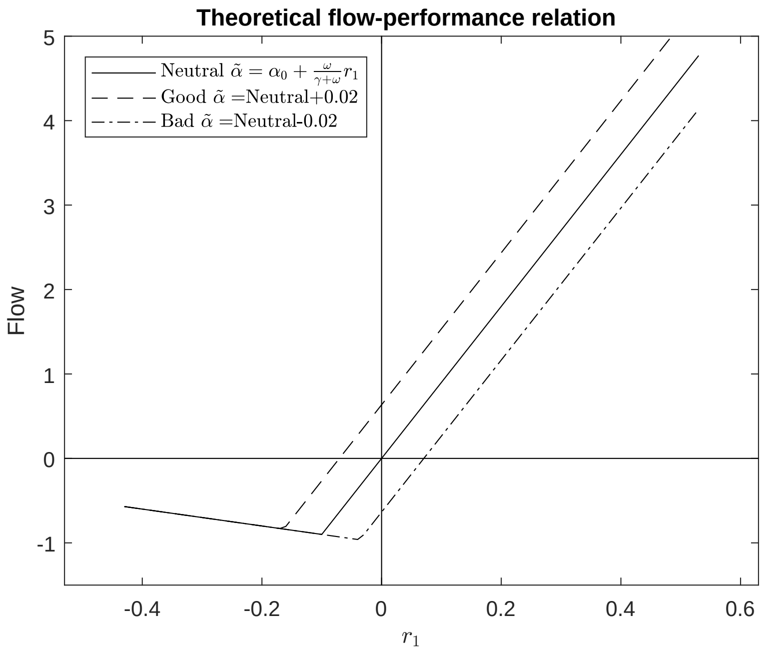


Figure 3.3: **Theoretical flow performance relation when qualitative information is neutral, good, or bad with respect to the posterior after the quantitative signal r_1.** The solid line corresponds to the case in which investors receive neutral qualitative information $\tilde{\alpha} = \alpha_0 + \frac{\omega}{\gamma+\omega} r_1$, the dashed line corresponds to the case in which investors receive a good qualitative signal with respect to the posterior expected ability after receiving the quantitative signal r_1, $\tilde{\alpha} = \alpha_0 + \frac{\omega}{\gamma+\omega} r_1 + 0.02$ and finally the dot-dashed line corresponds to bad qualitative with respect to the posterior expected ability after receiving the quantitative signal r_1 $\tilde{\alpha} = \alpha_0 + \frac{\omega}{\gamma+\omega} r_1 - 0.02$. The parameters are $a = 1$, $\omega = 39.06$, $\gamma = 156.25$, $\alpha_0 = 0.03$ and $f = 0.01$.The fund return is $r_1 = \alpha + \epsilon_1$ where $\alpha \sim N(\alpha_0, \frac{1}{\gamma})$ is the prior about managerial ability, and $\epsilon_1 \sim N(0, \frac{1}{\omega})$ is the i.i.d. noise over time and across funds.

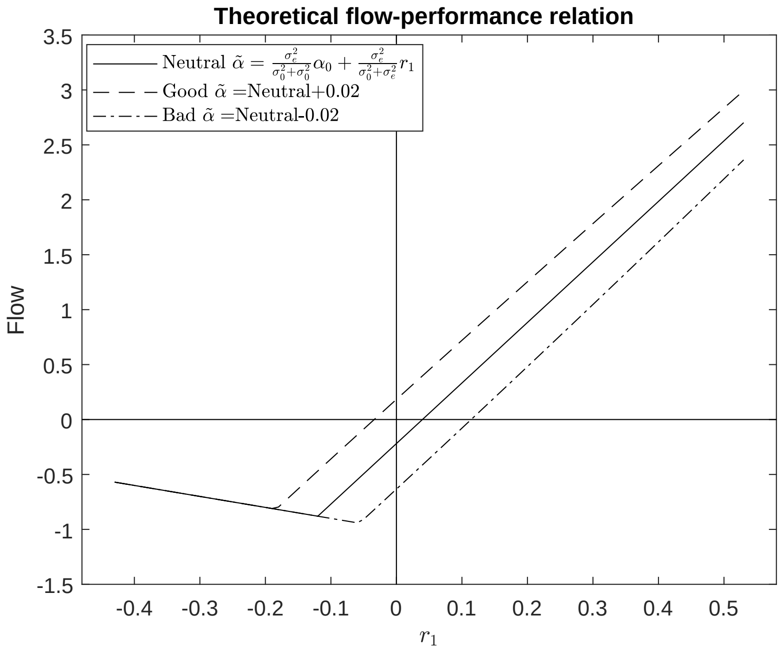


Figure 3.4: **Theoretical flow performance relation when qualitative information is neutral, good, or bad with respect to the posterior after the quantitative signal r_1.** The solid line corresponds to the case in which investors receive neutral qualitative information $\tilde{\alpha} = \frac{\sigma_\epsilon^2}{\sigma_0^2+\sigma_\epsilon^2}\alpha_0 + \frac{\sigma_0^2}{\sigma_0^2+\sigma_\epsilon^2}r_1$, the dashed line corresponds to the case in which investors receive a good qualitative signal with respect to the posterior expected ability after receiving the quantitative signal r_1, $\tilde{\alpha} = \frac{\sigma_\epsilon^2}{\sigma_0^2+\sigma_\epsilon^2}\alpha_0 + \frac{\sigma_0^2}{\sigma_0^2+\sigma_\epsilon^2}r_1 + 0.02$ and finally the dot-dashed line corresponds to bad qualitative with respect to the posterior expected ability after receiving the quantitative signal r_1 $\tilde{\alpha} = \frac{\sigma_\epsilon^2}{\sigma_0^2+\sigma_\epsilon^2}\alpha_0 + \frac{\sigma_0^2}{\sigma_0^2+\sigma_\epsilon^2}r_1 - 0.02$. The parameters are $\sigma_0 = 0.08$, $\sigma_e = 0.16$, $\alpha_0 = 0.03$ and $\gamma = 1$.The fund return is $r_1 = \alpha + \epsilon_1$ where $\alpha \sim N(\alpha_0, \sigma_0^2)$ is the prior about managerial ability, and $\epsilon_1 \sim N(0, \sigma_e^2)$ is the i.i.d. noise over time and across funds.

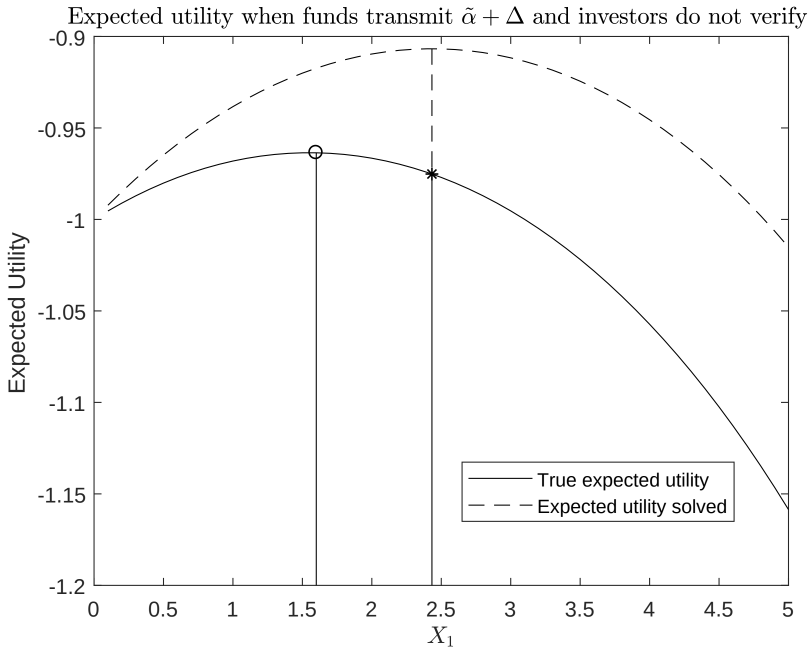


Figure 3.5: **Expected utility loss when a fund issues a qualitative signal with parameter $\tilde{\alpha} + \Delta$ and the investor takes this signal at face value.** The solid line corresponds to the true expected utility of the investor $EU(X, r_1, \tilde{\alpha})$, the dashed line corresponds to the expected utility being maximized by the investor $EU(X, r_1, \tilde{\alpha} + \Delta)$. The investor allocates too much of his wealth to the fund and losses the difference $EU(X(r_1, \tilde{\alpha}), r_1, \tilde{\alpha}) - EU(X(r_1, \tilde{\alpha} + \Delta), r_1, \tilde{\alpha})$. The parameters are: $\sigma_0 = 0.08$, $\sigma_\epsilon = 0.16$, $\alpha_0 = 0.03$, $\gamma = 1$, $r_1 = 0.1$, $W_1 = 0$, $\tilde{\alpha} = 0.05$ and $\Delta = 0.05$.

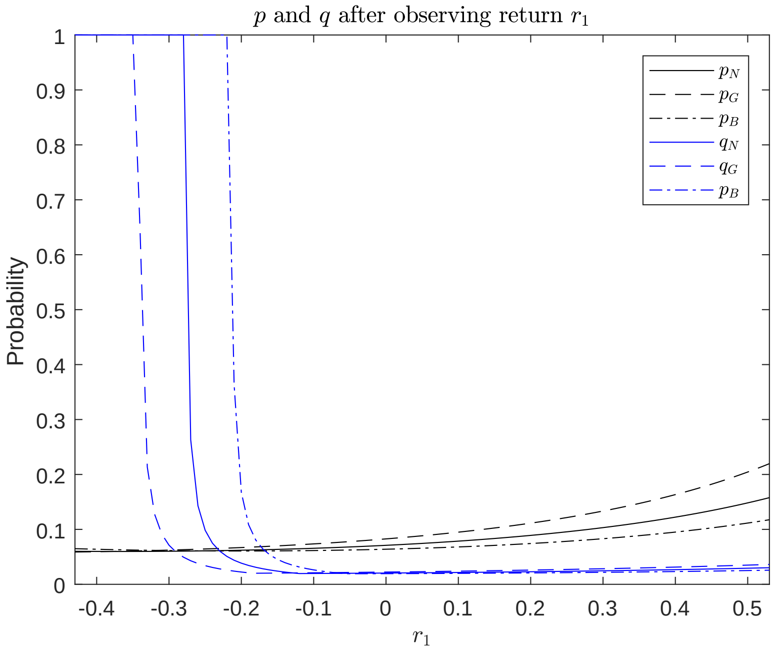


Figure 3.6: **Probabilities of the mixed Nash equilibrium after investors observe return r_1 and a signal $\tilde{\alpha}$.** The graph shows how the probability of funds manipulating information (p) and the probability of investors believing the signal issued by funds (q). The parameters are $\sigma_0 = 0.08$, $\sigma_\epsilon = 0.16$, $\alpha_0 = 0.03$, $\gamma = 1$, $W_0 = 0$, $\delta = 0.05$. The good qualitative signal is given by $\alpha_1 + 0.02$, the neutral qualitative signal is α_1, and the bad qualitative signal is $\alpha_1 - 0.02$. Probabilities p_i where $i \in \{G, N, B\}$ are the probabilities of a fund manipulating the qualitative information if it is good, neutral or bad respectively. Probabilities q_i where $i \in \{G, N, B\}$ are the probabilities that a investor process the signal at face value if the qualitative information is good, neutral or bad respectively.

Appendix A

Appendix - Stroke of a Pen: Investment and Stock Returns under Energy Policy Uncertainty

A.1 Energy Price Uncertainty and Investment - Alternative Formulation

In this section I develop a small model of investment based on Stewart (1978). Consider a representative firm that uses two inputs, capital (K) and energy (E) to produce a final product. The firm purchases energy in a competitive market at a price w_e per unit and combines energy with capital (e.g. property plant equipment PPE) to produce its final product, and all other factors required such as labor are maximized out of the equation. The firm's profit at any period is given by

$$\pi = pq - w_e E - rK \tag{A.1}$$

where p is the output price, q is the quantity produced, and r is the unit cost of capital including opportunity costs, or financing costs. I assume that input substitution between energy and capital is possible. More precisely, if the firm uses technology $q = f(K, E)$ to convert capital and energy into the final product, substitution implies that for any fixed output q and capital K, the quantity of energy required

$$E = g(q, K) \tag{A.2}$$

satisfies $\partial g / \partial K < 0$. I assume that the firm's PPE configuration is not instantaneously adjustable, so that capital has to be determined in advance, and that the price of energy w_e follows a mean preserving spread process

$$w_e = \beta\nu + \theta \tag{A.3}$$

where β and θ are constant shift parameters, and ν is a positive random variable. Finally, I assume the firm's manager is risk averse and maximizes expected utility over profit π with a standard von Neumann and Morgenstern (1947) utility function U. If we express the optimization problem in terms of capital and output

$$pq - w_e g(q, K) - rK \tag{A.4}$$

a necessary condition for utility maximization is therefore

$$\mathbb{E}[U'(\pi)(-w_e \frac{\partial g}{\partial K} - r)] = 0 \tag{A.5}$$

To study how uncertainty on energy prices impacts investment, I consider first the benchmark case in which the manager is risk neutral so $U'(\pi)$ is a constant. In this situation the first order condition of profit maximization implies the manager chooses next period capital satisfying

$$-\frac{\partial g}{\partial K} = \frac{r}{\mathbb{E}[w_e]} \tag{A.6}$$

equating the marginal rate of technical substitution to the expected factor price ratio. On the other hand, a risk averse manager with concave utility function will depart from this first order condition. Expressing (A.5) in covariance form yields

$$\mathbb{E}[U'(\pi)(-w_e \frac{\partial g}{\partial K} - r)] = \mathbb{E}[U'(\pi)]\mathbb{E}[(-w_e \frac{\partial g}{\partial K} - r)] + cov([U'(\pi), (-w_e \frac{\partial g}{\partial K} - r)) = 0 \tag{A.7}$$

U is a concave function so $U'(\pi)$ is increasing in w_e, and since the term $-w_e \frac{\partial g}{\partial K} - r$ is also increasing in w_e we have that the covariance term in (A.7) is strictly positive, which implies

$$\mathbb{E}[(-w_e \frac{\partial g}{\partial K} - r)] < 0 \rightarrow -\frac{\partial g}{\partial K} < \frac{r}{\mathbb{E}[w_e]} \tag{A.8}$$

so a risk averse manager demands more capital than its risk neutral counterparty. This results shows that in the presence of uncertainty regarding future energy prices w_e a risk averse manager increases investment.

