Investor Attention and Stock Market Outcomes

by

Yao Chen

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Warwick Business School

February 2017

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my PhD supervisors Dr. Chendi Zhang and Prof. Alok Kumar for their immense knowledge, encouragement, patience, and insightful comments. Without their continuous support during the past four years, it would not be possible for me to complete this thesis.

I thank Prof. Gulnur Muradoglu and Dr. Constantinos Antonio for their constructive comments and suggestions during my viva. Their suggestions improve this thesis greatly.

The members of the Finance group have contributed greatly to my personal and professional time at Warwick Business School. I am especially grateful to Prof. Michael Moore for providing me conference funding, and to Dr. Constantinos Antonio and Dr. Jana Fidrmuc for their constructive comments during upgrading and completion reviews which encouraged me to improve my research from various perspectives. I would also like to thank Dr. Arie Gozluklu, Dr. Daniele Bianchi, and Dr. Qian Wang for sharing job market experience with me, and thank our group secretary Nicki Pegg for making my life much easier during my hard working days.

My sincere thanks also go to the University of Warwick for recruiting me as a doctoral researcher, and in particular, for giving me the opportunity to meet my wife. I am also grateful to the British Economic and Social Research Council for funding my research throughout the past four years.

Chapter 2 of this thesis was presented at the following institutions and conferences: the University of Warwick, University of Miami, University of Exeter, University of Surrey, the European Financial Management Association Annual Meeting (Basel), China International Conference in Finance (Xiamen), and Financial Management Association Annual Meeting (Las Vegas). I thank the participants of these conferences and seminars for valuable comments and suggestions.

I thank my PhD colleagues Chengyu Yang, Harold Contreras Munoz, Zhun Liu, Ali Osseiran, Juan Carpio for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last four years. I also thank my senior PhD fellows Jeff Hung, Dr. Chunling Xia, Dr. Qi Xu, Dr. Zicheng Lei, Dr. Ilias Filippou, and Dr. Lu Li for their help and support.

Last but not least, I would like to thank my family. Words cannot express how grateful I am to my mother-in-law Ruolan Lin, father-in-law Zilong Chen, my mother Fengqin Hu, and my father Xiaoqing Chen for your unconditional support throughout my PhD. In the end, I would like to thank my beloved wife, Dr. Linquan Chen, for your continuous encouragement and for all of the sacrifices that you have made on my behalf.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy in Finance. It has been composed by myself and has not been submitted in any previous application for any degree. Chapters 2 to 4 are co-authored with Dr. Chendi Zhang and Prof. Alok Kumar. In all the three chapters, both the empirical analysis and writing were undertaken by myself.

Yao Chen

Feburary, 2017

Abstract

The first essay (Chapter 2) shows that changes in gambling attitudes affect asset prices and corporate decisions. Using the Internet search volume for lottery-related keywords to capture gambling sentiment shifts, we show that when the overall gambling sentiment is high, investor demand for lottery-like stocks increases, stocks with lottery-like characteristics earn positive abnormal returns in the short-run, managers are more likely to announce stock splits to cater to the increased demand for low-priced lottery stocks, and IPOs perceived as lotteries earn higher first-day returns. Further, the sentiment-return relation is stronger among low institutional-ownership firms and in regions where gambling is more acceptable.

The second essay (Chapter 3) examines the relation between social attributes and stock returns. As investors regularly update their beliefs on firm-level CSR records, they are likely to rebalance their portfolios to include firms with good social attributes. Using a novel measure to identify perceived social attributes, we demonstrate that stocks with good perceived social attributes have better future returns. A trading strategy that attempts to exploit demand-based return predictability generates an annualized risk adjusted performance of 14% and spans 15-36% of the market. Further, institutional trading results show that institutions have consistently higher demand for firms with good perceived social attributes. Our findings suggest that perceived social attributes predict stock returns.

The third essay (Chapter 4) investigates how social attributes affect mutual fund flows. Using social sensitivity estimates to capture fund-level social attributes perceived by the market, we show that mutual funds with good social attributes attract 0.13% higher monthly flows than their counterparts. In addition, these funds experience greater appreciation in flows following good performance and lower decline in flows following bad performance. When investors increase demand for corporate social responsibility, funds perceived to have poor social attributes experience 0.5% reduction in monthly fund flows. Overall, our findings are consistent with the view that mutual fund investors value social attributes when making investment decisions.

Abbreviations

ADR	American Depositary Receipt
ADS	Aruoba-Diebold-Scotti
AMEX	American Stock Exchange
ASVI	Abnormal search volume intensity
BSI	Buy-Sell Imbalance
CRSP	Center for Research on Security Prices
CSR	Corporate Social Responsibility
CUSIP	Committee on Uniform Security Identification Procedures
DEF	Default Spread
DGTW	Daniel, Grinblatt, Titman, and Wermers
DIV	Dividend Yield
EBSI	Excess Buy-Sell Imbalance
ESG	Environment Social and Corporate Governance
FACPR	Factor to Adjust Price
FACSHR	Factor to Adjust Shares Outstanding
HML	Value Factor
IPO	Initial Public Offering
KLD	Kinder Lydenberg and Domini
LIDX	Lottery Index
LIQ	Liquidity Factor
LTR	Long-Term Reversal Factor
MKTRF	Market Excess Return
MOM	Momentum Factor
MP	Monthly Growth in Industrial Production
NASDAQ	National Association of Securities Dealers Automated Quotations
NASPL	North American Association of State and Provincial Lotteries
NBER	National Bureau of Economic Research
NYSE	New York Stock Exchange
REC	Recession Indicator

REIT	Real Estate Investment Trust
RP	Monthly Default Risk Premium
S&L	Savings and Loan Association
SIC	Standard Industry Classification
SMB	Size Factor
SRI	Socially Responsible Investing
STR	Short-Term Reversal Factor
SVI	Search Volume Intensity
TERM	Term Spread
TNA	Total Net Asset
TS	Term Spread
UEI	Unexpected Inflation
UMD	Momentum Factor
UNEMP	Unemployment Rate
USD	U.S. Dollar
USSIF	U.S. Forum for Sustainable and Responsible Investment
VIX	Daily Market Volatility Index
YLD	Yield on the 90-day Treasury Bill

Chapter 1

Introduction

There is considerable interest among academics in examining the effect of investor attention on financial market outcomes. Among others, Barber and Odean (2008) shows that with limited attention, individual investors are likely to be net buyers of attention grabbing stocks. In addition, Antoniou, Kumar, and Maligkris (2016) show that attention to negative events could also affect the investment decision of more sophisticated investors such as equity analysts. Further, investor attention not only has influence on different types of investors, but also on various market outcomes such as analyst forecasts (Antoniou, Kumar, and Maligkris, 2016), earnings announcements reactions (e.g., DellaVigna and Pollet, 2009; Hirshleifer, Lim, and Teoh, 2009), and initial public offering (IPO) returns (Da, Engelberg, and Gao, 2011).

When examining the impact of attention on the stock market, early empiricists face a substantial challenge: there are no direct measures of investor attention. Therefore, they use indirect measures to proxy for investor attention. These indirect proxies include extreme returns (Barber and Odean, 2008), trading volume (Gervais, Kaniel, and Mingelgrin, 2001; Hou, Peng, and Xiong, 2009), news and headlines (Yuan, 2011), advertising expense (Grullon, Kanatas, and Weston, 2004; and Lou, 2014), and price limits (Seasholes and Wu, 2007).

More recently, Da, Engelberg, and Gao (2011) propose a novel measure to directly capture investor attention. Specifically, they use the search volume intensity (*SVI*) reported by Google Trends as a direct attention measure. As Google is the largest search engine in the world and accounts for about 70% of total search volume in the U.S. as of 2013, Google search volume is likely to represent the online search behavior of the general public.

In this thesis, I investigate the impact of investor attention on various financial market outcomes with a novel and objective measure for overall attention. Specifically, I use a new feature in Google Trends called *topic search*. According to Google, the topic search feature aggregates online search queries in different languages and different keywords as long as they are related to the topic of interest.[1]

Using the new topic search feature in Google Trends, I investigate the impact of investor attention on various financial market outcomes. Specifically, Chapter 2 examines how the time-variation in gambling attitudes affects asset prices and corporate decision. I use the *SVIs* for the topic "lottery" to capture shifts in investors' gambling attitudes. In Chapters 3 and 4, I change my perspective and investigate whether perceived social attributes affect stock returns and mutual fund flows. In these two chapters, I use the return sensitivity with respect to *SVIs* for the topic "corporate social responsibility" (CSR) to identify stocks and mutual funds that are likely to be perceived by investors as having good social attributes in the recent past. In the remaining part of this chapter, I briefly describe the motivations and key findings of the three essays.

Chapter 2 investigates whether changes in gambling attitudes affect stock returns and corporate decisions. An emerging literature in finance examines the potential link between gambling behavior and financial market outcomes. Recent theoretical studies predict that investors would be willing to accept a negative return premium for stocks with positively-skewed returns (e.g., Shefrin and Statman, 2000; Brunnermeier, Gollier, and Parker, 2007; Mitton and Vorkink, 2007; Barberis and Huang, 2008). Stocks with lottery-like payoffs are overpriced in the short-run and earn a negative average risk-adjusted return in the long-run. Empirical research on the effects of gambling attitudes has typically focused on the cross-sectional variation in gambling preferences and their impact on financial market outcomes (e.g., Bali, Cakici, and Whitelaw, 2011; Kumar, Page, and Spalt, 2011). For example, Kumar, Page, and Spalt (2011) find that investors' gambling preferences vary geographically impact stock returns as well as corporate policies.

In Chapter 2, I study how the time-variation in overall gambling attitudes affects various stock market outcomes.[2] I posit that attention to low-probability payoffs in one

---

[1] One potential limitation of Google search data is that the data are only available from 2004. Therefore, research using Google data tends to have a relatively short sample period.

[2] Google search volume mainly captures the attention of retail investors. Chapter 2 focuses on retail investors since institutional investors are more constrained when buying lottery-like stocks (e.g., Kumar, Page, and Spalt, 2011). Therefore, the Google measure is appropriate. In addition, since retail investors could invest through institutional investors like mutual funds, retail attention could also affect institutional trading. Further, Ben-Rephael, Da, and Israelsen (2017) show that Google measure is positively correlated with institutional investors' attention measure (i.e., search in Bloomberg).

setting may motivate individuals to overweight low-probability events in other related economic settings. Specifically, I conjecture that an increase in overall attitudes toward gambling (i.e., gambling sentiment) is likely to generate price pressure on stocks with lottery-like characteristics. Consequently, if arbitrage costs are high, the return of these stocks may be predictable in the short-run. In addition, corporate financial decisions could be affected as firms respond to changes in investors' gambling attitudes and their impact on asset prices.

Consistent with my conjecture, I find that when the overall gambling sentiment is high, investor demand for lottery-like stocks increases, stocks with lottery-like characteristics earn positive abnormal returns in the short-run, managers are more likely to announce stock splits to cater to the increased demand for low-priced lottery stocks, and IPOs perceived as lotteries earn higher first-day returns. Overall, these findings suggest that changes in investors' gambling attitudes have a positive spillover effect on stock market outcomes.

Chapter 3 examines whether perceived social attributes affect stock returns. Survey evidence suggests that investors are actively updating their beliefs about firm-level CSR records. Therefore, their recent perception on firm-level social attributes could affect their investment decisions. In particular, investors are likely to invest more in firms with good perceived social attributes. To investigate the effect of perceived social attributes on stock prices, I propose a novel measure to identify firms that are likely to be perceived as having good social attributes by the market. Specifically, I use social sensitivity, defined as the return sensitivity to the aggregate attention to CSR, to capture perceived social attributes. I conjecture that there will be predictable return patterns of socially sensitive firms and industries that can be identified *ex-ante.*

Consistent with the conjecture that perceived social attributes affect stock returns, I find that stocks with the most positive social sensitivity have better future returns. A trading strategy that attempts to exploit demand-based return predictability generates an annualized risk adjusted performance of 14% and spans 15-36% of the market. Further, institutional trading results show that institutions have consistently higher demand for firms with the most positive social sensitivity.

Relatedly, in Chapter 4, I investigate whether social attributes affect the investment decisions of mutual fund investors. Recent literature on mutual funds demonstrates that mutual fund investors care about fund attributes. Among others, Kumar, Niessen-Ruenzi, and Spalt (2015) document that investors are likely to avoid funds managed by individuals with foreign-sounding names even if these funds do not have inferior return performance. In addition, to pursue social objectives, mutual fund investors are willing to accept lower risk-adjusted returns (e.g., Renneboog, Ter Horst, and Zhang, 2008). Further, fund attributes could also influence the flow-return sensitivity of mutual funds (e.g., Bollen, 2007; Renneboog, Ter Horst, and Zhang, 2011).

Motivated by these studies, in Chapter 4, I investigate whether potential stereotypes associated with a fund's social attributes affect the investment decision of mutual fund investors. Since socially responsible investing (SRI) has become an important criterion to mutual fund investors, I ask whether investors are more likely to invest in funds that are perceived to have good social attributes. In addition, I also examine whether perceived social attributes influence flow-return sensitivity. My key conjecture is that funds perceived to have good social attributes attract higher flows even if these funds do not have superior return performance. In addition, I posit that these funds are likely to be rewarded more following good performance and punished less following the bad performance.