A.2 Robustness Analysis to the Information Set

A.2.1 Theoretical Setup

In this section I explain how I construct a measure of energy policy uncertainty using a large battery of macroeconomic, financial, and political data. To keep the notation self-contained I explain an extension of the forecasting model based on Jurado et al. (2015), including extra political variables in the forecasting exercise. For technical details I refer the reader to the original paper. Policy uncertainty arises from the impossibility of economic agents to perfectly forecast politician's decisions. I follow the literature on uncertainty and define energy policy uncertainty as the conditional volatility of the unforecastable component of the number of energy related executive orders from the point of view of an economic agent. In particular, given investor's information set I_t, I define the k-period ahead energy political uncertainty on topic i as

$$\mathcal{U}_t(k) = \sqrt{\mathbb{E}[(\mathrm{eo}_{t+k}^{energy} - \mathbb{E}[\mathrm{eo}_{t+k}^{energy}|I_t])^2|I_t]} \tag{A.9}$$

where eo_t^{energy} is the total number of executive orders signed by the U.S. President during month t. To model the information set I_t, I use a large battery of macroeconomic and financial time series, as well as political information.[1] The information set comprises variables that help forecast the number of executive orders within a one month, one quarter, and one year horizons. I include the current political agenda, the number of public laws and executive orders being passed on each topic. I use information about the party in power in each one of the two chambers as well as the president's affiliation. This allows me to control for differences in the agenda of both parties, which are reflected in next periods' political decisions. Finally I include the original 132 time-series of macroeconomic variables and the 147 financial variables presented in Ludvigson and Ng (2009) and Ludvigson and Ng (2007), and used in Jurado et al. (2015) for the uncertainty estimations.

The forecasting procedure is explained as follows. Let $\mathcal{P}_\mathbf{t} = (\mathrm{pl}_{1t}, ..., \mathrm{pl}_{Nt})$ be a vector containing the number of public laws for every topic, and $\mathcal{E}_\mathbf{t} = (\mathrm{eo}_{1t}, ..., \mathrm{eo}_{Nt})$ a vector containing the number of executive orders for every topic i. Also, let $\mathbf{X_t^{JLN}} = (X_{1t}^{JLN}, ..., X_{Nt}^{JLN})'$ denote the original

[1] As noted by Jurado et al. (2015), many proxies of economic uncertainty fail to account for the forecastable component of the time series being analysed, which pervades the uncertainty estimations with predictive variation.

macroeconomic and financial predictors used by Jurado et al. (2015) after suitable transformations to ensure the series are stationary. Let $\mathbf{X_t} = (\mathbf{X_t^{LN}}, \mathcal{P}_t, \mathcal{E}_t, \text{House}_t, \text{Senate}_t, \text{President}_t)$ be the whole set of predictors in the forecastable model consisting of the original 132+147 predictors in Jurado et al. (2015), 30 time series regarding the number of public laws, executive orders for every of the 20 topics defined in the Comparative Agenda project, plus three dummy variables House, Senate, and President which take the value of one if at month t the House of Representatives has a Republican Majority, the Senate has a Republican Majority, or the President is Republican. This last three variables are then transformed in first differences to ensure stationarity. It is assumed predictor $X_{it} \in \mathbf{X_t}$ has an approximate factor structure

$$X_{it} = \Lambda_i^{F'} \mathbf{F_t} + e_{it}^X \tag{A.10}$$

where $\mathbf{F_t}$ is a vector of r_F common factors, and e_{it}^X is an idiosyncratic error. To compute uncertainty on the number of executive order related to energy $\text{eo}_t^{energy} \in \mathcal{E}_t$

$$\text{eo}_{t+1}^{energy} = \phi_j^e(L)\text{eo}_t^{energy} + \gamma_j^F(L)\hat{\mathbf{F}}_\mathbf{t} + \gamma_j^W(L)\mathbf{W_t} + v_{t+1}^e \tag{A.11}$$

where $\mathbf{W_t}$ is a set of extra predictors including square terms of the principal components, and $\phi^e(L), \phi^F(L), \phi^W(L)$ are polynomials on the lag operator L of order n_e, n_f, n_w respectively. Let $\mathbf{Z_t} = (\hat{\mathbf{F}}_\mathbf{t}, \mathbf{W}_t)'$ and define $\mathcal{Z}_t = (\mathbf{Z'_t}, ..., \mathbf{Z'_{t-q+1}})$, as well as $E_t = (\text{eo}_t^{energy}, ..., \text{eo}_{t-q+1}^{energy})'$, the forecasting model can then be expressed as:

$$\begin{bmatrix} \mathcal{Z}_t \\ E_t \end{bmatrix} = \begin{bmatrix} \Phi^{\mathcal{Z}} & 0 \\ \Lambda' & \Phi^E \end{bmatrix} \begin{bmatrix} \mathcal{Z}_{t-1} \\ E_{t-1} \end{bmatrix} + \begin{bmatrix} \mathcal{V}_t^{\mathcal{Z}} \\ \mathcal{V}_t^E \end{bmatrix} \tag{A.12}$$

or in compact notation

$$\mathcal{Y}_t = \mathbf{\Phi}^{\mathcal{Y}} \mathcal{Y}_{t-1} + \mathcal{V}_t^{\mathcal{Y}} \tag{A.13}$$

and by the assumption of stationary and under quadratic loss the optimal k-period ahead forecast is the conditional mean

$$\mathbb{E}_t[\mathcal{Y}_{t+k}] = (\mathbf{\Phi}^{\mathcal{Y}})^k \mathcal{Y}_t \tag{A.14}$$

the forecast error variance-covariance matrix is

$$\Omega_t^{\mathcal{Y}}(k) = \mathbb{E}_t[(\mathcal{Y}_{t+k} - \mathbb{E}_t[\mathcal{Y}_{t+k}])(\mathcal{Y}_{t+k} - \mathbb{E}_t[\mathcal{Y}_{t+k}])'] \tag{A.15}$$

the $k > 1$ ahead forecast error variance matrix evolves accordingly to

$$\Omega_t^{\mathcal{Y}}(k) = \mathbf{\Phi}^{\mathcal{Y}} \Omega_t^{\mathcal{Y}}(k-1) \mathbf{\Phi}_{\mathbf{j}}^{\mathcal{Y}\prime} + \mathbb{E}[\mathcal{V}_{t+k}^{\mathcal{Y}} \mathcal{V}_{t+k}^{\mathcal{Y}\prime}] \tag{A.16}$$

where $\Omega_t^{\mathcal{Y}}(1) = \mathbb{E}[\mathcal{V}_{t+1}^{\mathcal{Y}} \mathcal{V}_{t+1}^{\mathcal{Y}\prime}]$. The political uncertainty estimation can then be estimated as:

$$\mathcal{U}_t(k) = \sqrt{1'_{energy} \Omega_t^{\mathcal{Y}}(k) 1_{energy}} \tag{A.17}$$

where 1_{energy} is an adequate selector operator. Finally the components in $\mathbb{E}[\mathcal{V}_{j,t+h}^{\mathcal{Y}} \mathcal{V}_{j,t+h}^{\mathcal{Y}\prime}]$ can be estimated imposing a stochastic volatility structure on the residuals in $\mathbf{Z_t}$ and E_t assuming an autoregressive behaviour of the elements of $\mathbf{Z_t}$:

$$Z_t = \Phi^Z Z_{t-1} + v_t^Z \tag{A.18}$$

where $Z_t \in \mathbf{Z_t}$, the residual term admits $v_t^Z = \sigma_t^Z \epsilon_t^Z$ and $\epsilon_t^Z \sim N(0,1)$ and the forecast residual $v_{t+1}^y = \sigma_{t+1} \epsilon_t^y$, the stochastic volatility model used assumes an AR(1) process on the square of the log volatility

$$\log(\sigma_t^Z)^2 = \alpha^Z + \beta^Z \log(\sigma_{t-1}^Z)^2 + \tau^Z \eta_t^Z \tag{A.19}$$

$$\log(\sigma_{t+1}^y)^2 = \alpha + \beta \log(\sigma_t^y)^2 + \tau^e \eta_{t+1} \tag{A.20}$$

where parameters $(\alpha^Z, \beta^Z, \tau^Z, \alpha, \beta, \tau^e)$ are estimated via MCMC. The stochastic volatility model allows us to express the volatility as:

$$\mathbb{E}_t[\sigma_{t+k}^Z] = \exp\left[\alpha^Z \sum_{s=0}^{k-1} (\beta^Z)^s + \frac{(\tau^Z)^2}{2} \sum_{s=0}^{k-1} (\beta^Z)^{2s} + (\beta^Z)^k \log(\sigma_t^Z)^2\right] \tag{A.21}$$

and

$$\mathbb{E}_t[\sigma_{j,t+k}^y] = \exp\left[\alpha_j^y \sum_{s=0}^{k-1} (\beta_j)^s + \frac{(\tau_j)^2}{2} \sum_{s=0}^{k-1} (\beta_j)^{2s} + (\beta_j)^k \log(\sigma_{jt}^y)^2\right] \tag{A.22}$$

elements in $\mathbb{E}[\mathcal{V}_{j,t+k}^{\mathcal{Y}} \mathcal{V}_{j,t+k}^{\mathcal{Y}\prime}]$ are estimated using the fact that $\mathbb{E}_t[\sigma_{j,t+k}^y]^2 = \mathbb{E}_t[v_{j,t+k}^y]^2$ and $\mathbb{E}_t[\sigma_{t+k}^Z]^2 = \mathbb{E}_t[v_{t+k}^Z]^2$.

A.2.2 Estimation

I follow Jurado et al. (2015) and set $r_F = 10$, and $\mathbf{Z}_t = [\mathbf{F}_t, \mathbf{F}_t^2, G_t]$ where G_t is the first principal component of $\mathbf{X}_t^2$. The polynomials on the lag operator used in the forecasting regressions are assumed to have degrees $n_y = 4, n_f = 2, n_w = 2$, and lags $q = \lfloor 4(\frac{T}{100})^{2/9} \rfloor$. On a first stage the elements of $\mathbf{Z}_t$ are pruned to keep only those ones that provide individual significance with t-statistics greater than 2.575 in the multivariate forecasting regression of y_{t+1} on the candidate predictors known at time t. Once residuals are estimated, the parameters in the stochastic volatility model are estimated via a Markov Chain Montecarlo Method.[2]

[2]I thank Serena Ng for making available the code used in Jurado et al. (2015) on her webpage.

Table A.1: Return predictability regressions - Complete Information Set

	(1) $R_{t\to t+1}$	(2) $R_{t\to t+3}$	(3) $R_{t\to t+12}$	(4) $R_{t\to t+36}$
$\mathcal{U}_{t\to t+1}$	-0.197	-0.484**	-2.246*	-2.216**
	(0.169)	(0.212)	(1.247)	(0.894)
$(d-p)_t$	2.350**	5.490***	28.952***	36.681**
	(1.045)	(2.019)	(7.410)	(14.889)
Republican President$_t$	-0.897**	-1.617**	-5.578**	-9.919
	(0.357)	(0.639)	(2.441)	(6.874)
Republican House$_t$	1.462**	3.154***	19.682***	24.346**
	(0.658)	(1.213)	(5.179)	(10.330)
Republican Senate$_t$	0.903**	1.325*	4.140	2.827
	(0.442)	(0.795)	(3.101)	(5.865)
smb*	-0.437	-0.442	-2.837	12.728
	(0.787)	(1.533)	(5.825)	(11.027)
hml*	0.467*	0.613	3.710**	6.925
	(0.278)	(0.477)	(1.777)	(4.435)
R_t^{oil}	0.925	2.663	-10.572	-7.796
	(2.491)	(3.993)	(8.468)	(9.615)
Constant	10.406***	22.849***	120.568***	180.340***
	(3.419)	(6.549)	(24.477)	(42.521)
Observations	715	715	715	691
Adjusted R^2 %	1.370	3.120	16.020	44.960
From	1959m6	1959m6	1959m6	1959m6
To	2018m12	2018m12	2018m12	2016m12

Note: This table presents results from the return predictability monthly regressions using the complete measure of energy policy uncertainty. $(d-p)_t$ is the natural logarithm of the aggregate VWCRSP dividend to price ratio, R_t^{oil} is the monthly return of the West Texas Intermediate price per barrel. Republican President$_t$ is a dummy variable that takes the value of 1 if the U.S. President at time t is Republican, and smb* and hml* are variables whose innovations correspond to factors smb and hml as in Maio and Santa-Clara (2012). Newey West standard errors for k lags reported in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Estimation sample is 1959m6 to 2018m12.

Table A.2: Consumption growth regressions - Complete Information Set

$$g(\theta)=\frac{1}{T}\sum_t\begin{Bmatrix}\ln\left(\frac{c_{t+k}}{c_t}\right)-\gamma_0-\gamma_1\mathcal{U}_{t\to t+1}-\gamma_2 R_{t\to t+k}-\gamma_3\text{term}_t\\ R_{t\to t+k}-\delta_0-\delta_1(d-p)_t-\delta_2\text{term}_t-\delta_3\text{def}_t\end{Bmatrix}\times\text{Instruments}_t=0$$

	(1) $\ln\left(\frac{c_{t\to c+12}}{c_t}\right), R_{t\to t+12}$	(2) $\ln\left(\frac{c_{t\to c+36}}{c_t}\right), R_{t\to t+36}$	(3) $\ln\left(\frac{c_{t\to c+60}}{c_t}\right), R_{t\to t+60}$	(4) $\ln\left(\frac{c_{t\to c+72}}{c_t}\right), R_{t\to t+72}$
γ_0	5.86***	17.07***	29.65***	36.23***
	(0.17)	(0.49)	(0.80)	(0.96)
γ_1	0.74***	1.28***	1.42***	1.21***
	(0.15)	(0.36)	(0.44)	(0.38)
γ_2	-0.61***	-2.67***	-4.77***	-5.87***
	(0.07)	(0.20)	(0.31)	(0.36)
γ_3	0.06***	0.12***	0.12***	0.12***
	(0.01)	(0.01)	(0.01)	(0.01)
δ_0	344.25***	342.41***	449.26***	477.57***
	(21.72)	(15.00)	(16.71)	(16.18)
δ_1	85.72***	83.01***	109.85***	114.58***
	(5.69)	(3.85)	(4.48)	(4.43)
δ_2	14.22***	16.35***	19.42***	18.65***
	(1.23)	(0.94)	(1.12)	(1.20)
δ_3	-38.87***	-27.41***	-22.63***	-23.65***
	(3.79)	(2.69)	(2.41)	(2.19)
Observations	672	648	624	612
Sample	1962m1-2017m12	1962m1-2015m12	1962m1-2013m12	1962m1-2012m12

Note: Parameters $\theta = \{\gamma_0, \gamma_1, \gamma_2, \gamma_3, \delta_0, \delta_1, \delta_2, \delta_3\}$ minimize $g(\theta)' I g(\theta)$ and Instruments: Constant, $\mathcal{U}_{t\to t+1}$, term, $d - p$, def. And, I is the identity matrix. This table presents results from the return and consumption growth predictability regressions on energy policy uncertainty. Parameters are estimated via GMM. $R_{t\to t+k}$ is the cumulative log return of the VWCRSP portfolio between month t and $t + k$. $\mathcal{U}_t$ is the level of energy policy uncertainty. Controls include the log dividend to price ration $(d - p)_t$, the term structure (term$_t$), the default spread (def$_t$). $\ln(c_{t+k}/c_t)$ is the growth on consumption between month t and $t + k$ measured as aggregate personal consumption expenses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Estimation sample of the predictability regression is 1962m1 to 2018m12.

Table A.3: Cross-sectional return regressions - Complete Information Set

	Size and Book-to-Market				
	(1)	(2)	(3)	(4)	(5)
μ_{mktrf}	0.007*** (0.002)	0.006*** (0.002)	0.006*** (0.002)	0.006*** (0.002)	0.007*** (0.002)
γ_{mktrf}	2.301* (1.359)	2.609* (1.507)	4.786** (1.941)	5.182*** (1.741)	8.620*** (2.290)
μ_{smb}		0.001 (0.001)	0.001 (0.001)		
γ_{smb}		2.575 (2.080)	5.586** (2.587)		
μ_{hml}		0.003** (0.001)	0.003** (0.001)		
γ_{hml}		3.941 (2.496)	0.806 (5.442)		
μ_{cma}			0.285*** (0.091)		
γ_{cma}			0.046 (0.105)		
μ_{rmw}			0.341*** (0.107)		
γ_{rmw}			0.107** (0.046)		
μ_{me}				0.002 (0.001)	0.002 (0.001)
γ_{me}				5.977*** (2.167)	10.443*** (2.761)
μ_{ia}				0.003*** (0.001)	0.003*** (0.001)
γ_{ia}				9.064** (4.349)	2.380 (5.291)
μ_{roe}				0.005*** (0.001)	0.006*** (0.001)
γ_{roe}				11.250*** (4.259)	-3.314 (6.553)
μ_{eg}					0.008*** (0.001)
γ_{eg}					42.471*** (14.441)
μ_{pol}	0.492*** (0.041)	0.491*** (0.041)	0.492*** (0.043)	0.489*** (0.043)	0.484*** (0.043)
γ_{pol}	-1.525*** (0.413)	-1.429*** (0.461)	-1.079*** (0.401)	-0.898** (0.385)	-0.890** (0.423)
Observations	468	468	468	468	468
MAE %	.26	.15	.12	.11	.09

Note: This table presents results from estimating the price of risk in expected return - covariance form by extending the CAPM model, Fama and French three and five factor models, and the q^4 and q^5 models. Estimations are performed via GMM in which factor loadings (covariances) and covariance prices of risk are estimated jointly. Factors smb, hml, cma, rmw, correspond to the Fama and French factors related to size, book-to-market, investment and profitability. Factors, me, ia, roe, and eg, correspond to factors related to size, investment, profitability, and expected investment growth. Factor pol corresponds to the innovations to energy policy uncertainty. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Estimation sample of the cross-sectional regression is 1956m1 to 2018m12.

Table A.4: Investment Cross-sectional Regressions - Complete Information Set

	(1) Inv_{it}	(2) Inv_{it}	(3) Inv_{it}	(4) Inv_{it}
$\text{Profitability}_{it}$	6.06***	2.73***	6.07***	2.74***
	(0.69)	(0.41)	(0.69)	(0.41)
$\log(Q_{it})$	0.54***	0.63***	0.52***	0.61***
	(0.02)	(0.02)	(0.02)	(0.02)
Size_{it}	-0.09***	-0.13***	-0.09***	-0.13***
	(0.01)	(0.01)	(0.01)	(0.01)
$\mathcal{U}_{t\to t+1}$	7.61**	6.31**	5.38**	3.81*
	(3.31)	(2.98)	(2.49)	(2.10)
$\mathcal{U}_{t\to t+1} \times \log(Q_{it})$			5.31*	5.95**
			(3.03)	(3.01)
Constant	1.94***	2.26***	1.94***	2.27***
	(0.06)	(0.06)	(0.06)	(0.06)
Observations	338253	338253	338253	338253
Adjusted R^2 (%)	4.46	24.63	4.47	24.64
Industry F.E.	No	Yes	No	Yes
From	2000m1	2000m1	2000m1	2000m1
To	2018m10	2018m10	2018m10	2018m10

Note: This table presents results from monthly cross-sectional investment regressions. Quarterly accounting variables are merged with pricing data with a two month lag to account for look ahead bias. Investment is defined as the difference between the cumulative quarterly capital expenditures (capxy) between quarters n and $n-1$ for $n > 1$ divided over total assets. Profitability is measured as operating income after depreciation (oiadpq) over the sum of book debt (dlcq+dlttq) and market equity (prccq $\times$ cshoq). Average Q is computed as the book value of debt plus equity (dlcq+dlttq+prc$\times$ cshoq) divided by total assets (atq), size is the natural logarithm of market equity (prc $\times$ cshoq), $\mathcal{U}_t$ corresponds to energy policy uncertainty. Standard errors clustered by gvkey reported in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

A.2.3 Mathematical Appendix

Proof. of proposition (6) The positive relation between investment and uncertainty in equation (1.2) can be written as

$$I_{it} = Y_{i,t+1}^{\frac{1-\beta}{\alpha}} \Big(\frac{r_{ft}}{\beta} (\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e) \Big)^{\frac{\beta}{\alpha}}$$

the derivative of investment with respect to energy policy uncertainty is

$$\frac{\partial I_{it}}{\partial \sigma^e} = \frac{\beta}{\alpha} Y_{i,t+1}^{\frac{1-\beta}{\alpha}} \Big(\frac{r_{ft}}{\beta} (\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e) \Big)^{\frac{\beta-\alpha}{\alpha}} \rho\sigma^M\sigma^e$$

given that $M_{t+1} > 0, w_{t+1} > 0$, we have that $\mathbb{E}[M_{t+1}w_{t+1}] = \mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e > 0$ so the sign of the partial derivative depends directly on the sign of ρ.

To compute the second derivative we forget about the term $\frac{\beta}{\alpha} Y_{i,t+1}^{\frac{1-\beta}{\alpha}} \rho\sigma^M$ which for the partial derivative is assumed constant and positive. The derivative of the remaining term $\Big(\frac{r_{ft}}{\beta} (\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e) \Big)^{\frac{\beta-\alpha}{\alpha}} \sigma^e$ is

$$\Big(\frac{r_{ft}}{\beta} (\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e) \Big)^{\frac{\beta-\alpha}{\alpha}} + \frac{\beta-\alpha}{\alpha} \Big(\frac{r_{ft}}{\beta} (\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e) \Big)^{\frac{\beta-2\alpha}{\alpha}} \rho\sigma^M\sigma^e \tag{A.23}$$

whose sign is ambiguous given that the term $\beta - \alpha$ can be either positive or negative. If $\beta - \alpha > 0$ the second derivative is always positive, however if $\alpha > \beta$ the second derivative is positive if

$$\begin{aligned}
&\Big(\frac{r_{ft}}{\beta} (\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e) \Big)^{\frac{\beta-\alpha}{\alpha}} + \frac{\beta-\alpha}{\alpha} \Big(\frac{r_{ft}}{\beta} (\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e) \Big)^{\frac{\beta-2\alpha}{\alpha}} \rho\sigma^M\sigma^e > 0 \\
&\frac{\alpha-\beta}{\alpha}\rho\sigma^M < \frac{r_{ft}}{\beta}(\frac{\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}]}{\sigma_e} + \rho\sigma^M) \\
&\rho\sigma^M(\frac{\beta}{r_{ft}}\frac{\alpha-\beta}{\alpha} - 1) < \frac{\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}]}{\sigma_e} \\
&\sigma_e > \frac{\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}]}{\rho\sigma^M(\frac{\beta}{r_{ft}}\frac{\alpha-\beta}{\alpha} - 1)} = \bar{\sigma}
\end{aligned} \tag{A.24}$$

where the last change in the inequality sign comes from the fact that $\frac{\beta}{r_{ft}}\frac{\alpha-\beta}{\alpha} < 1$. Since $\bar{\sigma} < 0$ the second derivative is positive for every value of σ_e. ☐

Proof. of proposition (7). If households consume the aggregate difference $Y - I$, expected consumption growth is given by

$$g_{t,t+1} = \mathbb{E}\Big[\frac{Y_{t+1}}{Y_t - I_t}\Big] \tag{A.25}$$

as all capital is depreciated by the end of $t+1$. Its derivative with respect to σ_e is

$$\frac{\partial g_{t,t+1}}{\partial \sigma_e} = \mathbb{E}\Big[\frac{\frac{\partial Y_{t+1}}{\partial \sigma_e}(Y_t - I_t) + Y_{t+1}\frac{\partial I_t}{\partial \sigma_e}}{(Y_t - I_t)^2}\Big] \tag{A.26}$$

Since the denominator is always positive, the sign depends on the term

$$\frac{\partial Y_{t+1}}{\partial \sigma_e}(Y_t - I_t) + Y_{t+1}\frac{\partial I_t}{\partial \sigma_e} \tag{A.27}$$

I will show that $\frac{\partial Y_{t+1}}{\partial \sigma_e}$ is non negative which concludes the proof. We can express future output Y_{t+1} as

$$\begin{aligned} \frac{\partial I_{it}}{\partial \sigma^e} &= \frac{\beta}{\alpha} Y_{i,t+1}^{\frac{1-\beta}{\alpha}} \Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e)\Big)^{\frac{\beta-\alpha}{\alpha}} \rho\sigma^M\sigma^e \\ Y_{t+1} &= \Bigg(\frac{\frac{\alpha}{\beta}\frac{\partial I_{it}}{\partial \sigma^e}}{\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e)\Big)^{\frac{\beta-\alpha}{\alpha}} \rho\sigma^M\sigma^e}\Bigg)^{\frac{\alpha}{1-\beta}} \end{aligned} \tag{A.28}$$

whose derivative is

$$\begin{aligned} \frac{\partial Y_{t+1}}{\partial \sigma_e} = & \underbrace{\frac{\alpha}{1-\beta}\Big[\frac{\frac{\alpha}{\beta}\frac{\partial I_{it}}{\partial \sigma^e}}{\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma_e)\Big)^{\frac{\beta-\alpha}{\alpha}} \rho\sigma^M\sigma^e}\Big]^{\frac{\alpha+\beta-1}{1-\beta}}}_{>0} \times \\ & \frac{1}{\Big(\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e)\Big)^{\frac{\beta-\alpha}{\alpha}} \rho\sigma^M\sigma^e\Big)^2} \times \Big[\frac{\alpha}{\beta}\frac{\partial^2 I_t}{\partial \sigma_e^2}\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e)\Big)^{\frac{\beta-\alpha}{\alpha}} \rho\sigma^M\sigma^e \\ & - \frac{\alpha}{\beta}\frac{\partial I_t}{\partial \sigma_e}\Big[\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e)\Big)^{\frac{\beta-\alpha}{\alpha}} + \frac{\beta-\alpha}{\alpha}\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e)\Big)^{\frac{\beta-2\alpha}{\alpha}} \rho\sigma^M\sigma^e\Big]\Big] \end{aligned} \tag{A.29}$$

Now, the sign of the term depends on the sign of

$$\begin{aligned} & \frac{\alpha}{\beta}\frac{\partial^2 I_t}{\partial \sigma_e^2}\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e)\Big)^{\frac{\beta-\alpha}{\alpha}} \rho\sigma^M\sigma^e \\ & - \frac{\alpha}{\beta}\frac{\partial I_t}{\partial \sigma_e}\Big[\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e)\Big)^{\frac{\beta-\alpha}{\alpha}} + \frac{\beta-\alpha}{\alpha}\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}] + \rho\sigma^M\sigma^e)\Big)^{\frac{\beta-2\alpha}{\alpha}} \rho\sigma^M\sigma^e\Big] \end{aligned}$$