Consistent with my conjecture, I find that monthly flows are around 0.1 percent higher for funds with good perceived social attributes. In addition, compared to equity funds with comparable characteristics, funds with good social attributes experience 0.19 percentage lower outflows when their recent performance is in the bottom decile of all mutual funds and 0.46 percentage higher inflows per month when their recent performance is in the top decile. Collectively, results in Chapters 3 and 4 show that investors value social attributes when making investment decisions, and their trading behavior affects stock returns and mutual fund flows.

Collectively, this thesis shows that investor attention affects investment decisions of both retail and institutional investors. Using the topic search data from Google Trends as a novel and direct measure for investor attention, I show that the time-variation in investors' gambling attitudes affect stock prices and corporate decisions such as stock splits and initial public offerings. In addition, using social sensitivity to capture the perceived social attributes of stocks and mutual funds, we show that investors demand for stocks with good perceived social attributes and are more tolerate toward mutual funds with good perceived social attributes. Overall, this these contributes to the literature by providing new evidence on the impacts of investor attention on different financial market outcomes from a time-varying perspective.

Chapter 2

Searching for Gambles: Investor Attention, Gambling Sentiment, and Stock Market Outcomes

2.1. Introduction

An emerging literature in finance examines the potential link between gambling behavior and financial market outcomes. Recent theoretical studies predict that investors would be willing to accept a negative return premium for stocks with positively-skewed returns (e.g., Shefrin and Statman, 2000; Brunnermeier, Gollier, and Parker, 2007; Mitton and Vorkink, 2007; Barberis and Huang, 2008). Stocks with lottery-like payoffs are overpriced in the short-run and earn a negative average risk-adjusted return in the long-run. Empirical research on the effects of gambling attitudes has typically focused on the cross-sectional variation in gambling preferences and their impact on financial market outcomes (e.g., Bali, Cakici, and Whitelaw, 2011; Kumar, Page, and Spalt, 2011). For example, Kumar, Page, and Spalt (2011) find that investors' gambling preferences vary geographically impact stock returns as well as corporate policies.

In this paper, we study how the time-variation in overall gambling attitudes affects various stock market outcomes. We posit that attention to low-probability payoffs in one setting may motivate individuals to overweight low-probability events in other related economic settings. Specifically, we conjecture that an increase in overall attitudes toward gambling (i.e., gambling sentiment) is likely to generate price pressure on stocks with lottery-like characteristics. Consequently, if arbitrage costs are high, the return of these stocks may be predictable in the short-run. In addition, corporate financial decisions could be affected as firms respond to changes in investors' gambling attitudes and their impact on asset prices.

To test these conjectures, we develop a novel measure of gambling sentiment of investors using Google's search volume intensity (*SVI*) for lottery-related keywords. We first examine the effects of gambling sentiment on stock returns. We focus on a segment of the U.S. stock market in which stocks have lottery-like return distributions. Following

Kumar, Page, and Spalt (2016), we define lottery-like stocks as those with low nominal share prices, high idiosyncratic skewness, and high idiosyncratic volatility. These stocks are also associated with low average returns, high return volatility and high turnover (Scheinkman and Xiong, 2003; Hong, Scheinkman, and Xiong, 2006; Grinblatt and Keloharju, 2009; Dorn and Sengmueller, 2009). We conjecture that lottery-like stocks are likely to be more affected by gambling sentiment than non-lottery stocks. Specifically, when the overall gambling sentiment is stronger, investor demand for lottery-like stocks would increase. If arbitrage costs are high,[3] this excess demand in turn could generate price pressure on lottery-like stocks and generate positive abnormal return in the short-run.

Consistent with our conjecture, we find that when gambling sentiment of investors becomes stronger, lottery-like stocks earn positive abnormal returns in the following month. In economic terms, a one standard deviation increase in investors' gambling sentiment is associated with an abnormal return of 47 basis points for the lottery-like stock portfolio in the following month. Further, this positive abnormal return is eventually arbitraged away within three months.

Next, we use attention-grabbing lottery jackpots to identify the source of time-variation in investors' gambling attitudes. Jackpot announcements are exogenous, attention-grabbing events that are likely to generate excitement among investors who may gamble in the stock market. Consistent with our expectation, we find abnormal stock returns among lottery-like stocks around these jackpots. The average abnormal return during month -1 to month +1 around the lottery jackpot events is 1.7% per month. In addition, during this period, the average abnormal trading volume is 17.2%.

To directly examine whether shifts in overall gambling attitudes have a positive spillover effect on the demand for lottery-type stocks, we use trading data from a major U.S. discount brokerage firm (Barber and Odean, 2000). Consistent with our conjecture of a positive spillover effect on investor demand, we find positive excess buy-sell imbalance on lottery-like stocks around the largest jackpot during the 1992-1996 period. The average excess buy-sell imbalance during month -1 to month +1 is 7%, which

[3] Given the low prices and high volatility of lottery-type stocks, the costs associated with arbitraging them are likely to be high.

indicates a 7% increase in net purchase of lottery-like stocks relative to non-lottery stocks. Similarly, large drawings are associated with excess buy-sell imbalance of 3% on the next trading day. This positive spillover effect is similar to the evidence in betting markets (e.g., Scott and Garen, 1994; Calcagno, Walker, and Jackson, 2010) and suggests that the demands for various gambling instruments in the U.S. are likely to be positively correlated.[4]

In the next set of tests, we examine the extent to which geographical differences in gambling sentiment influence the long-term performance of lottery-like stocks. As local investors' gambling sentiment varies across regions (Kumar, Page, and Spalt, 2011), we posit that the effects of gambling sentiment on stock returns would be stronger among U.S. states with stronger gambling sentiment. In these states, lottery-like stocks are more likely to be overpriced in the short-run and are likely to underperform in the long-run. To test our prediction, we use each firm's headquarter state to define its location and use the average state-level *SVI* to capture the gambling sentiment of local investors.

We find that in states with strong gambling sentiment, lottery-like stocks underperform non-lottery stocks (i.e., stocks with high stock price, low idiosyncratic skewness, and low idiosyncratic volatility) by 60 basis points per month. The results are stronger for stocks that are smaller or have lower institutional ownership. In contrast, in U.S. states with relatively weaker gambling sentiment, lottery-like stocks do not perform differently from non-lottery stocks.

Next, we change our perspective and investigate whether gambling sentiment affects corporate decisions. Low nominal share price is a salient feature of lottery-like stocks. Baker, Greenwood, and Wurgler (2009) show that retail investors' demand for stocks with low nominal share prices is time-varying. Further, firms cater to such demand by splitting stocks with high nominal share prices. We conjecture that the time-varying demand for low-priced stock would be related to the time-variation in investors' gambling attitudes. Consistent with this conjecture, we find that firms with high nominal share prices are more likely to split their shares when investors exhibit stronger gambling sentiment.

[4] Our findings are opposite to the evidence reported in Gao and Lin (2015). They use data from Taiwan and find a negative spillover effect. Our evidence indicates that the results from Taiwan are unlikely to generalize to the U.S. See Section 4.3 for additional details.

In the last set of tests, we examine the effects of gambling sentiment on the first-day returns of initial public offerings (IPOs). These tests are motivated by previous research, which demonstrates that IPOs are often perceived as lottery-like by retail investors (Barberis and Huang, 2008; Green and Hwang, 2011). Further, Loughran and Ritter (2004) show that the magnitude of the average first-day return for IPOs changes over time. We conjecture that IPOs would earn higher first-day returns when investors exhibit stronger gambling sentiment. Consistent with this conjecture, we find that a one standard deviation increase in investors' gambling sentiment is associated with a 1.6% increase of the average first-day IPO return in the following month.

Overall, these findings suggest that changes in investors' gambling attitudes have a positive spillover effect on stock market outcomes. In particular, when investors' gambling sentiment becomes stronger, stocks with lottery-like characteristics earn positive abnormal returns and firms with high nominal share prices are more likely to split their shares. In addition, initial public offerings earn higher first-day returns during these periods of high gambling sentiment.

Our findings contribute to at least four distinct strands of finance literature. First, we contribute to the finance literature on skewness and gambling. Recent literature shows that cross-sectional differences in gambling attitudes affect stock returns and corporate decisions (e.g., Bali, Cakici, and Whitelaw, 2011; Kumar, Page, and Spalt, 2011). For example, Bali, Cakici, and Whitelaw (2011) examine the role of extreme positive return in the cross-sectional pricing of stocks. They show that maximum daily stock return during the previous month is negatively correlated with returns in the following 11 months. This evidence suggests that investors are willing to pay more for stocks with extreme positive returns. In addition, Kumar, Page, and Spalt (2011) show that gambling propensity is stronger in U.S. states with high concentration of Catholics relative to Protestants. Investors located in regions with higher Catholic-Protestant ratios have a stronger preference for lottery-like stocks, broad-based employee stock option plans are more popular, and the initial first-day return following an initial public offering is higher. Our findings suggest that shifts in gambling attitudes over time also matter.

Second, our results provide new evidence on the possible link between the lottery market and the stock market. On the one hand, Kumar (2009) shows that state lotteries and lottery-like stocks attract similar socioeconomic clienteles. In addition, Doran, Jiang, and Peterson (2012) show that during periods with high participation rate in other forms of gambling activities (i.e., new year), investors also increase their demand for lottery-like stocks. Relatedly, the introduction of state lotteries has a positive spillover effect on participation in casino gaming and horse racing (Scott and Garen, 1994; Calcagno, Walker, and Jackson, 2010). In contrast, Dorn, Dorn, and Sengmueller (2015) and Gao and Lin (2015) find a substitution effect between lotteries and aggregate stock trading using international data. With a *direct* measure of gambling attitudes, we show a positive spillover effect to the stock market in the U.S.

Third, we provide new evidence on the economic effects of investor attention (e.g., Odean, 1999; Barber and Odean, 2008; Palomino, Renneboog, and Zhang, 2009; Da, Engelberg, and Gao, 2011 and 2015). Specifically, we show that salient lottery events trigger strong gambling sentiment, which generates return predictability among lottery-like stocks.

Fourth, we add to the catering literature in corporate finance. In particular, our results provide new insights into managerial motivation behind stock splits (e.g., Lakonishok and Lev, 1987; Angel, 1997; Baker, Greenwood and Wurgler, 2009) and provide an alternative explanation for the time-variation in first-day IPO returns (e.g., Loughran and Ritter, 2004). Specifically, we show that firms cater to the time-varying demand for lottery-like characteristics (e.g., low nominal share prices) by splitting stocks with high nominal share prices. These finding offer a potential explanation for the nominal share puzzle documented by Baker, Greenwood, and Wurgler (2009). In addition, we demonstrate that time-variation in investors' gambling sentiment is an important determinant of first-day IPO returns.

Collectively, this paper proposes a novel and direct measure of investors' time-varying gambling attitudes. Using this measure, we investigate the impact of the time-variation of gambling attitudes on various financial market outcomes, which has not yet been investigated in the existing literature due to data limitations.

2.2. Hypotheses development

We consider four different economic settings to study the impact of gambling sentiment on financial market outcomes. In the first setting, we focus on the short-term mispricing and correction pattern among lottery-like stocks. This analysis is motivated by recent studies, which find that investors are more likely to buy stocks that have recently captured their attention (Barber and Odean, 2008). Specifically, Da, Engelberg, and Gao (2011) show that a surge in attention could lead to temporal overpricing and predict short-term return reversals among the set of attention-grabbing stocks. Further, Google's daily search interest by retail investors is likely to capture market-level sentiment (Da, Engelberg, and Gao, 2015).

We extend this insight to lottery-like stocks that provide gambling opportunities to investors in the stock market. Kumar (2009) finds that state lottery players have similar behavior as investors who overweight lottery-like stocks, and the propensity to participate in state lotteries and purchase lottery-like stocks are positively correlated. When investors gamble in the stock market, they are likely to prefer stocks with low nominal share prices, especially those with positive idiosyncratic skewness for the possibility of extreme returns. Investors may also prefer stocks with high idiosyncratic volatility since extreme returns are more likely for these assets.

Kumar, Page, and Spalt (2016) construct a Lottery Index to categorize all stocks in the CRSP universe into lottery-like stocks, non-lottery stocks and other stocks. Lottery-like (non-lottery) stocks are those with low (high) price, high (low) idiosyncratic skewness and high (low) idiosyncratic volatility. Using this definition of lottery-type firms, we posit that when gambling sentiment is strong, investors are likely to invest disproportionally more in lottery-like stocks, leading to positive price pressure on these stocks. To summarize, our first hypothesis posits:

H1: Following periods of high gambling sentiment, lottery-like stocks would earn positive abnormal returns in the short-run.

Our second hypothesis focuses on the cross-sectional variation in the impact of shifts in gambling sentiment on stock returns. Barberis and Huang (2008) show that a security's idiosyncratic skewness would be priced. In particular, investors would be willing to accept lower returns for stocks with positive return skewness. Positive skewness could be a particularly important characteristic for investors with strong gambling attitudes. As investors are known to exhibit local bias (e.g., Coval and Moskowitz, 1999; Grinblatt and Keloharju, 2001; Hong, Kubik, and Stein, 2008; Seasholes and Zhu, 2010), the effects of gambling sentiment on stock returns would be stronger for stocks headquartered in states with stronger gambling attitudes. Further, we expect a larger impact on stocks that are more likely to be held by retail investors, i.e., stocks that are smaller or with lower institutional ownership. For stocks located in states with relatively weaker gambling attitudes, the negative lottery-like stock premium would be weaker or non-existent.