I will prove that this term is non negative by contradiction. Assume it is strictly negative, dividing both sides by $\frac{\alpha}{\beta}\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}]+\rho\sigma^M\sigma^e)\Big)^{\frac{\beta-\alpha}{\alpha}}\rho\sigma^M\sigma_e > 0$

$$\frac{\partial^2 I_t}{\partial\sigma_e^2}-\frac{\partial I}{\partial\sigma_e}\Big[\frac{1}{\rho\sigma^M\sigma_e}+\frac{\beta-\alpha}{\alpha}\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}]+\rho\sigma^M\sigma^e)\Big)^{-1}\Big]<0 \tag{A.30}$$

given that the first derivative of investment on energy policy uncertainty is positive,

$$\frac{\frac{\partial^2 I_t}{\partial\sigma_e^2}}{\frac{\partial I_t}{\partial\sigma_e}}<\frac{1}{\rho\sigma^M\sigma_e}+\frac{\beta-\alpha}{\alpha}\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}]+\rho\sigma^M\sigma^e)\Big)^{-1} \tag{A.31}$$

replacing the definition of the first and second derivatives

$$\frac{\frac{\beta}{\alpha}Y_{i,t+1}^{\frac{1-\beta}{\alpha}}\Big[\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}]+\rho\sigma^M\sigma^e)\Big)^{\frac{\beta-\alpha}{\alpha}}+\frac{\beta-\alpha}{\alpha}\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}]+\rho\sigma^M\sigma^e)\Big)^{\frac{\beta-2\alpha}{\alpha}}\rho\sigma^M\sigma^e\Big]}{\frac{\beta}{\alpha}Y_{i,t+1}^{\frac{1-\beta}{\alpha}}\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}]+\rho\sigma^M\sigma^e)\Big)^{\frac{\beta-\alpha}{\alpha}}\rho\sigma^M\sigma^e}<$$
$$\frac{1}{\rho\sigma^M\sigma_e}+\frac{\beta-\alpha}{\alpha}\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}]+\rho\sigma^M\sigma^e)\Big)^{-1} \tag{A.32}$$

which is true if

$$\begin{aligned}&\frac{1}{\rho\sigma^M\sigma_e}+\frac{\beta-\alpha}{\alpha}\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}]+\rho\sigma^M\sigma^e)\Big)^{-1}<\\&\frac{1}{\rho\sigma^M\sigma_e}+\frac{\beta-\alpha}{\alpha}\Big(\frac{r_{ft}}{\beta}(\mathbb{E}[M_{t+1}]\mathbb{E}[w_{t+1}]+\rho\sigma^M\sigma^e)\Big)^{-1}\end{aligned} \tag{A.33}$$

which is a contradiction, so it must be that $\frac{\partial Y_{t+1}}{\partial\sigma_e}\geqslant 0$ which means that $\frac{\partial g_{t+1}}{\partial\sigma_e}>0$. □

Proof. of propositions (8) and (9).

The proof proceeds by inspection of the term I_{it} inside the expression, it can be observed that increasing investment decreases the numerator given the term $I_{it}^{-\frac{\alpha}{\beta}}$ for $\alpha>0,\beta>0$, and equivalently investment increases the term in the denominator $\frac{I_{it}}{Y_{i,t+1}}\Big(1+a\frac{I_{it}}{K_{it}}\Big)$ of equation (1.4).

□

Appendix B

Appendix - Tweeting for Money: Social Media and Mutual Fund Flows

B.1 Tweet Classification and Examples

In this appendix we explain how we classify tweets and provide some examples. Classifying tweets into positive and negative requires knowing how these tweets are written, what information they contain, and who they are directed to. The positiveness of a tweet depends on whether the tweet contains financial information that can be translated into good or bad news for the fund family, or other type of information which, even if positive, may not be related to the business in which these companies operate. Loughran and McDonald (2011) describe in detail this issue when classifying good or bad information. Dictionaries, or other techniques used in textual analysis need to be corrected when considering financial information. Information related to finance and economics may contain specialized jargon, or even words with a different tone and interpretation than in other fields.[1] To overcome this issue we develop a two stage classification procedure in which tweets are first classified into one of two types: *financial* tweets, which are more likely to contain financial information; and *nonfinancial* tweets.

We start by explaining the data pre-processing process with an example. Most of the textual analysis applications in the literature consider a

[1] The best example is the word liability. While in dictionaries such as the Harvard IV-4 dictionary the word is classified as being negative, in a more financial context it has a different tone.

bag of words approach, in which the position of the word within a sentence is not relevant and the only important feature is the appearance or frequency of a word in the text. Since the language used in tweets is limited by the 140-character length restriction, just the presence of words might not be informative enough to classify them.[2] At the same time the *informality* of communications increases the use of words with less defined tonal categories, and provides less informative features to predict labels. To obtain a more informative set of features from tweets, we consider both the appearance of a word as well as its role in the sentence - also known as Part of Speech (POS).[3]

The in- and out-of-sample performance of machine learning depends on the algorithm chosen by the researcher. To avoid subjectivity on the choice of an algorithm we use six different algorithms: Naive-Bayes classifier, Multinomial-Naive-Bayes classifier, Bernoulli Naive-Bayes classifier, Stochastic Gradient Descent, Support Vector Machines, and Logistic Regression. We then consider a voting scheme that consists of classifying each tweet with the most voted label among the different algorithms. If three algorithms classify a tweet as positive and three as negative, we consider the tweet to have a neutral tone. The procedure also provides us with a measure of agreement between the algorithms. Even though this methodology increases the computational efforts, we believe it reduces the subjective judgement of the decision maker, and provides a natural way to quantify the confidence of a classification.

The selection of algorithms satisfies two criteria: their implementation is relatively fast using available programming packages, and the replication of results among platforms can be attained by providing detailed description of the algorithms. As argued by Loughran and McDonald (2016), the current state of textual analysis in Financial Economics needs to stand for replicability of results rather than using more sophisticated machinery in the analysis of textual data. We decide to rule out more accurate classifiers and keep the analysis replicable by presenting a detailed explanation of the algorithms.

Our analysis focuses on positiveness - a measure of how positive a text is - which can be decomposed as positiveness conditional on a tweet having financial information, positiveness conditional on a tweet having non-financial information, and the unconditional positiveness of tweets.

We start by extracting two important features of each tweet: Words,

[2]Twitter changed to 240 the character limit of tweets only starting in 2017.

[3]One of the grammatical groups, such as noun, verb, and adjective, into which words are divided depending on their use. Retrieved from: https://dictionary.cambridge.org/dictionary/english/part-of-speech

and Part of Speech (POS). We start by applying a tokenization based on regular expressions to automatically split the text into words. We proceed by using a POS tagger (an algorithm that tags each word with its more likely POS) to identify the role of each word within the sentence. The following example describes the procedure we use:

Figure B.1: Example of a financial tweet posted by Bloomberg @business on September 27 2017, 14:00.

The first step consists of tokenizing the tweet to isolate its components. We use a special tokenization procedure to account for hyperlinks, emoticons, and punctuation. The tokenization splits the tweet as follows:

Fidelity, Chairman, and, CEO, Abigail, Johnson, discusses, the, struggle, between, active and, passive, investing, https://www.bloomberg.com/...

Once the tweet is tokenized, we apply a POS tagger which applies an optimization algorithm that maximizes the likelihood of tuples of the form $(token, POS)$ to appear in a sentence. After the POS tagger is applied, the tweet becomes:

$$\underbrace{\text{Fidelity}}_{\text{proper noun, singular}}\ \underbrace{\text{Chairman}}_{\text{proper noun, singular}}\ \underbrace{\text{and}}_{\text{coordinating conjunction}}\ \underbrace{\text{CEO}}_{\text{proper noun, singular}}\ \underbrace{\text{Abigail}}_{\text{proper noun, singular}}$$

$$\underbrace{\text{Johnson}}_{\text{proper noun, singular}}\ \underbrace{\text{discusses}}_{\text{verb, 3rd person, singular, present}}\ \underbrace{\text{the}}_{\text{determiner}}\ \underbrace{\text{struggle}}_{\text{noun singular}}$$

$$\underbrace{\text{between}}_{\text{preposition/subordinating conjunction}}\ \underbrace{\text{active}}_{\text{adjective}}\ \underbrace{\text{and}}_{\text{coordinating conjunction}}\ \underbrace{\text{passive}}_{\text{adjective}}\ \underbrace{\text{investing}}_{\text{noun singular}}$$

$$\underbrace{\text{https://www.bloomberg.com/...}}_{\text{noun singular}}$$

After tokenizing and extracting the POS of every tweet in our database we proceed to extract the most common features. We do this by calculating the frequency of each tuple $(token, POS)$ and select the most common 4000 features.

We proceed by first classifying tweets depending on whether they contain financial or nonfinancial information, and then classify them into positive or negative. The machine learning algorithms we use are supervised algorithms. They require an initial set of tweets labelled according to whether they contain financial information or not, and whether they are positive or negative. The algorithms then find common patters which are applied to classify unlabelled tweets. We train our algorithms with a sample of 10,000 tweets manually classified by two research assistants (undergraduate students in economics and management science respectively). To ensure the training sample has enough tweets from all possible categories in both dimensions we randomly select them from the accounts of the Financial Times (FT) and The Wall Street Journal (WSJ). We use this source of tweets instead of those posted by fund families since negative information is less likely to be disclosed.

Table (B.1) presents shows the classification of tweets in the training sample. 42 percent of all tweets re classified as financial, and within this category 62 percent are classified as positive. To classify non-financial tweets into positive or negative tweets, we use a sample of 10,000 tweets randomly selected from the Sentiment Analysis Training Corpus Dataset from the University of Michigan which contains more than million and a half of tweets classified into positive or negative tweets. There is a large corpora of labelled tweets available for research, however this corpora contains Twitter information that covers a widely range of topics. Since fund families also post financial information, using a database with non-financial tweet to train our algorithms can lead to important biases in the results. Non-financial tweets on the other hand are not affected by this issue, and their tone can be inferred by training the same algorithms with a more general corpora of tweets. We train each one of the 6 algorithms using the same training samples. For a more robust analysis, we sort randomly the training sample and divide it into 4 sets of equal length. We then train and test each one of the algorithms using three of the sets as training set and 1 set as test.

		Sample 1	Sample 2	Total
Financial	Positive	1187	664	1851
	Negative	728	398	1126
Non financial		3085	3938	7023
Total				10000

Table B.1: Manual classification of tweets in the training sample

Using this methodology we classify each tweet into two topic categories {financial, nonfinancial} and two tone categories {positive, negative} using either our training sample if the tweet has been classified as containing financial information, or the more general training sample if the tweet has been classified as nonfinancial. For each classification, we obtain a measure of confidence based on the degree of agreement among the classifiers. The confidence of a tweet's classification as c is computed as:

$$w_i^c = \frac{\text{Number of algorithms that classify tweet } i \text{ with label } c}{6} \tag{B.1}$$

Table (B.2) shows the classification of tweets posted by fund families in the final sample by our procedure.

	Financial	**Non financial**	**Total**
Positive	217,444	549,735	767,179
Negative	63,168	135,807	198,975
Total	280,612	685,542	966,154

Table B.2: Algorithm classification of tweets by fund families

Below, we present four tweets posted by financial media, and three tweets posted by asset management companies to show the rationale behind estimating the confidence in our classifications. We start by presenting two tweets with a confidence of 1 and 0.67 in their topic classification, for the first tweet the 6 algorithms classified it as being financial, while the second one only 4 out of 6 algorithms coincided classifying it as financial. The first tweet (B.2), posted by The Wall Street Journal is systematically classified by financial by all the algorithms due to the presence of features such as (*stocks*, *Noun Plural*), (*hedge*, *Noun Singular*), (*funds*, *Noun Plural*), (*Apple*, *Noun Singular*), (*J.P.*, *Proper Noun Singular*), and (*Morgan*, *Proper Noun Singular*), which in the training set are more likely to represent a financial tweet. The second tweet (B.3), posted by Seeking Alpha, contains the features (*JPMorgan*, *Proper Noun Singular*), (*Numbers*, *Proper Noun Plural*), and (*CCAR*, *Proper Noun Singular*), which albeit Finance related, are not enough to make all algorithms infer it as a financial tweet. Although both tweets can be considered as financial by human classification, the weighting scheme places more trust on the correct classification of the first tweet rather than the second. This is because the second tweet contains more informative features extracted in the first step during the training stage.

The next two tweets written by our external sources, are classified according to their tone. Both tweets were classified in the first step as financial tweets, which requires using our own training sample to classify them according to their tone. The first tweet (B.4) posted by Financial News, is classified as positive with a confidence of one for containing the features (*Why*, *Adverb*), (*rule*, *Verb*), (*Wall*, *Noun Singular*), and (*Street*, *Noun Singular*), which are enough to make all algorithms coincide with a positive classification. The last tweet, posted also by Financial News (B.5), is classified as negative only with a confidence of only 4 out of 6 algorithms. The tweet can be understood as containing negative information, although from the machine learning algorithms the only word that is informative about the negative tone is *scorn*.

We repeat the analysis with some tweets posted by asset management companies in our sample. The first example (B.6) corresponds to a post written by Northern Trust which is classified as negative by the six algorithms. The feature (*challenge*, *Noun Singular*) together with (*growth*, *Noun Singular*) are informative enough to make all algorithms coincide with the classification. The second tweet (B.7) written by State Farm is classified as positive with a confidence of 1. In this tweet the features (*outperforms*, *Verb 3rd person singular*), (*industry*, *Noun Singular*), and (*average*, *Noun Singular*) are informative enough to make all algorithms coincide with the classification.

Finally, tweet (B.8) written by Pax World is classified as financial with a confidence of only 0.67 (4 out of 6 algorithms agreed on the classification). The tweet only contains features such as (*business*, *Noun Singular*) and (*work*, *Verb*) which are not informative enough to make all algorithms agree with the classification.

B.2 Data pre-processing and Machine Learning algorithms

B.2.1 Tweet Tokenization

The english-language that tweets have is not the same style of english that investors will find in SEC fillings, shareholder letters or press. Tweets are designed to provide a more informal type of communication. This makes the language in tweets more familiar to people, but it gives room to spelling mistakes, typos, words not separated by spaces, punctuation, and emoticons. To obtain a bag-of-words representation of a tweet before applying the POS algorithm we tokenize each one of the tweets. In this section we provide the exact procedure used for the tweet tokenization.

We provide the exact regular expressions used by the algorithm to deal with each procedure. [4]

1. URLs contain symbols that are not encoded in the standard utf-8 format. Before decoding non utf-8 characters we make sure those one present in the URL are kept.

2. usernames

```
(?:@[\w_]+)
```

3. Replace characters that are repeated more than 3 times consecutive with only 3 times by replacing the regular expression `\1\1\1` with `(.)\1{2,}`.

4. Shorten problematic sequences of non-alpha numeric characters to only 3 occurrences.

```
([^a-zA-Z0-9])\1{3,}'
```

5. HTML tags

```
<[^>\s]+>
```

6. ASCII Arrows

```
[\-]+>|<[\-]+
```

7. Hashtags

```
(?:\#+[\w_]+[\w\'_\-]*[\w_]+)
```

8. numbers, including fractions and decimals

```
(?:[+\-]?\d+[,/.:-]\d+[+\-]?)
```

9. words without apostrophes or dashes

[4]The procedure is explained based on our interpretation of the source code of the class TweetTokenize from the Natural Language Toolkit (NLTK) in python. For the original code visit the official documentation available at: http://www.nltk.org/_modules/nltk/tokenize/casual.html#TweetTokenizer

```
(?:[\w_]+)
```

10. Ellipsis dots

```
(?:\.(?:\s*\.){1,})
```

11. remaining characters that are not white space

```
(?:\S)
```

B.2.2 Part of Speech Tagger

We use the default Part of Speech Tagger function provided in the NLTK library. The tagger is trained on the Penn Treebank database provided by the Linguistic Data Consortium at the University of Pennsilvania. The Penn Treebank contains 2499 stories from a three year Wall Street Journal (WSJ) database of 98732 different stories. It provides approximately 7 million words of POS tagged among some other interesting resources. [5] [6]

The function implements a maximum-entropy tagger (Ratnaparkhi 1990). Details on the notation are presented in the next section of the appendix. Consider a tweet t with grammar elements of the form $(\text{word}, \text{tag}) \in \mathcal{G}(t)$. The maximum entropy tagger maximizes the conditional probability of having tags $\text{tag}_1 ... \text{tag}_N$ if the tweet has words $\text{word}_1 ... \text{word}_N$.

$$\arg \max_{\text{tag}_1 ... \text{tag}_N} P(\text{tag}_1 ... \text{tag}_N | \text{word}_1 ... \text{word}_N) = \arg \max_{\text{tag}_1 ... \text{tag}_N} \frac{P(\text{word}_1 ... \text{word}_N | P(\text{tag}_1 ... \text{tag}_N)}{P(\text{tag}_1 ... \text{tag}_N} \tag{B.2}$$

in order to maximize the above conditional probability the algorithm exploits the notion of contexts which we define by C. [7] The conditional

[5]Marcus, Mitchell, et al. Treebank-3 LDC99T42. Web Download. Philadelphia: Linguistic Data Consortium, 1999.

[6]http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.9.8216&rep=rep1&type=pdf

[7]Context $C_i$is defined as a group of words and relations between words with their respective POS that surround word_i, the context is estimated by analyzing a large corpora of documents in which word_i appears such like the Penn Treedatabase. The set of features can be arbitrarily complex and may include everything since encoding evidence of a word having a particular tag, to include information about the whole document or atomic features.

probability can be approximated as:

$$P(\text{tag}_1...\text{tag}_N|\text{word}_1...\text{word}_N) \approx \prod_{i=1}^{N} P(\text{tag}_i|C_i) \tag{B.3}$$

given a list of words $\{\text{word}_i\}_{i=1}^{N}$ and their respective contexts $\{C_i\}_{i=1}^{N}$. The algorithm estimates each probability as

$$P(\text{tag}|C) = \frac{1}{Z(C)} \exp\{\sum_{i=1}^{N} \lambda_i \text{feature}_i(C, \text{tag})\} \tag{B.4}$$

where feature_i is defined as:

$$f_i(C, \text{tag}) = \begin{cases} 1, & \text{if word}_i \in C \text{ and tag}_i = \text{tag} \\ 0, & \text{otherwise} \end{cases} \tag{B.5}$$

and $Z(C)$ is a normalization constant to ensure P is a probability distribution. The optimal sequence of tags is then calculated using *beam search*, an heuristic that prunes branches of the tree spanned by all sequences of tag_i.

B.2.3 Notation

For every tweet t denote $\mathcal{G}(t)$ the set of words tokenized and their respective POS. Define by $\mathcal{F}$ the set of relevant features. Every tweet is represented by the following set:

$$f(t) = \{(x, I_{\mathcal{G}(t)}(x)) : x \in \mathcal{F}\} \tag{B.6}$$

where $I_{\mathcal{G}(t)(x)}$ is the indicator function

$$I_{\mathcal{G}(t)}(x) = \begin{cases} 1, & \text{if } x \in \mathcal{G}(t) \\ 0, & \text{if } x \notin \mathcal{G}(t) \end{cases} \tag{B.7}$$

The set $f(t)$ contains information about the presence or absence of each feature in every tweet.

Features

The input of all of the classifiers used in the empirical analysis is a list of labelled tweets, their relevant features and the label. Since the amount of words used in tweeting is restricted to 140 characters, using just the presence of words in a tweet is not an informative feature. We expand the feature set by considering not only the appearance of words within tweets but also the Part of Speech (POS) of the word. The POS of a word determines its role within a sentence. Words are classified into one of the following categories:

Abbreviation	Part of Speech	Example
CC	coordinating conjunction	and, but, or
CD	cardinal digit	1,99
DT	determiner	a, the, every
EX	existential clause	there is a problem
FW	foreign word	laissez-faire
IN	preposition/subordinating conjunction	before, after, until
JJ	adjective	big
JJR	adjective, comparative	bigger
JJS	adjective, superlative	biggest
LS	list marker	1)
MD	modal	could, will
NN	noun, singular	desk
NNS	noun plural	desks
NNP	proper noun, singular	Harrison
NNPS	proper noun, plural	Americans
PDT	predeterminer	all the kids
POS	possessive ending	parent's
PRP	personal pronoun	I, he, she
PRP$	possessive pronoun	my, his, hers
RB	adverb	very, silently,
RBR	adverb, comparative	better
RBS	adverb, superlative	best
RP	particle	give up
TO	to go	to the store.
UH	interjection	Yoo-hoo
VB	verb, base form	take
VBD	verb, past tense	took
VBG	verb, gerund/present participle	taking
VBN	verb, past participle	taken
VBP	verb, singular. present, non-3rd person	take
VBZ	verb, 3rd person sing. present	takes
WDT	wh-determiner	which
WP	wh-pronoun	who, what
WP$	possessive wh-pronoun	whose
WRB	wh-abverb	where, when

Classification

This section explains the procedure used to classify tweets. We train our classifiers using a database of 10,000 tweets randomly selected from the accounts of the Financial Times (@FT), Wall Street Journal (@WSJ). We randomly give to two undergradutae students in economics and management science an independent sample of 5000 tweets each. The use of dictionaries is a common technique in textual analysis to infer the sentiment or tone from documents. We decided not to use standard dictionaries and rely our analysis in machine learning algorithms for the following reasons:

1. Dictionaries have been applied in studies concerning longer textual financial information. Tweets are limited to 140 characters which makes the vo-

cabulary used an endogenous decision by the person writing the tweet.

2. Financial dictionaries were made by analyzing financial disclosure documents such as 10-K fillings. Even though our tweets are financial and may contain similar words, the whole language used in Twitter is less formal.

However the accuracy of machine learning algorithms in understanding financial information has been widely criticized. Instead of relying our estimations in a single algorithm, we use a custom classifier based on a vote-scheme of 6 different classifiers. The algorithm returns the most voted label for a tweet together with a confidence indicator represented as the degree of consensus between the classifiers. There is more confidence of a tweet being positive if all 6 classifiers have agreed on the same label for the tweet than if 4 have agreed on the tweet having a positive tone and 2 of having a negative tone. We will use this confidence measure to weight our results when aggregating tweet labels over a period of time.

Naive Bayes Classifiers

The Naive Bayes Classifier is the oldest of the algorithms used to classify objects. The main assumption of this machine learning technique is that the appearance of words is independent, and therefore it helps reduce the dimensionality of the problem. In its simplest interpretation, the algorithm applies the Bayes rule to determine the likelihood that the features of a tweet $f(t)$ belong to a specific topic or tone. The probability of a tweet having label l can be calculated using Bayes rule: [8]

$$P(l|f(t)) = \frac{P(f(t)|l)}{P(l)} \tag{B.8}$$

where $P(f(t)|l)$ is the probability of observing features $f(t)$ in a tweet labeled as l, and $P(l)$ is the unconditional probability of observing a tweet labeled as l. The algorithm assumes features within a tweet are independent, and therefore given a label l the conditional probability $P(f(t)|l)$ can be calculated as:

$$P(f(t)|l) = \prod_{x \in f(t)} P(x|l) \tag{B.9}$$

where $P(x|l)$ is the probability of observing feature x in a tweet labeled as l.

[8] The classification can be extended to any finite set of classifications (e.g Positive, Neutral, or Negative)

The Naive Bayes classifier has a strong assumption of independence between features. The multinomial Naive Bayes classifier goes one step beyond and imposes a multinomial distribution on $P(x|l), x \in f(t)$, rather than calculating the probability based on the frequency of labeled tweets in the data.

Linear Support Vector Classification

To apply the linear SVC algorithm we convert the relevant features of a tweet $f(t)$ into an edge in a $|\mathcal{F}|$ dimensional hypercube. Using any arbitrary sort of the set $\mathcal{F} = \{f_1, f_2, ...f_{|\mathcal{F}|}\}$ the coordinates of the features $f(t)$ of a tweet are:

$$c(t) = \begin{bmatrix} 2 \times I_{\mathcal{G}(t)(f_1)} - 1 \\ 2 \times I_{\mathcal{G}(t)(f_2)} - 1 \\ \vdots \\ 2 \times I_{\mathcal{G}(t)(f_{|\mathcal{F}|})} - 1 \end{bmatrix} \tag{B.10}$$

If a tweet contains feature f_i a 1 will be assigned as the coordinate in the i dimensions, and a -1 otherwise. This parametrization allows us to map tweets into a high dimensional space, and have a defined gap between edges that will allow us to build a hyperplane splitting sets. If we have a training sample codified as points (c_i, y_i) where c_i corresponds to a set of coordinates for each training element, and y_i a variable that takes the value of 1 or -1 for each one of the labels. If the training set is linearly separable there exists a vector w such that the following two hyperplanes

$$wc - b = 1 \tag{B.11}$$

$$wc - b = -1 \tag{B.12}$$

are separated by a distance $\frac{2}{||w||}$. The goal of the algorithm is to minimize the distance between both hyperplanes, once the hyperplane is built ,the classification of each $c(t)$ is completely determined by its surrounding neighbours c_i which are called support vectors.

B.2.4 Tweets from financial media accounts

Figure B.2: Tweet classified as financial with a confidence of 1. The tweet corresponds to The Wall Street Journal (@wsj) written on February 22 of 2011. The tweet was included in the database after mentioning the asset management company J.P. Morgan (@jpmorgan) in the text.

Figure B.3: Tweet classified as financial with a confidence of 0.67 (4 of 6 algorithms agreed on the financial topic). The tweet corresponds to Seeking Alpha (@SeekingAlpha) written on June 27 2016. The tweet was included in the database after mentioning the asset management company J.P. Morgan (@jpmorgan) in the text.

Figure B.4: Tweet classified as positive with a confidence of 1. The tweet corresponds to Financial News (@FinancialNews) written on December 12 2016. The tweet was included in the database after mentioning the asset management company Goldman Sachs (@GoldmanSachs) in the text.

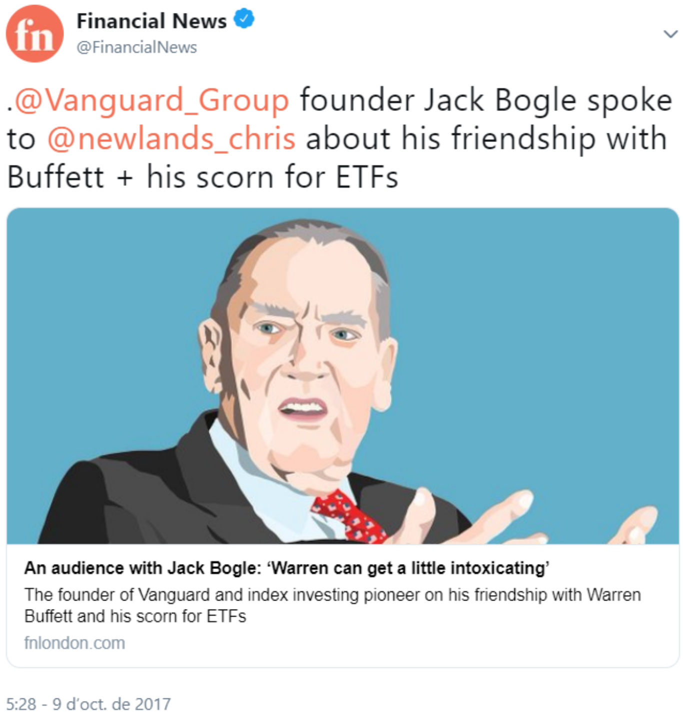


Figure B.5: Tweet classified as negative with a confidence of 0.67 (4 of 6 algorithms agreed on the negative tone). The tweet corresponds to Financial News (@FinancialNews) written on October 9 2017. The tweet was included in the database for mentioning Vanguard Group (@Vanguard_Group) in the text.

B.2.5 Tweets from asset management companies

Figure B.6: Tweet classified as negative with a confidence of 1. The tweet was written by asset management company Northern Trust (@Northern-Trust) on October 1 2013.