To summarize, our second hypothesis posits:

> *H2: The effects of gambling sentiment on stock returns would be stronger in U.S. states with stronger gambling attitudes. Further, this impact is likely to be amplified for smaller stocks and firms with lower institutional ownership.*

Our third hypothesis focuses on managerial response to changes in investors' gambling attitudes and their potential impact on asset prices. Weld, Michaely, Thaler and Benartzi (2009) show that firms keep their nominal share prices in a particular range by conducting stock splits. Baker, Greenwood, and Wurgler (2009) propose a catering theory of nominal share prices to explain this behavior. They find that the demand for low-priced stocks is time-varying and firms with high nominal share prices splits their shares when such demand is high. So far, the literature has not clearly identified what drives the time-varying demand for low-priced stocks.

We posit that the demand for low-priced stocks would at least partially be related to the gambling sentiment of retail investors.[5] Since low nominal share price is a salient feature of lottery-like stocks, stronger gambling sentiment would increase the demand for low-priced stocks and raise their share prices. Firms with high share prices would cater to this excess demand by splitting their shares. In contrast, firms with low nominal share prices would not split their shares.[6]

---

[5] We focus on the behavior of retail investors since past studies show that stocks splits are used to attract retail investors (Baker and Gallagher, 1980; Baker and Powell, 1993; Fernando, Krishnamurthy, and Spindt, 1999) and they are more likely to hold low-priced stocks than institutional investors (Lakonishok, Shleifer, and Vishny, 1992; Gompers and Metrick, 2001; Fernando, Krishnamurthy, and Spindt, 2004; Dyl and Elliott, 2006; Kumar and Lee, 2006).

[6] If the low-priced firms split shares, they would face substantial delisting risks. For example, for a firm with share price of $8, below the median share price of the CRSP universe of $14, a typical split ratio of 2 to 1 brings the share price down to $4. Practitioners often believe that firms with share price below $5 to

To summarize, our third hypothesis is:

H3: Firms with high share prices would exhibit a higher propensity to split their stocks when investors exhibit stronger gambling sentiment.

Our fourth hypothesis relates to another corporate finance anomaly, i.e., IPO underpricing. Loughran and Ritter (2004) show that the initial stock return after IPOs changes over time. The average first-day return doubled from 7% during 1980-1989 to 15% during 1990-1998 and surged to 65% during the 1999-2000 Internet bubble before reverting back to 12% during the 2001-2003 period. IPOs could be perceived as lotteries, given their positively-skewed returns (Barberis and Huang, 2008; Green and Hwang, 2011). Kumar, Page, and Spalt (2011) show that IPOs by firms located in regions with stronger gambling sentiment earn higher first-day returns.

Baker, Greenwood, and Wurgler (2009) show that firms choose a lower nominal offering price for IPOs when investors place relatively higher valuations on low-priced stocks. This raises post-IPO first-day return. Hence the time-variation in the IPO underpricing could be related to investors' gambling sentiment. In particular, if retail investors treat IPOs as lottery-like investment opportunities, they would be willing to pay a higher price for IPOs when their gambling sentiment is strong. This could generate a larger average first-day IPO return.

Overall, our fourth hypothesis posits that:

H4: The average first-day IPO return would be higher during periods of high gambling sentiment.

2.3. Data and methodology

To test these four hypotheses, we collect data from various sources. In this section, we describe those data sets and also describe our measure of gambling sentiment.

---

have substantial delisting risk (Market Watch: http://www.marketwatch.com/story/nyse-euronext-seeks-relax-minimum-bid). In addition, firms cannot easily undo their splits by undertaking reverse splits, as this would give negative signals to the market (e.g., Woolridge and Chambers, 1983; Campbell, Hilscher, and Szilagyi, 2008; Macey, O'Hara, and Pompilio, 2008).

2.3.1. Measures of gambling sentiment

Motivated by Da, Engelberg, and Gao (2011, 2015), we use the search volume intensity (*SVI*) for lottery-related keywords from Google to capture retail investors' gambling sentiment. Specifically, we use *SVIs* for the topic "lottery" from Google Trends,[7] at both national- and state-levels in the U.S. This includes searches in different languages and different text strings when they are lottery-related.[8] We choose the topic "lottery" for two reasons. First, the existing literature on gambling and skewness (e.g., Kumar, 2009) tends to use the payoff of state lottery to define stocks with lottery-like characteristics. Therefore, the topic "lottery" is more likely to capture investors' attitude toward stocks with similar payoff structure when compared to search topic of gambling activities with different payoff structures (e.g., casino gaming). Second, other relevant topics such as "gambling" do not have enough search volume for Google to report a complete time-series.

SVI measures the popularity of a particular search term relative to all other terms from the same location at the same time. An increase in *SVI* indicates that people pay more attention to the topic than they normally do. Google Trends reports *SVI* at weekly frequency. We aggregate this to the monthly frequency using linear interpolation, as in Da, Engelberg, and Gao (2011). In particular, we divide weekly *SVI* by 7 to obtain daily *SVI* by assuming that daily SVI is constant within the week. We then sum up daily *SVI* in a given calendar month to obtain monthly *SVI*.

To study the geographical variation in gambling attitudes, we use the average state-level *SVI* in the previous year to sort all U.S. states and the Washington D.C. into three groups with 17 states or district in each group. State-level *SVIs* are not directly comparable when downloaded separately. We deflate the *SVI* of each state by the corresponding national-level *SVI* to ensure they are comparable cross-sectionally and across time. We

[7] Google Trends reports weekly search volume intensity for various keywords. It is available at http://www.google.com/trends/.

[8] Google aggregate search volume in different languages into topics, so the search volume intensity of a topic represents the overall search interest in the selected region. In addition, to ensure that the search volume intensity is not driven by non-English search in the U.S., we also obtain the *SVI* of the following text strings: lottery. We find that the *SVI*s of the keyword search are highly correlated with the *SVI*s of the corresponding search topic. Therefore, *SVI*s of topic search are less likely to be biased by non-English search conducted in the U.S.

define strong (moderate) (weak) gambling sentiment states as the top (medium) (bottom) 17 states or Washington D.C. as measured by the average *SVI*.[9]

Following Da, Engelberg, and Gao (2011), our main variable is the abnormal search volume intensity (*ASVI*) for the topic "lottery":

$$ASVI_t = LogSVI_t - LogSVI_{t-1}, \tag{2.1}$$

where $ASVI_t$ is the abnormal search volume intensity for the topic "lottery" in month *t*. $LogSVI_t$ and $LogSVI_{t-1}$ represent the natural logarithm of *SVI*s in month *t* and month *t-1*, respectively.[10] The time-series of *ASVI* starts in March 2004 and it measures changes in people's attention toward lottery-related events.[11]

2.3.2. Validating the gambling sentiment measure

To test whether our measure of gambling sentiment is reasonable, we obtain the state lottery sales data from the North American Association of State and Provincial Lotteries (NASPL). The launch dates of state lotteries are collected from the websites of corresponding states. To measure per capita lottery sales, we obtain the demographics data from the U.S. Census Bureau. Population and education data are from the 2010 Census. We collect the news data from Factiva.

In the first validation test, we examine whether our measure of gambling sentiment correlates with news about state lotteries. Panel A of Figure 2.1 plots the "lottery" *SVI* for the U.S. Using Factiva, we find that nearly all peaks in the series coincide with the dates

[9] The purpose of deflating by state level is to measure how strong is gambling search interest in each U.S. state relative to the national level so that we could sort 51 U.S. states into 3 groups (i.e., strong, moderate and weak gambling sentiment group). We use this sorting to conduct our cross-sectional analysis in section 2.4.4 only. In other empirical analysis, we use the time-series of national level search volume intensity. Therefore, we are not giving equal weighting to all states in our analysis.

[10] Da, Engelberg and Gao (2011) define weekly frequency *ASVI* as the difference between the natural log of *SVI* in the current week and the natural log of the median value of *SVI*s during the prior 8 weeks. They argue that the median over a longer time window captures the "normal" level of attention in a way that is robust to recent jumps. This is the intuition of the name "abnormal search volume intensity". In this paper, since we use monthly frequency, so we define *ASVI* as the difference between natural log of *SVI*s between month *t* and month *t*-1 to ensure a longer period. We still call this variable *ASVI* because Da, Engelberg and Gao (2011) also call their monthly frequency measure (defined in the same way as ours) *ASVI*. We could also define abnormal *SVI* using the error term in a time-series regression. We choose the current definition to ensure the consistency with the existing literature.

[11] We also use two alternative methods to construct *ASVI*. First, we calculate *ASVI* as the log difference between *SVI* in month *t* and the median of *SVIs* in the previous three months (Da, Engelberg, and Gao, 2011). Second, we regress our baseline *ASVI* measure on month and year dummy variables and use residuals as a robustness check for any potential seasonality effects (Da, Engelberg, and Gao, 2015). Both *ASVI* measures yield quantitatively similar results.

of the largest lottery jackpots. For example, Points A to H correspond to record-breaking or near-record jackpots. These jackpots are independent of each other and are from two multi-state games, i.e., Mega Millions and Powerball. We use these attention-grabbing jackpots as natural experiments in Section 2.4.2.

Large jackpots receive greater media coverage. A search of lottery related news on Factiva illustrates this. On March 30, 2012, the drawing Friday for the $656 million jackpot, there were 1,045 lottery-related news stories in the U.S. The number of lottery-related news items reduced to 579 on the Friday one month later, almost a 50% drop in one month. This change in media coverage matches with our measure of gambling sentiment.

Next, we analyze how our state-level *SVI* relates to demographic characteristics of local investors. Panel B of Figure 2.1 depicts the geographical differences in gambling sentiment. It shows that jackpots of single-state lotto games raise mainly the *SVI* in that particular state, while jackpots of multi-state lotto games increase *SVIs* in all states. Panel C reports the regional search interest for each state. It is evident that the Internet search volume for the topic "lottery" is higher in the Western and Eastern coasts and is lower in the Central region.

Table 2.1 presents the top five and bottom five states during the 2004-2013 period. Florida and Georgia have the highest average *SVIs*, which is consistent with the fact that Powerball (Mega Millions) drawings are based in Florida (Georgia). Further, Massachusetts has one of the highest levels of Catholic concentration and it also has one of the highest average *SVIs*. In contrast, Utah has the highest level of Mormon concentration and it has one of the lowest average *SVIs*. This is consistent with the findings of Kumar, Page, and Spalt (2011) who show that Catholics are more likely to gamble while Mormons have a strong opposition to gambling.

In 2012, the median lottery sales value is $3,834 million for the top gambling sentiment states. This is 27 times greater than that of the bottom states. Obviously, these measures do not account for differences in state population. The median per capita lottery sales is $244 ($136) for the top (bottom) gambling sentiment states. In addition, we observe that the median percentage of the state population over the age of 25 that has a bachelor's degree or higher is 26.5% (29.5%) for top (bottom) states, which is consistent with Kumar

(2009) who shows that education is negatively related to the likelihood of lottery purchases.

Further, all of the top five states have legalized state lotteries. In contrast, three out of the bottom five states have not adopted state lotteries. This is similar to the findings of Kumar, Page, and Spalt (2011) who show that regions with stronger gambling propensity legalize state lotteries earlier. Overall, the validation test results indicate that our measure of gambling sentiment is reasonable.

2.3.3. Lottery-like stocks

To analyze the influence of shifts in retail investors' gambling sentiment on stock market outcomes, we focus on lottery-like stocks for our first two economic settings. Our definition of lottery-like stocks follows that of Kumar, Page, and Spalt (2016), which is a continuous measure of the "lotteriness" of stocks. This measure is based on the theoretical frameworks developed in Harvey and Siddique (2000) and Barberis and Huang (2008) and is also motivated by the empirical definition of lottery-type stocks in Kumar (2009). Specifically, we use the following three measures to construct the Lottery Index (LIDX): nominal stock price, idiosyncratic skewness, and idiosyncratic volatility. Stock price is the closing price in the last trading day of previous calendar year. Idiosyncratic skewness is the third moment of the residual obtained by fitting the following model using daily stock returns in the previous year:

$$r_i - r_f = \alpha + \beta_1\left(r_{mkt} - r_f\right) + \beta_2\left(r_{mkt} - r_f\right)^2 + \epsilon_t, \tag{2.2}$$

where r_i is the return of stock i, r_f is the risk free rate, and r_{mkt} is the market return. And, idiosyncratic volatility is the standard deviation of residual from the Carhart (1997) model using daily stock returns in the previous year:

$$r_i - r_f = \alpha + \tilde{\beta}_{mkt}\left(r_{mkt} - r_f\right) + \tilde{\beta}_{smb} r_{smb} + \tilde{\beta}_{hml} r_{hml} + \tilde{\beta}_{umd} r_{umd} + \epsilon_i, \tag{2.3}$$

where r_i is the realized return of stock i, r_f is the risk free rate, and r_{mkt} is the market return. r_{smb}, r_{hml}, *and* r_{umd} are size, market-to-book, and momentum factor returns. We obtain price, return, and market capitalization data at monthly and daily frequencies from the

Center for Research on Security Prices (CRSP). The size, market-to-book, and momentum factors are from Kenneth French's data library.[12]

Motivated by Harvey and Siddique (2000) and Kumar (2009), we use different models to estimate idiosyncratic volatility and skewness. Since the Carhart (1997) four-factor model is a standard benchmark for stock returns, we use it to estimate idiosyncratic volatility of stock returns. In addition, Harvey and Siddique (2000) suggest that excess return on an asset could be determined by its conditional covariance with both the market return and the square of the market return (conditional co-skewness). Therefore, idiosyncratic skewness could be better estimated using an asset pricing model that combines the multi-factor model with a simple nonlinear component derived from skewness. This choice is also consistent with the empirical finding in Ghysels (1998).