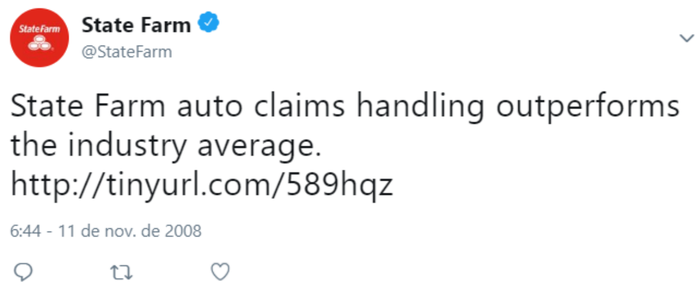


Figure B.7: Tweet classified as positive with a confidence of 1. The tweet was written by asset management company State Farm (@StateFarm) on November 11 2008.

Figure B.8: Tweet classified as financial with a confidence of 0.67 (4 of 6 algorithms agreed on the topic). The tweet was written by asset management company PaxWorld (@PaxWorld) on November 18 2015. PaxWorld funds are adviced by Impax Asset Management LLC, formerly Pax World Management LLC

B.2.6 Nonfinancial tweets

Figure B.9: Tweet classified as nonfinancial with a confidence of 0.67 (4 of 6 algorithms agreed on the tone). The tweet was written by asset management company JP Morgan (@jpmorgan) on January 10 2019.

Appendix C

Appendix - Learning from Quant (Qual)-itative Information

C.1 Mathematical Appendix

Proof. of Proposition (6). The extrema of a functional of the form $J[y] = \int_a^b H(y(x), y'(x), x)dx$ with boundary conditions $y(a) = y_a$ and $y(b) = y_b$ is given by the solution to the differential equation $H_y(y(x), y'(x), x) - \frac{d}{dx}H_{y'}(y(x), y'(x), x) = 0$ where $H_y, H_{y'}$ are the partial derivatives of H with respect to y and y' (See Dacorogna (1992, Theorem 2.1)). For the particular case in which the functional is of the form of equation (3.2) the Euler-Lagrange equation can be written as:

$$\frac{d^2G}{dx^2}\frac{dF}{dx} - \frac{dG}{dx}\frac{d^2F}{dx^2} = 0$$

or for a normal prior with mean μ and variance σ^2

$$\frac{d^2G}{dx^2} + (\frac{x-\mu}{\sigma^2})\frac{dG}{dx} = 0$$

The integrating factor $\exp\{\frac{x^2-2\mu x}{2\sigma^2} + C\}$ where C is a constant can be rewritten as $C_1 \exp\{\frac{1}{2}(\frac{x-\mu}{\sigma})^2\}$ where $C_1 = \exp\{-(\frac{\mu^2}{2\sigma^2} + C)\}$. This leads to the following general solution:

$$\frac{dG}{dx} = C_2 \exp\{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\} \tag{C.1}$$

The value of the constant C_2 can be calculated based on the boundary conditions. There is no guarantee for the function $\frac{dG}{dx}$ to be smooth, however we can split the differential equation into two differential equations with boundaries at $(-\infty, \tilde{\alpha}]$ and $(\tilde{\alpha}, \infty)$ as follows:

For the interval $(-\infty, \tilde{\alpha}]$

$$C_2 \int_{-\infty}^{\tilde{\alpha}} \exp\{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\}dx = \frac{1}{2} \to C_2 = \frac{1}{2\sqrt{2\pi\sigma}\Phi(\frac{\tilde{\alpha}-\mu}{\sigma})}$$

and the interval $(\tilde{\alpha}, \infty)$

$$C_2 \int_{-\infty}^{\tilde{\alpha}} \exp\{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\}dx = \frac{1}{2} \to C_2 = \frac{1}{2\sqrt{2\pi\sigma}(1-\Phi(\frac{\tilde{\alpha}-\mu}{\sigma}))}$$

replacing both solutions to C_2 in equation (C.1) gives us the p.d.f. in Proposition (6), finally integrating from $(-\infty, x)$ gives us the c.d.f. □

Proposition 14. $z\Phi(z) + \phi(z) \geqslant 0$

Proof. [1]

$$z\Phi(z) + \phi(z) = \int_{-\infty}^{z} z\phi(x)dx + \phi(z) \geqslant \int_{-\infty}^{z} x\phi(x)dx = -\phi(x)|_{-\infty}^{x} + \phi(x) = 0$$

□

Proof. of Proposition (7): If x has the p.d.f. in 3.7

$$\begin{aligned}
\Psi(t) &= \mathbb{E}\Big(e^{tx}\Big) \\
&= \int_{-\infty}^{\infty} e^{tx}\frac{dG(x)}{dx}dx \\
&= \frac{1}{\sqrt{8\pi\sigma}}\left(\int_{-\infty}^{\tilde{\alpha}} \frac{1}{\Phi(\frac{\tilde{\alpha}-\mu}{\sigma})}e^{tx-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}dx + \int_{\tilde{\alpha}}^{\infty} \frac{1}{(1-\Phi(\frac{\tilde{\alpha}-\mu}{\sigma}))}e^{tx-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}dx\right)
\end{aligned}$$

To derive the m.g.f. let us focus first on the term:

$$\int \exp\Big\{tx - \frac{1}{2\sigma^2}(x^2 - 2x\mu + \mu^2)\Big\}dx$$

$$\int \exp\Big\{-\frac{1}{2\sigma^2}(x^2 - 2x\mu - 2x\sigma^2\mu + \mu^2)\Big\}dx$$

$$\int \exp\Big\{-\frac{1}{2\sigma^2}(x^2 - 2x(\mu + \sigma^2 t) + \mu^2)\Big\}dx$$

$$\int \exp\Big\{-\frac{1}{2\sigma^2}(x^2 - 2x(\mu + \sigma^2 t) + (\mu + \sigma^2 t)^2 - (\mu + \sigma^2 t)^2 + \mu^2)\Big\}dx$$

$$\int \exp\Big\{-\frac{1}{2\sigma^2}((x - (\mu + \sigma^2 t))^2 - (\mu + \sigma^2 t)^2 + \mu^2)\Big\}dx$$

[1] I thank Dilip Sarwate for his elegant proof available in math.stackexchange.com

Moving out of the integral all terms that do not depend on x, and plugging it into the original expression:

$$\frac{e^{-\frac{(\mu^2-(\mu+\sigma^2 t)^2)}{2\sigma^2}}}{2}\left(\frac{1}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}\int_{-\infty}^{\bar{\alpha}}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}(\frac{x-(\mu+\sigma^2 t)}{\sigma})^2}dx+\frac{1}{(1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma}))}\int_{\bar{\alpha}}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}(\frac{x-(\mu+\sigma^2 t)}{\sigma})^2}dx\right)$$

$$\frac{e^{-\frac{(\mu^2-(\mu+\sigma^2 t)^2)}{2\sigma^2}}}{2}\left(\frac{\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}+\frac{1-\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{(1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma}))}\right)$$

$$\frac{e^{-\frac{\mu^2-\mu^2-2\mu\sigma^2 t+\sigma^4 t^2}{2\sigma^2}}}{2}\left(\frac{\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}+\frac{1-\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{(1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma}))}\right)$$

$$\Psi(t)=\frac{e^{\mu t+\frac{\sigma^2 t^2}{2}}}{2}\left(\frac{\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}+\frac{1-\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{(1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma}))}\right)$$

□

Proof. of Proposition (8): The first and second derivatives of $\Psi(t)$ are:

$$\frac{d}{dt}\Psi(t)=\frac{\exp\{\mu t+\frac{\sigma^2 t^2}{2}\}}{2}\times$$
$$\Big((\mu+\sigma^2 t)\Big(\frac{\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}+\frac{1-\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{(1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma}))}\Big)-\sigma\Big(\frac{\phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}-\frac{\phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}\Big)\Big)$$

$$\frac{d^2}{dt^2}\Psi(t)=\frac{\exp\{\mu t+\frac{\sigma^2 t^2}{2}\}}{2}\times$$
$$\Big(2(\mu+\sigma^2 t)\sigma\Big(\frac{\phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}-\frac{\phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}\Big)+\sigma^2\Big(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma}\Big)\Big(\frac{\phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}-\frac{\phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}\Big)$$
$$+\sigma^2\Big(\frac{\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}+\frac{1-\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{(1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma}))}\Big)+(\mu+\sigma^2 t)^2\Big(\frac{\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}+\frac{1-\Phi(\frac{\bar{\alpha}-(\mu+\sigma^2 t)}{\sigma})}{(1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma}))}\Big)\Big)$$

evaluating the first derivative at $t=0$

$$\tilde{\mu}=\mu+\frac{\sigma\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{2}\Big(\frac{1}{1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}-\frac{1}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}\Big)$$

now to get the second moment we evaluate the second derivative at $t=0$ which gives us:

$$\mu^2+\sigma^2+\mu\sigma\Big(\frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}-\frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}\Big)+\frac{\sigma^2}{2}\Big(\frac{\bar{\alpha}-\mu}{\sigma}\Big)\Big(\frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}-\frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}\Big)$$

Substracting the square of the first moment $\tilde{\mu}^2$ I obtain

$$\tilde{\sigma}^2 = \sigma^2 + \frac{\sigma^2}{2}\Big(\frac{\bar{\alpha}-\mu}{\sigma}\Big)\Big(\frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma})} - \frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}\Big) - \frac{\sigma^2}{4}\Big(\frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma})} - \frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}\Big)^2$$

and after some algebra

$$\tilde{\sigma}^2 = \sigma^2\Big(1 + \frac{1}{2}\Big(\frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma})} - \frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}\Big)\Big(\Big(\frac{\bar{\alpha}-\mu}{\sigma}\Big) - \frac{1}{2}\Big(\frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{1-\Phi(\frac{\bar{\alpha}-\mu}{\sigma})} - \frac{\phi(\frac{\bar{\alpha}-\mu}{\sigma})}{\Phi(\frac{\bar{\alpha}-\mu}{\sigma})}\Big)\Big)\Big)$$

□

Proof. of Proposition (9): After receiving the quantitative signal r_1 investors posterior at time 1 is normally distributed with mean $(\alpha_0 + \frac{\omega}{\gamma+\omega} r_1)$ and standard deviation $\frac{1}{\sqrt{\gamma+\omega}}$ (See. Berk and Green (2004) Proposition (1)). Replacing this as μ and σ in equation 3.9 gives the desired result. □

Proof. of Proposition (10): The sign of $\frac{\partial F}{\partial \tilde{\alpha}} = \frac{1}{a q_0}\frac{\partial \alpha_1}{\partial \tilde{\alpha}}$ is equal to the sign of $\frac{\partial \alpha_1}{\partial \tilde{\alpha}}$, defining $\Phi = \Phi(J(\tilde{\alpha}, r_1, \alpha_0)), \phi = \phi(J(\tilde{\alpha}, r_1, \alpha_0))$ and $z = J(\tilde{\alpha}, r_1, \alpha_0)$

$$\frac{\partial \alpha_1}{\partial \tilde{\alpha}} = \frac{1}{2} \times \Big(\frac{\big(-z\Phi\big(-2\Phi^2+3\Phi-1\big)\phi + \big(2\Phi^2-2\Phi+1\big)\phi^2\big)}{(\Phi-1)^2\Phi^2}\Big)$$

since $\phi > 0$ the sign of the derivative depends on the sign of the following expression:

$$f(z) = \Phi z\big(2\Phi^2 - 3\Phi + 1\big) + \big(2\Phi^2 - 2\Phi + 1\big)\phi$$

which can be factorized as:

$$f(z) = (z\Phi + \phi)(2\Phi^2 - 2\Phi + 1) - z\Phi^2$$

In order to prove that the expression is strictly positive I will prove that the function is (i) continuous,(ii) it tends to zero in the limits $\pm\infty$, (iii) there $\exists z \in R : f(z) > 0$ and (iv) the function has no real roots. This implies that the function never crosses the x-axis which means it is strictly positive everywhere.
(i) f is continuous since it is a composition of continuous functions ϕ, Φ and z.
(ii) Its behaviour approaching $\pm\infty$ is:

$$\begin{aligned}\lim_{z\to\pm\infty} f(z) &= \lim_{z\to\pm\infty} 2z\Phi^3 - 2z\Phi^2 + z\Phi + 2\phi\Phi^2 - 2\phi\Phi + \phi - z\Phi^2 \\ &= \lim_{z\to\pm\infty} z\Phi(1-\Phi)(1-2\Phi) + 0\end{aligned}$$

since the limit of the term $2\phi\Phi^2 - 2\phi\Phi + \phi$ is 0 when $z \to \pm\infty$.
I use the fact that if $z \geqslant 1 \to (1-\Phi(z)) \leqslant \phi(z)$ and that $\lim_{z\to\infty} z^k\phi(z) = 0$ for $k > 0$, we know that $\lim_{z\to\infty} z(1-\Phi(z)) = 0$ and since $\Phi(z) = 1-\Phi(-z)$ we know that $\lim_{z\to-\infty} z\Phi = 0$, plugging this limits into the limit of $z\Phi(1-\Phi)(1-2\Phi)$ shows that the limit of $f(z)$ is zero in $\pm\infty$.
(iii, iv) I can show that the function is strictly positive for all $z < 0$ and since the function is even there are no real roots. I start by looking at the sign of the following functions: $z\Phi + \phi \geqslant 0$ from Proposition (14), $2\Phi^2 - 2\Phi + 1 > 0$ since it has no real roots and its convex, and $-z\Phi^2 > 0$ when $z < 0$. Finally we just need to prove that the function is even, we can use the fact that ϕ is even and that $\Phi(-z) = 1 - \Phi(z)$.

$$
\begin{aligned}
f(-z) &= (-z(1-\Phi)+\phi)(2(1-\Phi)^2 - 2(1-\Phi)+1) + z(1-\Phi)^2 \\
&= (z\Phi + \phi - z)(2(1-2\Phi+\Phi^2) + 2\Phi - 2 + 1) + z(1-2\Phi+\Phi^2) \\
&= (z\Phi + \phi - z)(2\Phi^2 - 4\Phi + 2 + 2\Phi - 2 + 1) + z(1-2\Phi+\Phi^2) \\
&= (z\Phi + \phi - z)(2\Phi^2 - 2\Phi + 1) + z(1-2\Phi+\Phi^2+\Phi^2-\Phi^2) \\
&= (z\Phi + \phi)(2\Phi^2 - 2\Phi + 1) - z\Phi^2 \\
&= f(z)
\end{aligned}
$$