In January of each year, we assign all common stocks (with a share code of 10 or 11) in the CRSP universe into twenty groups based on each criterion. We conduct the three sorting independently and create 60 groups. Group 20 (1) contains the stocks with the highest (lowest) idiosyncratic skewness, highest (lowest) idiosyncratic volatility, or lowest (highest) price. We then add up the group numbers of each stock to a score between 3 and 60 and standardize this score to a value between 0 and 1 using LIDX = (Score-3) / (60-3).[13] Finally, we define lottery-like stocks as stocks with a top 30% LIDX value, non-lottery stocks as those with a bottom 30% LIDX value, and remaining stocks as other stocks. We update this list in January of each year.

Panel A of Table 2.2 presents the main characteristics of lottery-like stocks. For comparison, we also report the characteristics of non-lottery stocks, other stocks, and all stocks in the CRSP universe. The average price of lottery-like stocks is $5.67, which is comparable in magnitude to the price of lottery tickets.[14] Lottery-like stocks have a small average market capitalization of $266 million. They have higher market-to-book ratio

[12] The factors are obtained from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[13] For example, if stock A is in group 1 for idiosyncratic skewness, group 20 for idiosyncratic volatility, and group 20 for price. The score for stock A equals to 1 + 20 + 20 = 51. We standardize this score to a value between 0 and 1: LIDX = (51 - 3)/60 – 3 = 0.84.

[14] For instance, the ticket prices for the two largest lotto games in the U.S., Mega Millions and Powerball, are $1 and $2, respectively. *Source:* Mega Millions (http://www.megamillions.com/), and Powerball (http://www.powerball.com/pb_home.asp).

than non-lottery stocks. As expected, they also have significantly higher volatility and skewness.

2.3.4. Brokerage data and macroeconomic variables

To directly examine the potential spillover effects of jackpots on lottery-like stocks, we obtain trading data from a major U.S. discount brokerage house. This data set contains all trades of a set of individual investors during the 1991-1996 period. We examine trades on common stocks.[15] During this period, the only available multi-state lottery game is Powerball, which started from April 22, 1992. We obtain draw date, winners, and jackpot prize of each drawing from the Multi-State Lottery Association (i.e., the operator of Powerball).[16]

Additionally, we use five commonly used macroeconomic variables to capture the effects of business cycles: U.S. monthly unemployment rate (UNEMP), unexpected inflation (UEI), monthly growth in industrial production (MP), monthly default risk premium (RP), and the term spread (TS). We obtain UNEMP from the Bureau of Labor Statistics web site. UEI is the difference between the current month inflation and the average of the past 12 realizations. We obtain MP from the Federal Reserve web site. RP is the difference between Moody's Baa-rated and Aaa-rated corporate bond yields. TS is the difference between the yields of a constant maturity 10-year Treasury bond and 3-month Treasury bill. Summary statistics for these variables are reported in Panel B of Table 2.2.

2.3.5. Institutional ownership data

Our second hypothesis posits that gambling sentiment would have greater impact on stocks that are more likely to be held by retail investors, i.e., small stocks or stocks with low institutional ownership. Small stocks are defined as stocks in the bottom 30th percentile by market capitalization in the CRSP universe. Firm-level institutional ownership data are collected from FactSet and based on Ferreira and Matos (2008). We measure a firm's institutional ownership in a year by its average quarterly total institutional ownership. The mean of total institutional ownership is 55% for our sample

[15] Details on the brokerage data are available in Barber and Odean (2000).

[16] We thank Multi-State Lottery Association for providing the historical Powerball information.

period (see Panel B of Table 2.2). Low institutional ownership stocks are stocks with less than ten percent total institutional ownership.

2.3.6. Stock splits data

Our third economic setting focuses on the implication of time-varying gambling attitudes on stock splits. We include all common stocks in the U.S. and identify splitters as those with a CRSP distribution code of 5523. We study stock splits that reduce stock prices, so reverse stock splits are not included in our sample. Following Lin, Singh, and Yu (2009), we require splitters to have a CRSP Factor to Adjust Price (FACPR) greater than or equal to one and equal to the CRSP Factor to Adjust Shares Outstanding (FACSHR). After dropping stocks without COMPUSTAT data, our sample includes 490 stock splits from January 2005 to December 2013. The average monthly probability of stock splits is 0.12% (see Panel B of Table 2.2).

2.3.7. IPO data

In our fourth economic setting, we analyze the effects of time-varying gambling attitudes on first-day returns of IPOs. We obtain the monthly average first-day return on the "net IPOs" from Jay Ritter's website.[17] Net IPOs are IPOs excluding closed-end funds, REITs, acquisition companies, stocks with offer prices below $5, ADRs, limited partnerships, units, banks and S&Ls, and those not listed on CRSP, as defined in Ibbotson, Sindelar, and Ritter (1994). The first-day return is calculated as the percentage return from the offering price to the first closing bid price. The monthly average first-day return is calculated as the equal-weighted average of the first-day returns on all the offerings in a particular calendar month. During our sample period, the average post-IPO first-day return is 13.5% (see Panel B of Table 2.2).

2.4. Empirical results

2.4.1. Stock return predictability

Our first hypothesis focuses on the impact of time-varying gambling attitudes on stock returns. If elevated gambling sentiment increases the demand for lottery-like stocks and

[17] See http://bear.warrington.ufl.edu/ritter/ipodata.htm.

generates price pressure on these stocks, *ASVI* should have a positive impact on the abnormal return of lottery-like stocks in the short-run. Our tests examine whether this short-term return predictability exists.

To measure the abnormal return performance of lottery-like stocks, we use the Carhart (1997) four-factor model to account for size, market-to-book, and past performance. We estimate 36-month rolling-window regressions and require all stocks to have at least 12 months of return data. After estimating the factor loadings using equation (2.3), we calculate the abnormal return for each stock as:

$$AR_{i,t} = r_i - r_f - \tilde{\beta}_{mkt,t-1}(r_{mkt} - r_f) - \tilde{\beta}_{smb,t-1} r_{smb} - \tilde{\beta}_{hml,t-1} r_{hml} - \tilde{\beta}_{umd,t-1} r_{umd}, \quad (2.4)$$

where $AR_{i,t}$ is the abnormal return of stock *i* in month *t*. Factor loadings are estimated from month *t-36* to *t-1*. The abnormal returns are then value-weighted to obtain the portfolio return.[18]

Following Da, Engelberg, and Gao (2011), we estimate the following regression to determine if stock returns are predictable in the short-run:

$$AR_{portfolio,t+n} = \alpha + \beta_n \times ASVI_t + \epsilon_t, \quad n = 0, 1, 2, 3, \quad (2.5)$$

where $AR_{portfolio,t+n}$ is the average abnormal return in month *t+n* of a stock portfolio weighted by market capitalization in month *t+n-1*. The coefficient β_n measures the predictive power of *ASVI* with *n* lags.

The coefficient estimates in Table 2.3 support our conjecture. The β_n coefficients are positive in months 0 and 1 for lottery-like stock portfolio. In economic terms, a one standard deviation increase (i.e., 20%) in the *ASVI* for the topic "lottery" is associated with a significantly positive price change of 47 basis points in month 1.[19] The coefficient estimates become negative from month two onward, indicating a subsequent price reversal as the mispricing get corrected. In economic terms, a one standard deviation increase in *ASVI* significantly reduces lottery-like stocks' abnormal returns in month 3 by 33 basis points. In contrast to lottery-like stock portfolio, *ASVI* does not have any power to predict the return of non-lottery stock and other stock portfolios. Further, the estimates

[18] Our results are similar if we first form portfolios and then estimate abnormal returns at the portfolio-level.
[19] The price impact of large jackpots on lottery-like stocks is likely to occur in month 1 as large jackpots are claimed at the end of month 0.

in Column 5 show that the return predictability is stronger when we Long lottery-like stocks and Short non-lottery stocks simultaneously. The abnormal return of lottery-like stocks disappears after three months.

Overall, the results in Table 2.3 support our first hypothesis. Lottery-like stocks earn significantly positive abnormal returns when investors have stronger gambling sentiment. This is consistent with our conjecture that retail investors' gambling sentiment would generate short-term overpricing among lottery-like stocks.

2.4.2. Attention-grabbing jackpots and predictability

One potential explanation for our evidence of short-term return predictability among lottery-like stocks could be that those stocks experience positive abnormal returns when investors pay more attention to lottery-related events. In this section, we use lottery jackpots that grab investors' attention to identify the sources of time-variation in investors' gambling attitudes. These are exogenous events and do not require the gambling sentiment measure from Google.[20]

We define attention-grabbing jackpots in three different ways. First, the *all-time largest* jackpot of $656 million was awarded on March 30, 2012, which is the largest jackpot in the U.S. history. Second, *record-breaking* jackpots are those that break national record at the time. This criterion gives us the *all-time largest* jackpot and two other jackpots announced on February 18, 2006 and March 6, 2007. Third, we use a broader definition of *attention-grabbing* jackpots. In addition to the three *record-breaking* jackpots, we include near-record jackpots that are either the second largest jackpot at the time or the largest jackpot over the past 24 months. If two jackpots selected are next to each other such that event windows are contaminated, we drop both of them from our sample. The above criterion leads to eight attention-grabbing jackpots in our sample period. The sizes and dates of eight attention-grabbing jackpots are reported in the Appendix (Table 2.1A). These jackpots take on average two months from the first to the last drawing dates. Their values are between $336 million and $656 million.

[20] The occurrences of large jackpots are random and are unlikely to be driven by factors that affect the stock market. For example, the winning odds for Mega Millions and Powerball jackpots in 2015 are as low as 1 in 258,890,850 and 1 in 175,223,510, respectively. *Source*: Mega Millions and Powerball web sites.

After identifying the lottery jackpots, we use a calendar-time portfolio approach to measure the impact of attention-grabbing jackpots on stock returns:

$$r_{portfolio,t} - r_{f,t} = \alpha + \beta_1 D^{Jackpot}_{[-1,+1]} + \beta_2 D^{Post}_{[+2,+3]} + \beta_3 (r_{mkt} - r_f) + \beta_4 r_{smb} + \beta_5 r_{hml} + \beta_6 r_{umd} + \epsilon_{t,} \quad (2.6)$$

where the dependent variable is the average excess return of lottery-like or non-lottery stocks in month *t*. The final drawing dates of attention-grabbing jackpots are in month 0.[21] $D^{Jackpot}_{[-1,+1]}$ is a dummy variable that equals to one from month -1 to month +1 and zero otherwise. β_1 measures average abnormal return during the (-1, +1) period. $D^{Post}_{[+2,+3]}$ is a dummy variable that equals to one during the (+2, +3) period and zero otherwise. β_2 measures average abnormal return during the (+2, +3) period. Standard errors are adjusted for auto-correlation using the Newey and West (1987) method.

Table 2.4 reports the estimated coefficients of β_1 (Panel A) and β_2 (Panel B). We find that lottery-like stocks earn significantly positive returns around attention-grabbing jackpots. The average abnormal return during the (-1, +1) period is between 1.55% and 1.75% per month. This short-term mispricing is partially corrected during the (+2, +3) period. The price reversal in months (+2, +3) is -1.23% for the all-time largest jackpot and -1.10% for the record-breaking jackpots. These results are consistent with our return predictability results in Table 2.3. Again, abnormal return of lottery-like stocks becomes statistically insignificant beyond month +3.

Next, we examine the abnormal trading volumes around jackpots. Following Chae (2005), the abnormal trading volume is calculated as:

$$Abnormal\ Trading\ Volume_{i,t} = \tau_{i,t} - \bar{\tau}_i, \quad (2.7)$$

where $\tau_{i,t}$ is the log-transformed turnover (i.e., trading volumes divided by outstanding shares) for stock *i* in month *t* and $\bar{\tau}_i$ is the average log-transformed turnover during the estimation period, which has a length of 36 months and ends three months before the event.[22]

[21] We include month -1 as a run-up period because investors pay attention to lotteries when jackpot prizes pass certain thresholds (see Williams and Siegel, 2014).

[22] We also use an alternative approach to estimate abnormal trading volume. We adjust the log-transformed turnover by the market model where the market volume is the value-weighted log turnover of stocks listed on NYSE, AMEX and NASDAQ, as in Campbell and Wasley (1996). The results are quantitatively similar.

Table 2.5 presents the results. For attention-grabbing jackpots, lottery-like stocks experience significantly positive abnormal trading volume of about 17% during the (-1, +1) period and about 21% during the (+2, +3) period. This evidence suggests that changes in gambling sentiment induce changes in trading activity among lottery-like stocks. We find that the abnormal trading volume of lottery-like stocks is not significant around the all-time largest jackpot. One possible explanation is that it may be difficult to make conclusions about individual-level behavior using aggregate level data (e.g., trading volume). Therefore, in the next section, we use the actual trades of retail investors to directly examine their investment activities around large jackpots.

Overall, our results in this section suggest that lottery-like stocks experience short-term overpricing and abnormal trading volume due to an increase in gambling sentiment among retail investors. This spillover effect on lottery-like stocks is consistent with evidence from the economics literature on betting markets.[23]

2.4.3. Lottery jackpots and investor trading: direct link

Our results so far suggest that lottery-type stocks are overpriced when the overall gambling sentiment is high. However, we have not yet shown that shift in gambling attitudes is positively correlated with investor trading. Without this direct link, our results may have alternative explanations, especially because evidence in recent studies (i.e., Dorn, Dorn, Sengmueller, 2015; Gao and Lin, 2015) indicates that gambling and stock market trading activities may be negatively correlated, i.e., lotteries and stocks serve as substitutes rather than complements.[24]

---

[23] A related literature in economics finds that state lotteries are complements to other forms of gambling. For example, the introduction of state lotteries increases the participation in casino gaming and horse racing (Scott and Garen, 1994; Calcagno, Walker, and Jackson, 2010). Increases in gambling expenditure by households are associated with reductions in non-gambling expenditure, rather than reductions in other types of gambling expenditure (Kearney, 2005). In addition, different types of U.S. lotteries complement each other (Grote and Matheson, 2007).

[24] Using data from Taiwan, Gao and Lin (2015) find a negative impact of lottery jackpots (i.e., large jackpot drawings) on stock trading volume. Further, using the same brokerage data as ours in the U.S., Dorn, Dorn, and Sengmueller (2015) report similar findings as Gao and Lin (2015). But their results are insignificant for the 2004- 2008 period, which overlaps with our main sample period, using the TAQ trading data. Our paper is different from the previous studies because we analyze (i) US multi-state jackpots rather than local jackpots in California or Taiwan as in previous studies; (ii) the realization of large attention-grabbing jackpots rather than daily balance of all jackpots; and (iii) the excess buy-sell imbalance of lottery-like stocks rather than trading volume of all stocks. Our results are opposite, which suggests that results from

In this section, we test directly whether retail investors increase aggregate demand for lottery-like stocks around large jackpots. We use the actual trades of retail investors from a large discount brokerage house during the 1991-1996 period. We use two types of lottery measures. First, we study the largest jackpot during the 1992-1996 period (i.e., the $111 million prize announced on July 7, 1993). Second, following Gao and Lin (2015), we examine large drawings, which include either claimed jackpots or unclaimed balances.

To examine the impact of large jackpots on investor trading, we measure the aggregate demand for lottery-like stocks as the excess buy-sell imbalance (*EBSI*) defined as the difference in buy-sell imbalance between lottery-like and non-lottery stock portfolios (Kumar, 2009).[25] This measure captures the change in investors' bullishness towards lottery-like stocks relative to their change in bullishness towards non-lottery stocks. Specifically, we estimate the following time-series regression:

$$\begin{aligned} EBSI_t = \alpha + \beta_1 D_{[-1,+1]}^{Jackpot} + \beta_2 MKTRET_t + \beta_3 MKTRET_{t-1} + \beta_4 LOTRET_t \\ + \beta_5 LOTRET_{t-1} + \beta_6 EBSI_{t-1} + Controls_{t-1} + \epsilon_t. \end{aligned} \tag{2.8}$$

The dependent variable is the monthly excess buy-sell imbalance for lottery-like stocks. Lottery-like stocks are defined as in Section 2.3.3. The set of independent variables includes contemporaneous and one month lagged market returns, contemporaneous and one month lagged returns of the lottery-like stock portfolio. We also include lagged *EBSI* to control for potential serial correlation in that measure. Additionally, we include *UNEMP*, *UEI*, *MP*, *RP* and *TS* as control variables to account for potential business cycle effects since investors are known to have stronger gambling sentiment during economic recessions (Kumar, 2009). The sample period is from April 1992 to November 1996. Standard errors are calculated using the method in Newey and West (1987).

Taiwan cannot be directly extrapolated to the U.S. setting. More importantly, our results suggest that it may be difficult to make conclusions about individual-level behavior using aggregate market-level data.

[25] The buy-sell imbalance (*BSI*) of portfolio *p* in month *t* is defined as $BSI_{pt} = \frac{100}{N_{pt}} \sum_{i=1}^{N_{pt}} BSI_{it}$, where the *BSI* for stock *i* in month *t* is defined as $BSI_{it} = \frac{\sum_{j=1}^{D_t}(VB_{ijt} - VS_{ijt})}{\sum_{j=1}^{D_t}(VB_{ijt} + VS_{ijt})}$. Here, D_t is the number of days in month *t*. VB_{ijt} (VS_{ijt}) is the dollar buying (selling) volume of stock *i* on day *j* in month *t*, and N_{pt} is the number of traded stocks in portfolio *p* in month *t*. Kumar and Lee (2006) show that an equal-weighted *BSI* measure is more appropriate for capturing shifts in investor sentiment than a value-weighted measure.

The key variable of interest is $D_{[-1,+1]}^{jackpot}$, which equals to one from month -1 to month +1 around the event jackpot, and zero otherwise. A positive and significant coefficient on $D_{[-1,+1]}^{jackpot}$ would indicate that trading in lottery-like stocks increases around lottery jackpot announcements.[26] Table 2.6 presents the results. Consistent with our expectation, we find a significantly positive coefficient of 7% on $D_{[-1,+1]}^{jackpot}$, which suggests that jackpot announcement generates 7% higher net purchase of lottery-like stocks relative to non-lottery stocks.

Next, we examine the effect of large drawings on daily *EBSI*. We have 228 large drawings during our sample period. Since Powerball drawings were held on Wednesday and Saturday evenings at 10:59 pm, we examine the spillover effect on the next trading day (i.e., Thursdays and Mondays). Specifically, we estimate the following time-series regression:

$$EBSI_t = \alpha + \beta_1 D_t^{Drawing} + \beta_2 MKTRET_t + \beta_3 LOTRET_t + \beta_4 EBSI_{t-1} + \beta_5 VIX_{t-1} + \beta_6 ADS_{t-1} + Controls + \epsilon_t. \quad (2.9)$$

The dependent variable is daily excess buy-sell imbalance for lottery-like stocks. The set of independent variables includes market return, return of the lottery-like stock portfolio, and lagged *EBSI*. We also include the lagged Chicago Board Options Exchange daily market volatility index (*VIX*) to account for investor fear and market sentiment, and include lagged Aruoba-Diebold-Scotti business conditions index (*ADS*) to capture the economic condition at the daily-level (Da Engelberg, and Gao, 2015). Control variables include lagged market and lottery-like stock portfolio returns (up to five lags) and day-of-the-week dummies. Standard errors are calculated using the method in Newey and West (1987).

The key variable of interest is $D_t^{Drawing}$, which equals to one on the next trading day following a large drawing, and zero on days with no drawings or small drawings. A large (small) drawing has above (below) median drawing value during the April 22, 1992 to

[26] Although the brokerage data cover trades of a set of retail investors from all U.S. states, only 14 states participated in Powerball at its inception in 1992. Large states such as New York and California did not join Powerball during the 1992-1996 period. Therefore, our estimate is conservative. We expect a smaller effect using the brokerage data.

November 30, 1996 period. Table 2.7 presents the results. We find a significantly positive coefficient of 3% on $D_t^{Drawing}$, which suggests that large drawings generate 3% more net purchase of lottery-like stocks on the following day relative to non-lottery stocks. The magnitude of *EBSI* is smaller because large drawings attract less attention than lottery jackpots.

Overall, the results in Tables 2.6 and 2.7 show that attention-gabbing jackpots motivate retail investors to increase their demand for lottery-like stocks. This evidence suggests that shifts in gambling attitudes have a positive spillover effect on the stock market in the U.S.

2.4.4. Cross-sectional variation in return predictability

Next, we study whether cross-sectional differences in gambling sentiment shifts affect stock performance in the long-term. We test whether the negative lottery stock premium is stronger among states with stronger gambling sentiment. Barberis and Huang (2008) show that a security's idiosyncratic risk could be priced. To get exposure to idiosyncratic skewness, investors with strong gambling attitudes are willing to accept a lower risk adjusted return for stocks with positive return skewness. We conjecture that this effect is more likely to exist among investors in U.S. states with stronger gambling attitudes. Since previous literature has documented that investors have local biases, we use the average state-level *SVI* for the topic "lottery" to form portfolios based on gambling sentiment of the state in which a firm is headquartered. Specifically, we first define value-weighted portfolios for lottery-like, non-lottery, and other stocks. Then, we estimate the alphas of these portfolios using monthly return regressions that include the four factors as the benchmark.

Table 2.8 shows that lottery-like stocks significantly underperform non-lottery stocks. A trading strategy that goes Long in lottery-like stocks and Short in non-lottery stocks significantly underperforms the four-factor benchmark by 43 basis points per month. Interestingly, this effect is driven by firms headquartered in states with strong gambling sentiment. In contrast, lottery-like and non-lottery stocks do not have significantly different performance when they are headquartered in U.S. states with moderate or weak gambling sentiment.

Next, we further investigate whether the lottery-like stock premium is stronger for stocks that are more likely to be held by retail investors since these investors are more likely to gamble in the stock market. We focus on stocks with low price and low institutional ownership. We focus on the cross-sectional differences among firms that are headquartered in states with strong gambling sentiment. Table 2.9 reports the long-term performance of stocks sorted by institutional ownership or firm size. Panel A shows that lottery-like stocks in high gambling sentiment states have larger underperformance when the institutional ownership is lower. Specifically, lottery-like stocks underperform non-lottery stocks by about 1.3% per month when institutional ownership is below ten percent. The abnormal return difference decreases to 53 basis points for stocks with above ten percent institutional ownership.

Panel B shows that among stocks ranked in the bottom 30% by size, lottery-like stocks significantly underperform non-lottery stocks by 1.4% per month. In contrast, such underperformance does not exist for large stocks. These findings suggest that gambling sentiment has a larger impact on stocks returns for stocks that are more likely to be held by retail investors.

Collectively, the results in Tables 2.8 and 2.9 support our second hypothesis. Consistent with our conjecture, we find that in regions with strong gambling sentiment, local investors are willing to accept more negative risk-adjusted return for lottery-like stocks. This evidence is more pronounced among stocks with low institutional ownership or low market capitalization. In contrast, in U.S. regions with weak gambling sentiment, lottery-like stocks do not significantly underperform non-lottery stocks.

2.4.5. Gambling sentiment and stock splits

Previous finance literature has shown that firms are more likely to split their shares when stock prices are high (e.g., Baker and Powell, 1992; Dyl and Elliott, 2006; Minnick and Raman, 2014). Our third hypothesis posits that an increase in gambling sentiment would lead to a higher probability of stock splits for stocks with high nominal prices.

In this section, we use logistic regression to estimate the influence of gambling sentiment on stock splits. The dependent variable in this regression is equal to one if the company splits its shares in a given month, and zero otherwise. We control for stock return, lagged firm size, and market-to-book ratio. We also control for split activities in the previous year. Specifically, we run the following logistic regression:

$$Logit(Split_{i,t}) = \alpha + \beta_1 D_{ASVI\,t-1} + \beta_2 D_{p\,i,t-1} + \beta_3 D_{ASVI\,t-1} \times D_{p\,i,t-1} + \beta_4 return_{i,t} + \beta_5 size_{i,t-1} + \beta_6 MTB_{i,t-1} + \beta_7 splitter_{i,t-12} + \epsilon_{t,} \qquad (2.10)$$

where $D_{ASVIt-1}$ is a dummy variable that equals one if investors have strong gambling sentiment. We define strong gambling sentiment as *ASVI* values above the 75th percentile value of the time-series. $Dp_{i,t-1}$ is a dummy variable of price that equals to one if a given stock is a high-priced stock. A high-priced stock has share prices above the 75th price percentile of all common stocks in the CRSP universe in a given month.[27] $D_{ASVIt-1} \times Dp_{i,t-1}$ is the interaction between the price and the gambling sentiment dummy variables.

Among other variables, $return_{i,t}$ is the return excluding dividends of stock *i* over the course of the month *t*, as in Baker, Greenwood, and Wurgler (2009). $Size_{i,t-1}$ is the natural logarithm of the market capitalization of stock *i* in month *t-1* while $MTB_{i,t-1}$ is the market-to-book ratio defined as the market value of the firm over its book value. Market value equals to market equity at calendar year-end plus book debt, while book value is calculated as stockholders' equity minus preferred stock plus deferred taxes and investment tax credits and post retirement assets. $Splitter_{i,t-12}$ is a dummy variable that equals to one if a firm split its stocks in the previous year. Standard errors are clustered by firm and by date.

The key variable of interest is the interaction between the price and gambling sentiment dummy variables. We expect that the splitting propensity would be high when share price is high and gambling sentiment is strong.

Table 2.10 reports the estimation results. We find that the interaction between price and gambling sentiment dummy variables is positively significant in all specifications, which supports our third hypothesis. The price dummy variable is significant at the 1%

[27] During our sample period, the average value of the 75th price percentiles is \$30 with a minimum value of \$15 and a maximum of \$43. Our definition of high-priced stock is similar to the 70th price percentile used in Baker, Greenwood, and Wurgler (2009). Our definition of high-priced stocks is also motivated by the minimum bid price requirements of major stock exchanges. Both NYSE and NASDAQ require listed firms to have a share price of at least \$1. Firms that fail to meet this requirement can be delisted. During the 2007 financial crisis, hundreds of firms traded below \$2. In 2008 alone, 85 firms (10% of all listed firms in NASDAQ) were delisted from NASDAQ, mostly for not meeting the \$1 price requirement. In general, firms trading in the sub-\$5 range face substantial delisting risk.

level, which suggest that high-priced firms are more likely to conduct stock splits regardless of the level of gambling sentiment. In contrast, the gambling sentiment dummy variable is insignificant, which suggests that not all firms are likely to have a high propensity to split during high gambling sentiment periods. This is expected because low-priced firms are not likely to split shares during high gambling sentiment periods for two reasons: (i) these stocks already have the low price lottery-like characteristics, and (ii) further split will incur substantial delisting risks. A positively significant interaction effect suggests that gambling sentiment affects the split probability only when share price is high.