Which means that there are no real roots, and the function is strictly positive also for $z > 0$, which concludes the proof. □

Proof. of Proposition 11. Defining:

$$
\begin{aligned}
\mu &= W_1 + X_1\alpha_1 \\
\sigma^2 &= X_1^2(\sigma_1^2 + \sigma_\epsilon^2) \\
\bar{\alpha} &= W_1 + X_1\tilde{\alpha} \\
\sigma_\alpha^2 &= (\sigma_1^2 + \sigma_\epsilon^2)
\end{aligned}
$$

And replacing $t = -\gamma$ we have that the expected utility of the investor is equal to:

$$
\frac{\exp\{-(W_1 + X_1\alpha_1)\gamma + \frac{X_1^2\sigma_\alpha^2\gamma^2}{2}\}}{2}\left(\frac{\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha} + X_1\sigma_\alpha\gamma)}{\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha})} + \frac{1-\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha} + X_1\sigma_\alpha\gamma)}{(1-\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha}))}\right)
$$

Taking the derivative with respect to X_1 gives us:

$$\left(\frac{\gamma\sigma_\alpha\phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha}+X_1^*\gamma\sigma_\alpha)}{\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha})}-\frac{\gamma\sigma_\alpha\phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha}+X_1^*\gamma\sigma_\alpha)}{1-\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha})}\right)$$
$$+(X_1^*\sigma_\alpha^2\gamma^2-\alpha_1\gamma)\left(\frac{1-\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha}+X_1^*\sigma_\alpha\gamma)}{1-\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha})}+\frac{\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha}+X_1^*\sigma_\alpha\gamma)}{\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha})}\right)=0$$
$$\phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha}+X_1^*\gamma\sigma_\alpha)\left(\frac{1}{\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha})}-\frac{1}{1-\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha})}\right)$$
$$+(X_1^*\sigma_\alpha\gamma-\frac{\alpha_1}{\sigma_\alpha})\left(\frac{1-\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha}+X_1^*\sigma_\alpha\gamma)}{1-\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha})}+\frac{\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha}+X_1^*\sigma_\alpha\gamma)}{\Phi(\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha})}\right)=0$$

Defining $\frac{\tilde{\alpha}-\alpha_1}{\sigma_\alpha}=z_\alpha$, $\Phi^\alpha=\frac{1}{\Phi(z_\alpha)}-\frac{1}{1-\Phi(z_\alpha)}$, $\Phi_1^\alpha=\frac{1}{\Phi(z_\alpha)}$, and $\Phi_2^\alpha=\frac{1}{1-\Phi(z_\alpha)}$ the first order condition becomes:

$$\phi(z_\alpha+X_1\gamma\sigma_\alpha)\Phi^\alpha+(X_1\sigma_\alpha\gamma-\frac{\alpha_1}{\sigma_\alpha})(\Phi_2^\alpha(1-\Phi(z_\alpha+X\sigma_\alpha\gamma))+\Phi_1^\alpha(\Phi(z_\alpha+X\sigma_\alpha\gamma)))=0$$

or

$$\phi(z_\alpha+X_1\gamma\sigma_\alpha)\Phi^\alpha+(X_1\sigma_\alpha\gamma-\frac{\alpha_1}{\sigma_\alpha})(\Phi_2^\alpha+\Phi^\alpha\Phi(z_\alpha+X\sigma_\alpha\gamma))=0$$

re arranging terms the optimal holding is the solution to the following system of equations

$$\phi(y)\Phi^\alpha+(y-\frac{\tilde{\alpha}}{\sigma_\alpha})(\Phi_2^\alpha+\Phi^\alpha\Phi(y))=0$$
$$y=z_\alpha+X_1\sigma_\alpha\gamma$$

□

Proof. of Proposition 12: The Nash equilibrium is the solution to the equations:

$$qF(r_1,\tilde{\alpha}+\Delta)+(1-q)F(r,\tilde{\alpha})=F(r,\tilde{\alpha})$$

and,

$$pEU(W_2(X(r_1,\tilde{\alpha}+\Delta)),r_1,\tilde{\alpha})+(1-p)EU(W_2(X(r_1,\tilde{\alpha})),r_1,\tilde{\alpha})=EU(W_2(X(r_1,\tilde{\alpha}))-C,r_1,\tilde{\alpha})$$

solving for p and q gives the desired result.

□

Proof. of Proposition 13:

- $\frac{\partial q}{\partial K} = \frac{F(r_1,\tilde{\alpha}+\Delta)-F(r,\tilde{\alpha})}{(F(r_1,\tilde{\alpha}+\Delta)-F(r,\tilde{\alpha})+K)^2} > 0$

- The sign of $\frac{\partial p}{\partial C}$ can be inferred by observing that the larger C is, the difference

$$EU(W_2(X(r_1,\tilde{\alpha})), r_1, \tilde{\alpha}) - EU(W_2(X(r_1,\tilde{\alpha})) - C, r_1, \tilde{\alpha})$$

  becomes larger and therefore the derivative is positive.

- $\frac{\partial q}{\partial \tilde{\alpha}} = \frac{K(\frac{\partial F(r_1,\tilde{\alpha})}{\partial \tilde{\alpha}} - \frac{\partial F(r_1,\tilde{\alpha}+\Delta)}{\partial \tilde{\alpha}})}{(F(r_1,\tilde{\alpha}+\Delta)-F(r,\tilde{\alpha})+K)^2} < 0$ if F is convex on $\tilde{\alpha}$ (e.g. $\frac{\partial F(r_1,\tilde{\alpha})}{\partial \tilde{\alpha}} - \frac{\partial F(r_1,\tilde{\alpha}+\Delta)}{\partial \tilde{\alpha}} < 0$) and 0 if F is linear on $\tilde{\alpha}$.

$\square$

Bibliography

Abel, A. and Eberly, J. (1994). A unified model of investment under uncertainty. *American Economic Review*, 84(5):1369–84.

Abramowitz, M. J. (2017). Stop the manipulation of democracy online. *New York Times*, December 11.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *Journal of Finance*, 59(3):1259–1294.

Aral, S. and Eckles, D. (2019). Protecting elections from social media manipulation. *Science*, 365(6456):858–861.

Bachmann, R. and Bayer, C. (2014). Investment dispersion and the business cycle. *American Economic Review*, 104(4):1392–1416.

Bai, Y., Kehoe, P., and Arellano, C. (2011). Financial Markets and Fluctuations in Uncertainty. Technical report.

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

Bali, T. G., Del Viva, L., El Hefnawy, M., and Trigeorgis, L. (2019). The value uncertainty premium. Available at SSRN: https://ssrn.com/abstract=3299582.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59(4):1481–1509.

Barber, B., Huang, X., and Odean, T. (2016). Which factors matter to investors? Evidence from mutual fund flows. *Review of Financial Studies*, 29(10):2600–2642.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343.

Barnett, S. A. and Sakellaris, P. (1998). Nonlinear response of firm investment to q: Testing a model of convex and non-convex adjustment costs. *Journal of Monetary Economics*, 42(2):261–288.

Barradale, M. J. (2010). Impact of public policy uncertainty on renewable energy investment: Wind power and the production tax credit. *Energy Policy*, 38(12):7698–7709.

Basu, S. and Bundick, B. (2012). Uncertainty shocks in a model of effective demand. NBER Working Papers 18420, National Bureau of Economic Research, Inc.

Berk, J. B. and Green, R. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112(6):1269–1295.

Berk, J. B. and van Binsbergen, J. (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics*, 119(1):1–23.

Bertomeu, J. and Marinovic, I. (2016). A theory of hard and soft information. *The Accounting Review*, 91(1):1–20.

Bialkowski, J., Gottschalk, K., and Wisniewski, T. (2008). Stock market volatility around national elections. *Journal of Banking & Finance*, 32(9):1941–1953.

Bianchi, D., Büchner, M., and Tamoni, A. (2020). Bond Risk Premiums with Machine Learning. *The Review of Financial Studies*, 34(2):1046–1089.

Blankespoor, E., Miller, G. S., and White, H. D. (2014). The role of dissemination in market liquidity: Evidence from firms' use of Twitter$^{\text{TM}}$. *The Accounting Review*, 89(1):79–112.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3):623–685.

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I., and Terry, S. J. (2018). Really uncertain business cycles. *Econometrica*, 86(3):1031–1065.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

Bolton, P., Chen, H., and Wang, N. (2011). A unified theory of tobin's q, corporate investment, financing, and risk management. *Journal of Finance*, 66(5):1545–1578.

Bolton, P. and Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*. forthcoming.

Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics*, 7(3):265–296.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–51.

Brennan, M. J., Wang, A. W., and Xia, Y. (2004). Estimation and test of a simple model of intertemporal capital asset pricing. *The Journal of Finance*, 59(4):1743–1776.

Brøgger, A. and Kronies, A. (2020). Skills and sentiment in sustainable investing. Working Paper.

Brinkerink, J., Chegut, A., and Letterie, W. (2019). Energy performance and capital expenditures in manufacturing industries. *Energy Efficiency*, 12:2011–2038.

Brogaard, J. and Detzel, A. (2015). The asset-pricing implications of government economic policy uncertainty. *Management Science*, 61(1):3–18.

Bulusu, N. and Gómez Biscarri, J. (2012). Frequency of consumption adjustment and the equity premium. Working Paper.

Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, 63(6):2899–2939.

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1):27 – 60.

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

Chassot, S., Hampl, N., and Wüstenhagen, R. (2014). When energy policy meets free-market capitalists: The moderating influence of worldviews on risk perception and renewable energy investment decisions. *Energy Research & Social Science*, 3:143–151.

Chen, J., Hong, H., Huang, M., and Kubik, J. D. (2004). Does fund size erode mutual fund performance? the role of liquidity and organization. *American Economic Review*, 94(5):1276–1302.

Chevalier, J. and Ellison, G. (1997). Risk taking by mutual funds as a response to incentives. *Journal of Political Economy*, 105(6):1167–1200.

Chiang, I.-H. E., Hughen, W. K., and Sagi, J. S. (2015). Estimating oil risk factors using information from equity and derivatives markets. *The Journal of Finance*, 70(2):769–804.

Christiano, L., Motto, R., and Rostagno, M. (2010). Financial factors in economic fluctuations. Working Paper Series 1192, European Central Bank.

Christiano, L. J., Motto, R., and Rostagno, M. (2014). Risk shocks. *American Economic Review*, 104(1):27–65.

Christoffersen, S. E., Musto, D. K., and Wermers, R. (2014). Investor Flows to Asset Managers: Causes and Consequences. *Annual Review of Financial Economics*, 6(1):289–310.

Christoffersen, S. E. K., Evans, R., and Musto, D. K. (2013). What Do Consumers' Fund Flows Maximize? Evidence from Their Brokers' Incentives. *Journal of Finance*, 68(1):201–235.

Cochrane, J. (2009). *Asset Pricing: Revised Edition*. Princeton University Press.

Cochrane, J. H. (1991). Production-based asset pricing and the link between stock returns and economic fluctuations. *The Journal of Finance*, 46(1):209–237.

Cochrane, J. H. (2008). The Dog That Did Not Bark: A Defense of Return Predictability. *Review of Financial Studies*, 21(4):1533–1575.

Cochrane, J. H. and Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, 95(1):138–160.

Colak, G., Durnev, A., and Qian, Y. (2017). Political uncertainty and ipo activity: Evidence from u.s. gubernatorial elections. *Journal of Financial and Quantitative Analysis*, 52(06):2523–2564.

Cooper, R. and Ejarque, J. (2003). Financial frictions and investment: Requiem in q. *Review of Economic Dynamics*, 6(4):710–728.

Cornelli, F., Kominek, Z., and Ljungqvist, A. (2013). Monitoring managers: Does it matter? *Journal of Finance*, 68(2):431–481.

Cronqvist, H. (2006). Advertising and portfolio choice. *Working Paper*.

Dacorogna, B. (1992). *Introduction to the Calculus of Variations*.

DeMiguel, V., Gil-Bazo, J., Nogales, F. J., and A. P. Santos, A. (2021). Can machine learning help to select portfolios of mutual funds?

Demir, E., Gozgor, G., Lau, C. K. M., and Vigne, S. A. (2018). Does economic policy uncertainty predict the bitcoin returns? an empirical investigation. *Finance Research Letters*, 26:145 – 149.

Dickey, D. and Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root. *JASA. Journal of the American Statistical Association*, 74.

Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, 7(2):197 – 226.

Dou, W., Ji, Y., and Wang, P. (2020a). Granular uncertainty and disastrous misallocation. Working Paper.

Dou, W., Kogan, L., and Wu, W. (2020b). Common fund flows: Flow hedging and factor pricing. Working Paper.

Dou, W. W. (2017). Embrace or Fear Uncertainty: Growth Options, Limited Risk Sharing, and Asset Prices. Technical report, University of Pennsylvania.

Döpke, J. and Pierdzioch, C. (2006). Politics and the stock market: Evidence from germany. *European Journal of Political Economy*, 22(4):925–943.

Duffie, D. and Epstein, L. (1992). Stochastic differential utility. *Econometrica*, 60(2):353–94.

Dumitrescu, A. and Gil-Bazo, J. (2016). Information and investment under uncertainty. *Economics Letters*, 148(C):17–22.

Dumitrescu, A. and Gil-Bazo, J. (2017). Market Frictions, Investor Sophistication, and Persistence in Mutual Fund Performance. *Journal of Financial Markets forthcoming*.

Dye, R. A. and Sridhar, S. (2004). Reliability-relevance trade-offs and the efficiency of aggregation. *Journal of Accounting Research*, 42(1):51–88.