We use the Marginal Effects at the Means for discrete variables to calculate the marginal effect of high gambling sentiment on the probability of stock splits. We focus on the marginal effect when the stock has a high price while holding other variables at their means. We find that that, in economic terms, for a high-priced stock with average values on size, book-to-market ratio, return, and splitting activities in the past 12 months, a one-unit increase in the dummy variable for gambling sentiment raises the split probability by 0.10% per month.[28] This effect is economically significant since the average monthly split probability is 0.12%. In addition, we find that small stocks and stocks with higher returns are more likely to split, which is consistent with the existing evidence in the literature.

Overall, the evidence in Table 2.10 provides support to our third hypothesis. We demonstrate that retail investors' gambling sentiment plays an important role in explaining the time-varying demand for low-priced stocks. When investors' gambling sentiment is strong, high-priced firms are more likely to split stocks to cater to the excess demand for low-priced stocks.

2.4.6. Underpricing of IPOs

In this section, we focus on our fourth hypothesis and examine whether gambling sentiment helps to explain the time-variation in first-day IPO returns. We regress the average monthly first-day return of the net IPOs against lagged *ASVI*. Following Baker,

[28] In specification 5, when the price dummy variable equals to one and the *ASVI* dummy variable equals to zero, the predicted split probability is 0.51% per month keeping control variables at their mean. In contrast, the predicted split probability increases to 0.61% if the *ASVI* dummy variable increases to one.

Greenwood, and Wurgler (2009), we control for the average log price at the beginning of the month and the value-weighted market return excluding dividends over the course of the month. We also include the hotness of IPO market and the monthly number of net IPOs as additional controls. Our sample period for the test is from January 2005 to June 2014. This gives us a time-series of 107 monthly observations with IPO data.

We estimate the following regression:

$$r_t^{IPO} = \alpha + \beta_1 ASVI_{t-1} + \beta_2 p_{t-1} + \beta_3 VWMKT_t + \beta_4 hotness_t + \beta_5 IPOnumber_t + \epsilon_{t,} \quad (2.11)$$

where p_{t-1} is the equally-weighted average log price in the previous month by using all common stocks in the CRSP universe, and $VWMKT_t$ is the value-weighted return of all common stocks over the course of month *t*. Following Ibbotson, Sindelar, and Ritter (1994), $Hotness_t$ is the percentage of deals that priced above the midpoint of the original file price range in month *t*. $IPO\ number_t$ is the natural logarithm of the monthly number of net IPOs in month *t*. We do not examine the long-run performance of IPOs because our sample period gets significantly shortened.

Table 2.11 reports the estimation results. After controlling for the market return, average price level and the hotness of IPO market, the coefficient estimates of *ASVI* are positive and significant at the 1% level. Economically, a one standard deviation increase in *ASVI* (i.e., 20%) is associated with a 1.62% increase in the average first-day IPO return. Relative to the mean first-day return of 13.44%, this reflects an economically meaningful 12.05% increase. Consistent with our fourth hypothesis, this evidence suggests that when investors have stronger gambling sentiment, IPOs experience higher average first-day returns.

2.4.7. Robustness checks and alternative explanations

In this section, we conduct a number of robustness checks for our baseline results. First, we include five commonly used macroeconomic variables in our return predictability regressions to account for business cycle effects since investors are more likely to gamble during economic recessions (Kumar, 2009). The estimation results are reported in Panel A of Table 2.12, which are similar to the baseline return predictability estimates in Table 2.3. This evidence suggests that business cycles in the U.S. cannot explain the predictive power of our gambling sentiment measure.

Second, we test whether our findings can be explained by other investor sentiment proxies. In particular, Baker and Wurgler (2006, 2007) construct an investor sentiment index by using the first principal component of six sentiment proxies, where each of the proxies has been orthogonalized with respect to a set of macroeconomic variables. These data are available until 2010. In Panel B of Table 2.12, Column 1 (2) reports the return predictability results without (with) the investor sentiment variable from Baker and Wurgler (2007). Our results remain similar, which suggests that our results on gambling sentiment cannot be explained by the other investor sentiment measures.

Third, search volume intensity from Google was publicly available only after June 2006. Column 3 shows that our return predictability results are similar for the sub-period that starts in June 2006. In Column 4, we also control for the market-wide investor sentiment for this sub-period and our results remain similar. Thus, the stock return predictability patterns that we document exist even after Google's search volume intensity data are made public.

Fourth, we conduct several robustness checks for our stock split analysis. These results are summarized in Panel C of Table 2.12. In Column 1, we include firm fixed effects to control for unobservable firm characteristics. This allows us to focus on firms that have time-series variations in stock splits. Our results remain similar. In addition, in Columns 2 to 7, we include macroeconomic controls and our results still hold. In untabulated tests, we also include returns over the past three, six, or twelve months as additional control variables, since past literature suggests that split decisions are a function of a long period of lagged returns. Our results remain similar after including these lagged return variables.

In the last robustness check, we reconsider the first-day IPO returns analysis, where we include five macroeconomic variables in the regression specification. Our results remain robust (see Panel D of Table 2.12).

2.5. Summary and Conclusion

This study investigates how changes in overall attitudes toward gambling affect financial market outcomes. Using a novel measure of gambling sentiment based on lottery-related Internet search volume, we show that the time-variation in gambling attitudes predicts the returns of lottery-like stocks. Further, using attention-grabbing lottery jackpots as exogenous shocks to gambling sentiment, we show that our results do not reflect potential reverse causality. We find that large lottery jackpots not only increase people's participation in lotteries, but also enhance investors' propensity to purchase stocks with lottery-like characteristics. Analyzing trades of retail investors from a major U.S. discount brokerage firm, we also show directly that investors increase aggregate demand for lottery-like stocks around large jackpots and large drawings.

Examining geographical differences in the impact of gambling sentiment on market outcomes, we find that in U.S. states where gambling attitudes are strong, lottery-like stocks underperform stocks that are otherwise similar in the long-run. These effects are stronger for smaller firms and firms with lower institutional ownership.

The time-variation in gambling attitudes also affects corporate financial decisions. Specifically, firms with high nominal share prices are more likely to split their shares when investors' gambling sentiment becomes stronger. Stronger gambling sentiment is also associated with higher first-day returns of initial public offerings. Collectively, these results suggest that shifts in overall gambling attitudes have spillover effects on financial markets.

These findings contribute to the growing finance literature that examines the role of gambling in financial markets. Our paper adds a new dimension to this literature by demonstrating that time-variation in gambling attitudes generates short-term mispricing and also affects corporate decisions. In future work, it may be interesting to examine whether time-varying gambling attitudes influence mutual fund flows. Mutual funds that hold more lottery-like stocks could experience more cash inflows when gambling sentiment is strong. It would also be interesting to examine the influence of time-varying gambling sentiment on other lottery-like securities such as options.

Table 2.1: Top and bottom gambling sentiment states

This table reports characteristics of the top and bottom five states in terms of the average search volume intensity for the topic "lottery" from 2004 to 2013. *Annual sales* (reported in million $) presents the total lottery sales value in fiscal year 2012. *Population* shows the state-level total population according to the 2010 Census. *Per capital sales* is calculated as lottery sales divided by population. *Education* reports the proportion of state population over the age of 25 that has obtained a bachelor's degree or higher. *Launch year* reports the year when the first state lottery ticket is on sale. *Average SVI* is the average annual search volume intensity, which is aggregated from weekly SVIs.

Panel A: Top five states in gambling sentiment

States	(1) *Annual sales*	(2) *Population*	(3) *Per capital sales*	(4) *Education*	(5) *Launch year*	(6) *Average SVI*
Florida	4,449.90	18,801,310	236.68	26.52	1988	147.34
Georgia	3,834.70	9,687,653	395.83	29.52	1993	92.23
Massachusetts	4,741.40	6,547,629	724.14	39.05	1972	77.04
Michigan	2,413.46	9,883,640	244.19	26.31	1972	75.42
Tennessee	1,311.00	6,346,105	206.58	23.42	2004	73.14
Average	3,350.09	10,253,267	361.48	28.96		93.03
Median	3,834.70	9,687,653	244.19	26.52		77.04

Panel B: Bottom five states in gambling sentiment

States	(1) *Annual sales*	(2) *Population*	(3) *Per capital sales*	(4) *Education*	(5) *Launch year*	(6) *Average SVI*
Hawaii	N/A	1,360,301	N/A	29.52	N/A	9.96
Utah	N/A	2,763,885	N/A	30.29	N/A	10.15
Alaska	N/A	710,231	N/A	26.67	N/A	11.37
Idaho	175.84	1,567,582	112.17	25.70	1989	12.04
Vermont	100.93	625,741	161.30	34.82	1978	13.69
Average	138.39	1,405,548	136.73	29.40		11.44
Median	138.39	1,360,301	136.73	29.52		11.37

Table 2.2: Summary statistics

This panel reports the characteristics for lottery-like stocks, non-lottery stocks, and other stocks. Variables are calculated as the monthly average from 2005 to 2013. Lottery-like stocks are defined as stocks within the upper 30 percentiles of Lottery Index (LIDX) in each year. Similarly, Non-lottery stocks are defined as stocks in the bottom 30 percentiles of LIDX in each year. Other stocks are defines as the rest of stocks in CRSP. *Number of stocks* reports the average number of lottery-like, non-lottery, other and all common stocks in the CRSP universe in each year. *Stock price* is the average price. *Stock return* is the monthly realized return. *Firm size* in million U.S. dollar is calculated as stock price multiplied by shares outstanding. *MTB Ratio* is defined as the market value of the firm over its book value. Market value equals to market equity at calendar year-end plus book debt while book value is calculated as stockholders' equity minus preferred stock plus deferred taxes and investment tax credits and post retirement assets. *Trading volume* is the log-transformed turnover (i.e., total shares traded divided by outstanding shares). *Idiosyncratic volatility* is the standard deviation of the residual from the four-factor model. *Total skewness* (*kurtosis*) is the third (fourth) moment of monthly stock returns. *Idiosyncratic skewness* is the scaled measure of the third moment of the residual from a two factor model (i.e., equation (2.2)). *Observations* is the number of firm-month observations.

Panel A: Characteristics of lottery-like stocks

Variables	(1) Lottery-like stocks	(2) Non-lottery stocks	(3) Other stocks	(4) CRSP all stocks
Number of stocks	1,269	1,288	1,721	4,278
Stock price	5.67	135.03	19.42	51.10
Stock return	0.87%	0.80%	0.84%	0.84%
Firm size ($M)	266.54	8,995.79	1,694.37	3,534.46
Total volatility	19.63%	8.61%	12.40%	13.39%
Idiosyncratic volatility	18.80%	7.62%	11.40%	12.48%
Total skewness	0.52	0.08	0.23	0.27
Idiosyncratic skewness	0.54	0.15	0.28	0.32
Kurtosis	1.19	0.51	0.67	0.77
MTB ratio	2.22	1.68	1.73	1.85
Trading Volume	-2.71	-2.14	-2.37	-2.40
Observations	127,858	137,200	180,136	445,194

Table 2.2 (Cont'd)

This panel reports the summary statistics of other variables for our empirical analyses. *IPO return* is the monthly average first-day return (in percentage) on the net IPOs. *Hotness* reports the percentage of deals that priced above the midpoint of the original file price range. *IPO number* is the natural logarithm of the monthly number of net IPOs. All the above three variables are obtained from Jay Ritter's website. *Market return* (i.e., $VWMKT_t$ in equation (2.11)) reports the value-weighted monthly percentage return excluding dividends for all stocks in the CRSP universe. *UNEMP* reports the U.S. monthly unemployment rate obtained from the Bureau of Labor Statistics. *UEI* is the unexpected inflation (i.e., current month inflation minus the average of the past 12 realizations). *MP* is the monthly growth in industrial production obtained from the Federal Reserve. *RP* is the monthly default premium (i.e., difference between Moody's Baa-rated and Aaa-rated corporate bond yields) obtained from the Federal Reserve Bank of St. Louis. *TS* is the term spread (i.e., difference between the yields of a constant maturity 10-year Treasury bond and 3-month Treasury bill). *Split* is the monthly average splitting probability in percentage. *IO* is the annual total institutional ownership. *Stock* r*eturn* (i.e., $return_{i,\, t}$ in equation (2.10)) reports the stock-level monthly percentage return excluding dividends. *Ln (Size)* is the natural logarithm of market capitalization. *Std dev* reports the standard deviation. We also report the 25, 50, and 75 percentiles. N reports the number of observations. The sample period is from 2005 to 2013 for most variables.

Panel B: Other variables

Variables	(1) Mean	(2) Std dev	(3) 25th Pctl	(4) 50th Pctl	(5) 75th Pctl	(6) N
IPO return	13.44	9.68	8.40	12.50	18.00	107
Hotness	37.79	23.31	22.00	40.00	50.00	107
IPO number	2.21	0.68	1.61	2.40	2.71	107
Market return	0.85	3.80	-1.47	1.15	3.40	107
UNEMP	7.04	1.98	5.00	7.40	9.00	108
UEI	-0.01	0.47	-0.24	0.01	0.29	108
MP	0.07	0.81	-0.20	0.20	0.55	108
RP	1.19	0.55	0.90	0.98	1.29	108
TS	1.82	1.21	0.77	1.97	2.78	108
Split	0.12	0.14	0.03	0.07	0.19	108
IO	55.14	32.61	25.41	58.93	85.40	33,095
Stock return	0.54	16.40	-6.66	0.00	6.60	373,847
Ln (Size)	12.91	2.01	11.45	12.81	14.24	373,847

Table 2.3: Stock return predictability regression estimates

This table reports the predictive power of our Google gambling sentiment measure. We regress portfolio abnormal returns on the abnormal search volume intensity for the topic "lottery":

$$AR_{portfolio,\, t+n} = \alpha + \beta_n \times ASVI_t + \varepsilon_t,\ (n=0, 1, 2, 3).$$

ASVI is the abnormal search volume intensity based on the time-series difference in log search volume intensities (see equation (2.1)). We estimate the abnormal return of individual stocks by 36 months rolling window regressions. We use the Carhart (1997) four-factor model as benchmark. We then form value-weighted portfolios of lottery-like, non-lottery, and other stocks. Lottery-like stocks are defined as stocks within the upper 30 percentiles of Lottery Index (LIDX) in each year. Non-lottery stocks are defined as stocks in the bottom 30 percentiles of LIDX in each year. Other stocks are the rest of stocks in the CRSP universe. β_n measure the predictive power of *ASVI* with n lags. Column (1) indicates the *month n* (n=0, 1, 2, 3). Columns (2) to (4) report the regression coefficients on *ASVI* (β_n) for lottery-like, non-lottery, and other stock portfolios, respectively. Firms in the three portfolios are rebalanced in every January while portfolio weights are adjusted in every month according to market capitalization in the previous month. Column 5 reports the coefficient estimates of a portfolio strategy that goes long in lottery-like stocks and goes short in non-lottery stocks. The sample period is from January 2005 to December 2013. *N months* reports the number of months. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

(1) Months	(2) Lottery-like stocks	(3) Non-lottery stocks	(4) Other stocks	(5) Long-short portfolio
0	0.847	-0.140	0.416	0.987
	(0.925)	(0.145)	(0.453)	(0.961)
1	2.339**	-0.025	-0.586	2.365**
	(0.947)	(0.137)	(0.509)	(1.000)
2	-1.372	-0.003	0.308	-1.369
	(0.976)	(0.199)	(0.360)	(1.021)
3	-1.653**	0.199	-0.522*	-1.852**
	(0.787)	(0.136)	(0.307)	(0.827)
N months	108	108	108	108

Table 2.4: Stock performance around attention-grabbing jackpots

This table reports the average abnormal return for lottery-like and non-lottery stocks around attention-grabbing jackpots. We define attention-grabbing jackpots in three different ways. First, the *all-time largest* jackpot is the $656 million jackpot announced on March 30, 2012, which is the largest jackpot in the U.S. history. Second, *record-breaking* jackpots are the jackpots that break national record at the time. They include the all-time largest jackpot and two other jackpots announced on February 18, 2006 and March 6, 2007. Third, we use a broader definition of *attention-grabbing* jackpots. In addition to the above three record-breaking jackpots, we include five near-record jackpots that are either the second largest jackpot at the time or the largest jackpot over the past 24 months. We estimate the following regression:

$$r_{portfolio,t} - r_{f,t} = \alpha + \beta_1 D^{Jackpot}_{[-1,+1]} + \beta_2 D^{Post}_{[+2,+3]} + \beta_3(r_{mkt} - r_f) + \beta_4 r_{smb} + \beta_5 r_{hml} + \beta_6 r_{umd} + \epsilon_{t,}.$$

Dependent variable is the average excess return of lottery-like or non-lottery stocks in month *t*. The final drawing dates of attention-grabbing jackpots are in month 0. $D^{Jackpot}_{[-1,+1]}$ is a dummy variable that equals to one from month -1 to month 1 and zero otherwise. β_1 measures average monthly abnormal return for the (-1, +1) period. $D^{Post}_{[+2,+3]}$ is a dummy variable that equals to one for the (+2, +3) period and zero otherwise. β_2 measures average monthly abnormal return for the (+2, +3) period. Panel A (B) reports the estimated coefficients of β_1 (β_2). *Long-short portfolio* is a portfolio strategy that goes long in the lottery-like stock portfolio and goes short in the non-lottery stock portfolio. The sample period is from January 2005 to December 2013. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Average monthly abnormal return during months (-1, +1)

Jackpot type	Lottery-like stocks	Non-lottery stocks	Long-short portfolio
All-time largest	1.750***	0.047	1.704***
	(0.416)	(0.189)	(0.519)
Record-breaking	1.546**	-0.137	1.683***
	(0.597)	(0.122)	(0.605)
Attention-grabbing	1.694*	-0.164	1.858*
	(0.865)	(0.139)	(0.941)

Panel B: Average monthly abnormal return during months (+2, +3)

Jackpot type	Lottery-like stocks	Non-lottery stocks	Long-short portfolio
All-time largest	-1.225*	-0.090	-1.135*
	(0.678)	(0.106)	(0.638)
Record-breaking	-1.100*	0.114	-1.214**
	(0.610)	(0.199)	(0.578)
Attention-grabbing	-0.395	0.236	-0.630
	(0.752)	(0.147)	(0.805)

Table 2.5: Trading volume around attention-grabbing jackpots

This table reports the abnormal trading volume of lottery-like and non-lottery stocks around attention-grabbing jackpots. We define attention-grabbing jackpots in three different ways. First, the *all-time largest* jackpot is the $656 million jackpot announced on March 30, 2012, which is the largest jackpot in the U.S. history. Second, *record-breaking* jackpots are the jackpots that break national record at the time. They include the all-time largest jackpot and two other jackpots announced on February 18, 2006 and March 6, 2007. Third, we use a broader definition of *attention-grabbing* jackpots. In addition to the above three record-breaking jackpots, we include five near-record jackpots that are either the second largest jackpot at the time or the largest jackpot over the past 24 months. Abnormal trading volume is estimated as the difference between log-transformed turnover (the total number of shares traded divided by shares outstanding) in month *t* and the average log-transformed turnover in the estimation period. The estimation period has a length of 36 months and ends three months before the event month. Panel A reports the average abnormal trading volume in months (-1, +1) for the all-time largest, recording-breaking, and attention-grabbing jackpots. Panel B reports the average abnormal trading volume in months (+2, +3). The sample period is from January 2005 to December 2013. Standard errors (reported in parentheses) are clustered by events and by firms. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Average abnormal trading volume during months (-1, +1)

Jackpot type	Lottery-like stocks	Non-lottery stocks	Difference
All-time largest	1.556	-15.338***	16.893***
	(1.604)	(0.668)	(1.738)
Record-breaking	14.040**	7.046	6.994
	(6.807)	(10.039)	(15.709)
Attention-grabbing	17.170***	-4.602	21.772**
	(5.987)	(6.079)	(8.532)

Panel B: Average abnormal trading volume during months (+2, +3)

Jackpot type	Lottery-like stocks	Non-lottery stocks	Difference
All-time largest	1.536	-1.765*	3.301
	(2.166)	(0.934)	(2.359)
Record-breaking	18.086**	19.186*	-1.100
	(7.909)	(10.694)	(26.367)
Attention-grabbing	20.975***	-0.244	21.219**
	(5.930)	(7.927)	(9.900)

Table 2.6: Aggregate demand for lottery-like stocks around large jackpots

This table reports the excess buy-sell imbalance (*EBSI*) of lottery-like stocks around the largest Powerball jackpot announced on July 7, 1993. We run the following time-series regression:

$$EBSI_t = \alpha + \beta_1 D^{Jackpot}_{[-1,+1]} + \beta_2 MKTRET_t + \beta_3 MKTRET_{t-1} + \beta_4 LOTRET_t + \beta_5 LOTRET_{t-1} + \beta_6 EBSI_{t-1} + Controls + \epsilon_t.$$

$EBSI_t$ is the month t difference in buy-sell imbalance between lottery-like and non-lottery stocks. $D^{Jackpot}_{[-1,+1]}$ is a dummy variable that equals one from month -1 to month 1 around the largest jackpot, and zero otherwise. Other independent variables include contemporaneous and one month lagged market returns ($MKTRET_t$, $MKTRET_{t-1}$), contemporaneous and one month lagged returns of the lottery-like stock portfolio ($LOTRET_t$, $LOTRET_{t-1}$), and *EBSI* in the previous month. Control variables include five macroeconomic variables: U.S. monthly unemployment rate (*UNEMP*), unexpected inflation (*UEI*), monthly growth in industrial production (*MP*), monthly default risk premium (*RP*), and the term spread (*TS*). The sample period is from April 1992 to November 1996. *N months* reports the number of months. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
$D_{Jackpot}$	9.476***	9.253***	10.473***	9.343**	7.059*
	(3.159)	(3.183)	(3.887)	(4.135)	(4.045)
$MKTRET_t$	0.863**	0.845**	-1.177	-0.922	-1.107
	(0.410)	(0.388)	(1.264)	(1.349)	(1.234)
$MKTRET_{t-1}$		0.144	-0.261	-1.182	-1.526
		(0.430)	(0.450)	(1.215)	(1.278)
$LOTRET_t$			1.417	1.307	1.549*
			(0.871)	(0.911)	(0.845)
$LOTRET_{t-1}$				0.667	0.720
				(0.716)	(0.777)
$EBSI_{t-1}$					0.257**
					(0.115)
Constant	30.370**	31.153**	32.223**	33.576**	27.603*
	(13.657)	(14.122)	(15.171)	(16.221)	(14.766)
Controls	Yes	Yes	Yes	Yes	Yes
N months	56	56	56	56	56
Adjusted R^2	0.167	0.152	0.201	0.200	0.227

Table 2.7: Aggregate demand for lottery-like stocks following large drawings

This table reports the daily buy-sell imbalance (*EBSI*) of lottery-like stocks following large Powerball drawings. We run the following time-series regression:

$$EBSI_t = \alpha + \beta_1 D_t^{Drawing} + \beta_2 MKTRET_t + \beta_3 LOTRET_t + \beta_4 EBSI_{t-1} + \beta_5 VIX_{t-1} + \beta_6 ADS_{t-1} + Controls + \epsilon_t.$$

$EBSI_t$ is the day *t* difference in buy-sell imbalance between lottery-like and non-lottery stocks. $D_t^{Drawing}$ is a dummy variable that equals to one on the next trading day following a large drawing, and zero on days with no drawings or small drawings. A large (small) drawing has above (below) median drawing value during the sample period. Other independent variables include market return, return of the lottery-like stock portfolio, and lagged *EBSI*. VIX_{t-1} is the lagged Chicago Board Options Exchange (CBOE) daily market volatility index. ADS_{t-1} is the lagged Aruoba-Diebold-Scotti business conditions index. Control variables include lagged market and lottery-like stock portfolio returns (up to five lags) and day-of-the-week dummies. The sample period is from April 22, 1992 to November 30, 1996. *N days* reports the number of trading days. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
$D_{Drawing}$	3.423*	3.293*	3.178*	3.170*	3.080*
	(1.785)	(1.783)	(1.716)	(1.721)	(1.693)
$MKTRET_t$	3.114***	6.856**	6.047**	6.061**	5.673**
	(1.189)	(2.740)	(2.800)	(2.804)	(2.769)
$LOTRET_t$		-3.273	-2.769	-2.780	-2.570
		(2.064)	(2.131)	(2.140)	(2.119)
$EBSI_{t-1}$			0.179***	0.179***	0.157***
			(0.040)	(0.040)	(0.040)
VIX_{t-1}				-0.054	0.287
				(0.383)	(0.410)
ADS_{t-1}					-6.737***
					(1.719)
Constant	2.260*	2.519*	1.963	2.744	-0.800
	(1.316)	(1.338)	(1.287)	(5.793)	(5.976)
Controls	Yes	Yes	Yes	Yes	Yes
N days	1,167	1,167	1,167	1,167	1,167
Adjusted R^2	0.021	0.022	0.053	0.052	0.068

Table 2.8: Stock performance among U.S. states sorted by gambling sentiment

This table reports the performance of a value-weighted portfolio of lottery-like or non-lottery stocks. Abnormal return is measured as the intercept of monthly return regressions by using the Carhart (1997) four factor model as benchmark. *Full sample* reports the abnormal portfolio returns for all stocks in our sample. *Strong (moderate) (weak) sentiment* reports the abnormal portfolio returns of stocks headquartered in U.S. states with strong (moderate) (weak) gambling sentiment. *Strong-weak* (*strong-moderate*) measures the abnormal return difference between stocks located in states with strong and weak (moderate) gambling sentiment. Strong (moderate) (weak) gambling sentiment state group includes 17 states with top (medium) (bottom) average search volume intensity for the topic "lottery". The three groups of states are updated in January of each year. *Long-short portfolio* is a portfolio strategy that goes long in the lottery-like stock portfolio and goes short in the non-lottery stock portfolio. The sample period is from January 2005 to December 2013. *N months* reports the number of months. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) Lottery-like stocks	(2) Non-lottery stocks	(3) Long-short portfolio
Full sample	-0.422*	0.010	-0.432*
	(0.222)	(0.033)	(0.242)
Strong sentiment	-0.542**	0.060	-0.602**
	(0.225)	(0.054)	(0.253)
Moderate sentiment	-0.143	-0.045	-0.098
	(0.284)	(0.080)	(0.285)
Weak sentiment	0.146	-0.056	0.202
	(0.447)	(0.151)	(0.477)
Strong – weak	-0.687*	0.116	-0.803*
	(0.394)	(0.159)	(0.427)
Strong – moderate	-0.399	0.105	-0.504**
	(0.251)	(0.113)	(0.240)
N months	108	108	108