Edelstein, P. and Kilian, L. (2009). How sensitive are consumer expenditures to retail energy prices? *Journal of Monetary Economics*, 56(6):766–779.

Epstein, L. G. and Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969.

Evans, R. B. and Sun, Y. (2020). Models or stars: The role of asset pricing models and heuristics in investor risk adjustment. *The Review of Financial Studies*.

Fama, E. (1991). Efficient capital markets: Ii. *Journal of Finance*, 46(5):1575–617.

Fama, E. and French, K. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–65.

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

Fama, E. and French, K. (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance*, 50(1):131–55.

Fama, E. and French, K. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, 51(1):55–84.

Fama, E. F. and French, K. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–25.

Fama, E. F. and French, K. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1):23–49.

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

Füss, R. and Bechtel, M. M. (2008). Partisan politics and stock market performance: The effect of expected government partisanship on stock returns in the 2002 german federal election. *Public Choice*, 135(3):131–150.

Gala, V. D., Gomes, J. F., and Liu, T. (2019). Investment without q. *Journal of Monetary Economics*.

Gallaher, S. T., Kaniel, R., and Starks, L. T. (2015). Advertising and mutual funds: From families to individual funds. *Working Paper*.

Gennaioli, N. and Shleifer, A. (2010). What comes to mind. *The Quarterly Journal of Economics*, 125(4):1399–1433.

Gholampour, V. and van Wincoop, E. (2017). What can we learn from euro-dollar tweets? Working Paper 23293, National Bureau of Economic Research.

Gil-Bazo, J. and Ruiz-Verdú, P. (2009). The relation between price and performance in the mutual fund industry. *Journal of Finance*, 64(5):2153–2183.

Gilje, E., Ready, R., and Roussanov, N. (2016). Fracking, Drilling, and Asset Pricing: Estimating the Economic Benefits of the Shale Revolution. NBER Working Papers 22914, National Bureau of Economic Research, Inc.

Gomes, J. F. (2001). Financing investment. *American Economic Review*, 91(5):1263–1285.

Goodell, J. and Vahamaa, S. (2013). Us presidential elections and implied volatility: The role of political uncertainty. *Journal of Banking & Finance*, 37(3):1108–1117.

Grotteria, M. (2019). Follow the money. Technical report, London Business School.

Gu, C. and Kurov, A. (2020). Informational role of social media: Evidence from twitter sentiment. *Journal of Banking and Finance*, 121:105969.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33(5):2223–2273.

Gulen, H. and Ion, M. (2015). Policy Uncertainty and Corporate Investment. *The Review of Financial Studies*, 29(3):523–564.

Gustafson, A., Goldberg, M. H., Kotcher, J. E., Rosenthal, S. A., Maibach, E. W., Ballew, M. T., and Leiserowitz, A. (2020). Republicans and democrats differ in why they support renewable energy. *Energy Policy*, 141:111448.

Ha, Y. and Ko, K. (2019). Misspecifications in the fund flow-performance relationship. *Journal of Financial Intermediation*, 38:69–81.

Haller, S. A. and Hyland, M. (2014). Capital–energy substitution: Evidence from a panel of irish manufacturing firms. *Energy Economics*, 45:501 – 510.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–54.

Hansen, L. and Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy*, 99(2):225–62.

Harvey, C. (1988). The real term structure and consumption growth. *Journal of Financial Economics*, 22(2):305–333.

Harvey, C., Liu, Y., and Zhu, H. (2016). Editor's choice ... and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68.

Harvey, C. R., Liu, Y., and Saretto, A. (2019). An evaluation of alternative multiple testing methods for finance applications. Available at ssrn: https://ssrn.com/abstract=3480087 or http://dx.doi.org/10.2139/ssrn.3480087.

Hassett, K. A. and Metcalf, G. E. (1993). Energy conservation investment: Do consumers discount the future correctly? *Energy Policy*, 21(6):710–716.

Hayashi, F. (1982). Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica*, 50(1):213–224.

Hillert, A., Niessen-Ruenzi, A., and Ruenzi, S. (2016). Mutual fund shareholder letters: Flows, performance, and managerial behavior. Technical report.

Hong, H., Li, F. W., and Xu, J. (2019). Climate risks and market efficiency. *Journal of Econometrics*, 208(1):265–281.

Hortaçsu, A. and Syverson, C. (2004). Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *Quarterly Journal of Economics*, 119(2):403–456.

Hou, K., Mo, H., Xue, C., and Zhang, L. (2020). An Augmented q-factor Model with Expected Growth. *Review of Finance*. rfaa004.

Hou, K., Xue, C., and Zhang, L. (2014). Digesting Anomalies: An Investment Approach. *The Review of Financial Studies*, 28(3):650–705.

Huang, J., Wei, K. D., and Yan, H. (2007). Participation costs and the sensitivity of fund flows to past performance. *The Journal of Finance*, 62(3):1273–1311.

Huang, J. C., Wei, K. D., and Yan, H. (2012). Investor learning and mutual fund flows. Technical report.

Ippolito, R. A. (1992). Consumer reaction to measures of poor quality: Evidence from the mutual fund industry. *Journal of Law and Economics*, 35(1):45–70.

Jain, P. C. and Wu, J. S. (2000). Truth in mutual fund advertising: Evidence on future performance and fund flows. *Journal of Finance*, 55(2):937–958.

Jin, Y. and Jorion, P. (2006). Firm value and hedging: Evidence from u.s. oil and gas producers. *The Journal of Finance*, 61(2):893–919.

Jones, C. M. and Kaul, G. (1996). Oil and the stock markets. *Journal of Finance*, 51(2):463–91.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3):1177–1216.

Kaldor, N. (1966). Marginal productivity and the macro-economic theories of distribution: Comment on samuelson and modigliani. *Review of Economic Studies*, 33(4):309–319.

Karnizova, L. and Li, J. C. (2014). Economic policy uncertainty, financial markets and probability of us recessions. *Economics Letters*, 125(2):261 – 265.

Kelly, B., Pastor, L., and Veronesi, P. (2016). The price of political uncertainty: Theory and evidence from the option market. *Journal of Finance*, 71(5):2417–2480.

Khorana, A. (2001). Performance changes following top management turnover: Evidence from open-end mutual funds. *Journal of Financial and Quantitative Analysis*, 36(3):371–393.

Kim, O. and Verrecchia, R. (1991). Trading volume and price reactions to public announcements. *Journal of Accounting Research*, 29(2):302–321.

Klößner, S. and Sekkel, R. (2014). International spillovers of policy uncertainty. *Economics Letters*, 124(3):508 – 512.

Kogan, L. and Papanikolaou, D. (2012). Economic activity of firms and asset prices. *Annual Review of Financial Economics*, 4(1):361–384.

Kogan, L. and Papanikolaou, D. (2013). Firm Characteristics and Stock Returns: The Role of Investment-Specific Shocks. *The Review of Financial Studies*, 26(11):2718–2759.

Kogan, L. and Papanikolaou, D. (2014). Growth opportunities, technology shocks, and asset prices. *The Journal of Finance*, 69(2):675–718.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2–3):221 – 247. Economic Consequences of Alternative Accounting Standards and Regulation.

Li, J. and Born, J. A. (2006). Presidential election uncertainty and common stock returns in the united states. *Journal of Financial Research*, 29(4):609–622.

Li, Z. and Zhong, J. (2020). Impact of economic policy uncertainty shocks on china's financial conditions. *Finance Research Letters*, 35:101303.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *Journal of Finance*, 20(4):587–615.

Liu, L. and Zhang, T. (2015). Economic policy uncertainty and stock market volatility. *Finance Research Letters*, 15:99 – 105.

Liu, L. X., Whited, T. M., and Zhang, L. (2009). Investment-Based Expected Stock Returns. *Journal of Political Economy*, 117(6):1105–1139.

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

Loughran, T. and McDonald, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance*, 69(4):1643–1671.

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

Lucas, R. (1978). Asset prices in an exchange economy. *Econometrica*, 46(6):1429–45.

Ludvigson, S. and Ng, S. (2007). The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics*, 83(1):171–222.

Ludvigson, S. C. and Ng, S. (2009). A Factor Analysis of Bond Risk Premia. NBER Working Papers 15188, National Bureau of Economic Research, Inc.

Lynch, A. W. and Musto, D. K. (2003). How investors interpret past fund returns. *Journal of Finance*, 58(5):2033–2058.

Maio, P. and Santa-Clara, P. (2012). Multifactor models and their consistency with the icapm. *Journal of Financial Economics*, 106(3):586–613.

Margolis, R. M. and Kammen, D. M. (1999). Evidence of under-investment in energy r&d in the united states and the impact of federal policy. *Energy Policy*, 27(10):575–584.

Mattozzi, A. (2008). Can we insure against political uncertainty? evidence from the u.s. stock market. *Public Choice*, 137(1):43–55.

McCracken, M. W. and Ng, S. (2015). FRED-MD: A Monthly Database for Macroeconomic Research. Working Papers 2015-12, Federal Reserve Bank of St. Louis.

Merton, R. (1973). An intertemporal capital asset pricing model. *Econometrica*, 41(5):867–87.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, 34(4):768–783.

Mullainathan, S. (2002). A memory-based model of bounded rationality. *The Quarterly Journal of Economics*, 117(3):735–774.

Mullainathan, S., Schwartzstein, J., and Shleifer, A. (2008). Coarse Thinking and Persuasion. *The Quarterly Journal of Economics*, 123(2):577–619.

Nanda, V., Wang, Z. J., and Zheng, L. (2004). Family values and the star phenomenon: Strategies of mutual fund families. *Review of Financial Studies*, 17(3):667–698.

Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–08.

Palmer, M. A. (2002). Kenneth R. Mayer. With the Stroke of the Pen: Executive Orders and Presidential Power. Princeton: Princeton University Press. 2001. *The American Historical Review*, 107(5):1553–1554.

Papanikolaou, D. (2011). Investment shocks and asset prices. *Journal of Political Economy*, 119(4):639–685.

Pastor, L. and Stambaugh, R. (2012). On the size of the active management industry. *Journal of Political Economy*, 120(4):740 – 781.

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2015). Scale and skill in active management. *Journal of Financial Economics*, 116(1):23–45.

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2019). Sustainable Investing in Equilibrium. NBER Working Papers 26549, National Bureau of Economic Research, Inc.

Pástor, L. and Veronesi, P. (2006). Was there a nasdaq bubble in the late 1990s? *Journal of Financial Economics*, 81(1):61 – 100.

Pástor, L. and Veronesi, P. (2009). Technological revolutions and stock prices. *The American Economic Review*, 99(4):1451–1483.

Pástor, L. and Veronesi, P. (2012). Uncertainty about government policy and stock prices. *Journal of Finance*, 67(4):1219–1264.

Pástor, L. and Veronesi, P. (2013). Political uncertainty and risk premia. *Journal of Financial Economics*, 110(3):520 – 545.

Pástor, L. and Veronesi, P. (2017). Political Cycles and Stock Returns. CEPR Discussion Papers 11864, C.E.P.R. Discussion Papers.

Pástor, L. and Veronesi, P. (2018). Inequality Aversion, Populism, and the Backlash Against Globalization. NBER Working Papers 24900, National Bureau of Economic Research, Inc.

Popp, D. (2002). Induced innovation and energy prices. *American Economic Review*, 92(1):160–180.

Rabin, M. (2002). Inference by believers in the law of small numbers. *The Quarterly Journal of Economics*, 117(3):775–816.

Rabin, M. and Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics*, 114(1):37–82.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., and Mozetič, I. (2015). The effects of twitter sentiment on stock price returns. *PloS one*, 10(9):e0138441.

Ready, R. C. (2017). Oil Prices and the Stock Market*. *Review of Finance*, 22(1):155–176.

Reuter, W. H., Szolgayová, J., Fuss, S., and Obersteiner, M. (2012). Renewable energy investment: Policy and market impacts. *Applied Energy*, 97:249–254.

Riem, M. (2016). Corporate investment decisions under political uncertainty. ifo Working Paper Series 221, ifo Institute - Leibniz Institute for Economic Research at the University of Munich.

Rubinstein, M. (1976). The valuation of uncertain income streams and the pricing of options. *Bell Journal of Economics*, 7(2):407–425.

Santa-Clara, P. and Valkanov, R. (2003). The presidential puzzle: Political cycles and the stock market. *The Journal of Finance*, 58(5):1841–1872.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

Sirri, E. R. and Tufano, P. (1998). Costly search and mutual fund flows. *Journal of Finance*, 53(5):1589–1622.

Snowberg, E., Wolfers, J., and Zitzewitz, E. (2007). Partisan impacts on the economy: evidence from prediction markets and close elections. *The Quarterly Journal of Economics*, 122(2):807–829.

Stambaugh, R. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.

Stambaugh, R. F. and Yuan, Y. (2016). Mispricing Factors. *The Review of Financial Studies*, 30(4):1270–1315.

Stewart, M. B. (1978). Factor-price uncertainty with variable proportions. *The American Economic Review*, 68(3):468–473.

Tobin, J. (1969). A general equilibrium approach to monetary theory. *Journal of Money, Credit and Banking*, 1(1):15–29.

Tovar, M. A. and Iglesias, E. M. (2013). Capital-Energy Relationships: An Analysis when Disaggregating by Industry and Different Types of Capital. *The Energy Journal*, 0(Number 4).

von Neumann, J. and Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton University Press.

Yogo, M. (2006). A consumption-based explanation of expected stock returns. *The Journal of Finance*, 61(2):539–580.

You, W., Guo, Y., and Peng, C. (2017). Twitter's daily happiness sentiment and the predictability of stock returns. *Finance Research Letters*, 23:58–64.

Zhang, L. (2005). The value premium. *The Journal of Finance*, 60(1):67–103.

Zhang, L. (2017). The investment capm. *European Financial Management*, 23(4):545–603.

Zhao, C. (2020). Pseudo-bayesian updating. Working paper.

Zhu, M. (2018). Informative fund size, managerial skill, and investor rationality. *Journal of Financial Economics*, 130(1):114–134.