Table 2.9: Performance of stocks headquartered in U.S. states with strong gambling sentiment

This table reports the performance of a value-weighted portfolio of stocks located in U.S. states with strong gambling sentiment. Abnormal return is measured as the intercept of monthly return regressions by using the Carhart (1997) four-factor model as benchmark. Panel A reports the long-term performance of stocks with different levels of institutional ownership (*IO*). *Low (high) IO* is the abnormal return of a value-weighted portfolio of lottery-like or non-lottery stocks with less (more) than ten percent institutional ownership. Panel B reports the long-term performance of stocks with different market capitalizations. *Small (large)* is the abnormal return of stocks in the bottom (top) 30% by size. *Low- high (Small- large)* reports the abnormal return difference between the same types of stocks with different levels of institutional ownership (market capitalizations). *Long-short portfolio* reports the abnormal return earned by a portfolio strategy that goes long in lottery-like stocks and goes short in non-lottery stocks. The sample period is from January 2005 to December 2013. *N months* reports the number of months. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Stocks sorted by institutional ownership

	(1) Lottery-like stocks	(2) Non-lottery stocks	(3) Long-short portfolio
Low IO	-1.237***	0.112	-1.349***
	(0.334)	(0.405)	(0.503)
High IO	-0.465**	0.064	-0.529**
	(0.223)	(0.056)	(0.250)
Low – high	-0.772**	0.048	-0.821*
	(0.331)	(0.406)	(0.455)
N months	108	108	108

Panel B: Stocks sorted by firm size

	(1) Lottery-like stocks	(2) Non-lottery stocks	(3) Long-short portfolio
Small	-3.427***	-1.989***	-1.438***
	(0.384)	(0.443)	(0.390)
Large	-0.050	0.062	-0.111
	(0.248)	(0.054)	(0.272)
Small - large	-3.377***	-2.050***	-1.327***
	(0.451)	(0.456)	(0.422)
N months	108	108	108

Table 2.10: Gambling sentiment and stock splits

This table reports the results of our logistic estimate. We run the following regressions:

$$Logit\ (Split_{i,t}) = \alpha + \beta_1 D_{ASVIt-1} + \beta_2\ Dp_{i,t-1} + \beta_3 D_{ASVIt-1} \times Dp_{i,t-1} + \beta_4 Return_{i,t} + \beta_5 Size_{i,t-1} + \beta_6 MTB_{i,t-1} + \beta_7 Spllitter_{i,t-12} + \varepsilon_{t.}.$$

Dependent variable is equal to one if a company splits its shares in a given month. Independent variables include a dummy variable of the abnormal search volume intensity for the topic" Lottery" ($D_{ASVIt-1}$), a dummy variable of stock prices ($Dp_{i,t-1}$), and their interaction term ($D_{ASVIt-1} \times Dp_{i,t-1}$). We use 75th percentile as the break points for both dummies. $D_{ASVIt-1}$ is equal to one if it has a value above the 75th percentile of the time-series. Similarly, D_{pt-1} is equal to one if a firms' price is above the 75th percentile of all stock in the CRSP universe in a given month. Control variables include size ($Size_{i,t-1}$) and market-to-book ratio ($MTB_{i,t-1}$) at the beginning of the month and return ($Return_{i,t}$) over the course of the month. $Size_{i,t-1}$ is the natural logarithm of the market capitalization of stock *i* in month *t-1* while $MTB_{i,t-1}$ is defined as the market value of the firm over its book value. Market value equals to market equity at calendar year-end plus book debt while book value is calculated as stockholders' equity minus preferred stock plus deferred taxes and investment tax credits and post retirement assets. $Splitter_{i,t-12}$ is equal to one if a firm splits its share in the previous year. The sample period is from January 2005 to December 2013. *Observations* is the number of firm-month observations. Standard errors (reported in parentheses) are clustered by firm and by time. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Variables	(1)	(2)	(3)	(4)	(5)
$D_{ASVIt-1} \times Dp_{t-1}$	1.118**	1.116**	1.116**	1.119**	1.120**
	(0.514)	(0.517)	(0.518)	(0.520)	(0.520)
$D_{ASVIt-1}$	-0.927	-0.930	-0.928	-0.932	-0.932
	(0.595)	(0.597)	(0.597)	(0.601)	(0.601)
Dp_{t-1}	3.713***	3.730***	3.905***	3.926***	3.928***
	(0.224)	(0.228)	(0.239)	(0.244)	(0.244)
$Return_t$		0.006***	0.006***	0.006***	0.006***
		(0.001)	(0.001)	(0.001)	(0.001)
$Size_{t-1}$			-0.066**	-0.067**	-0.063**
			(0.031)	(0.031)	(0.031)
MTB_{t-1}				0.016***	0.016***
				(0.003)	(0.003)
$Splitter_{t-12}$					-0.494
					(0.410)
Constant	-9.109***	-9.129***	-8.334***	-8.375***	-8.415***
	(0.258)	(0.262)	(0.480)	(0.483)	(0.484)
Observations	373,910	373,847	373,847	373,847	373,847
Pseudo R^2	0.143	0.143	0.144	0.145	0.145

Table 2.11: Gambling sentiment and first-day IPO returns

This table reports the underpricing of initial public offerings (IPOs). We run the following regressions:

$$r_t^{IPO} = \alpha + \beta_1 ASVI_{t-1} + \beta_2 p_{t-1} + \beta_3 VWMKT_t + \beta_3 Hotness_t + \beta_4 IPOnumber_t + \varepsilon_{t}.$$

The dependent variable is the monthly average first-day return on the net IPOs obtained from Jay Ritter's website. Net IPOs are IPOs excluding closed-end funds, REITs, acquisition companies, stocks with offer prices below $5, ADRs, limited partnerships, units, banks and S&Ls, and those not listed on CRSP. First-day return is calculated as the percentage return from the offering price to the first closing bid price. The monthly average first-day return is calculated as the equal-weighted average of the first-day returns on all the offerings in a particular calendar month. Independent variables are the abnormal search volume intensity for the topic "lottery" in the previous month ($ASVI_{t-1}$). Following Baker, Greenwood, and Wurgler (2009), we also include the average log price in the previous month (p_{t-1}) and the value-weighted return excluding dividends of all common stocks in the CRSP universe ($VWMKT_t$) as control variables. In addition, we control for the hotness of IPO market ($Hotness_t$, i.e., the percentage of deals that priced above the midpoint of the original file price range) and the natural logarithm of monthly number of net IPOs ($IPOnumber_t$). The sample period is from January 2005 to June 2014. *N months* reports the number of months. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Variables	(1)	(2)	(3)	(4)	(5)
$ASVI_{t-1}$	6.445**	7.022**	7.233**	8.402**	8.093***
	(3.163)	(3.417)	(3.406)	(3.367)	(3.090)
$VWMKT_t$		0.516**	0.487**	0.405**	0.414**
		(0.222)	(0.221)	(0.193)	(0.190)
p_{t-1}			13.167***	14.310***	14.978***
			(4.335)	(2.807)	(2.781)
$Hotness_t$				0.166***	0.168***
				(0.041)	(0.040)
$IPOnumber_t$					-0.726
					(1.766)
Constant	13.432***	12.993***	-38.949**	-49.672***	-50.776***
	(1.290)	(1.307)	(17.384)	(11.448)	(10.672)
N months	107	107	107	107	107
Adjusted R^2	0.010	0.042	0.125	0.281	0.276

Table 2.12: Robustness checks

This table reports results for various robustness tests. Panels A and B consider the robustness with respect to the predictive power of our gambling sentiment measure. The dependent variables are the contemporaneous and future abnormal returns of a long-short portfolio. In Panel A, Columns 1 to 5 include U.S. monthly unemployment rate (*UNEMP*), unexpected inflation (*UEI*, i.e., current month inflation minus the average of the past 12 realizations), monthly growth in industrial production (*MP*), monthly default risk premium (*RP*, i.e., difference between Moody's Baa-rated and Aaa-rated corporate bond yields), or term spread (*TS*, i.e., difference between the yields of a constant maturity 10-year Treasury bond and 3-month Treasury bill) respectively as macroeconomic control. *ALL* (Column 6) reports the estimates by including all the five macroeconomic controls. Panel B considers subsets of data. Columns 1 and 2 consider a subsample from 2005 to 2010 when data on monthly investor sentiment index are available (Baker and Wurgler, 2007). Columns 3 and 4 consider the sample after June 2006, when Google's search volume intensity data become publicly available. We use Column 5 of Table 2.3 as the baseline specification. Panel C considers the robustness of our results for stock splits. Dependent variable is equal to one if the company splits its shares in a given month. Column 1 includes firm-level fixed effects. Columns 2 to 7 include the five macroeconomic variables as control variables. We use Column 5 of Table 2.8 as the baseline specification. Panel D considers the robustness of results related to IPO first-day return. The dependent variable is the monthly average first-day return on the net IPOs obtained from Jay Ritter's website. Columns 1 to 6 include macroeconomic controls. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method (Panels A, B, and D) or clustered by firm and by time (Panel C). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Return predictability with macroeconomic controls

Months	(1) UNEMP	(2) UEI	(3) MP	(4) RP	(5) TS	(6) ALL
0	0.985	1.008	1.035	1.070	0.991	1.084
	(0.962)	(0.996)	(0.969)	(0.974)	(0.964)	(1.026)
1	2.364**	2.368**	2.440**	2.493**	2.357**	2.506**
	(1.005)	(1.021)	(0.965)	(1.029)	(1.006)	(1.038)
2	-1.369	-1.369	-1.269	-1.264	-1.372	-1.223
	(1.027)	(1.022)	(1.115)	(1.017)	(1.024)	(1.110)
3	-1.851**	-1.864**	-1.800**	-1.755**	-1.852**	-1.739*
	(0.831)	(0.847)	(0.896)	(0.829)	(0.832)	(0.878)
N months	108	108	108	108	108	108

Table 2.12 (Cont'd)

Panel B: Return predictability with subsamples and investor sentiment control

Months	(1) 2005-2010	(2) 2005-2010	(3) After June 06	(4) After June 06
0	1.234	1.235	0.830	0.871
	(1.830)	(1.829)	(1.060)	(2.500)
1	3.437***	3.433***	2.413**	4.127***
	(1.215)	(1.246)	(1.138)	(1.521)
2	-1.836	-1.846	-1.013	-1.088
	(1.522)	(1.540)	(1.104)	(2.028)
3	-2.413	-2.420	-2.315**	-3.876**
	(1.484)	(1.455)	(0.933)	(1.804)
Sentiment Control	NO	YES	NO	YES
N months	72	72	91	55

Panel C: Stock split with firm-level fixed effects and macroeconomic controls

Variables	(1) Firm FE	(2) UMEMP	(3) UEI	(4) MP	(5) RP	(6) TS	(7) ALL
$D_{ASVIt-1} \times Dp_{t-1}$	1.248**	1.115**	1.121**	1.121**	1.116**	1.117**	1.114**
	(0.634)	(0.519)	(0.520)	(0.520)	(0.519)	(0.519)	(0.518)
Controls	YES	YES	YES	YES	YES	YES	YES
Observations	38,100	373,847	373,847	373,847	373,847	373,847	373,847
Pseudo R^2	0.186	0.167	0.146	0.146	0.171	0.161	0.180

Panel D: IPO first-day return with macroeconomic controls

Variables	(1) UMEMP	(2) UEI	(3) MP	(4) RP	(5) TS	(6) ALL
$ASVI_{t-1}$	7.982***	8.066**	7.900**	8.000***	7.765***	7.561**
	(3.000)	(3.133)	(3.113)	(2.952)	(2.903)	(3.040)
Controls	YES	YES	YES	YES	YES	YES
N months	107	107	107	107	107	107
Adjusted R^2	0.276	0.270	0.271	0.269	0.288	0.270

Figure 2.1: Search volume intensity for "lottery"

This figure plots the time-series of the search volume intensity (*SVI*) for the topic "lottery" at the national level from January 2005 to December 2013. Points A to H correspond to the eight attention-grabbing jackpots. We define attention-grabbing jackpots in three different ways. First, the *all-time largest* jackpot is the $656 million jackpot announced on March 30, 2012, which is the largest jackpot in the U.S. history. Second, *record-breaking* jackpots are the jackpots that break national record at the time. They include the all-time largest jackpot and two other jackpots announced on February 18, 2006 and March 6, 2007. Third, we use a broader definition of *attention-grabbing* jackpots. In addition to the above three record-breaking jackpots, we include near-record jackpots that are either the second largest jackpot at the time or the largest jackpot over the past 24 months. Prize and date of the eight jackpots are reported in Appendix Table 2.1A. Source: Google Trends.

Panel A: National-level search volume intensity for "lottery"

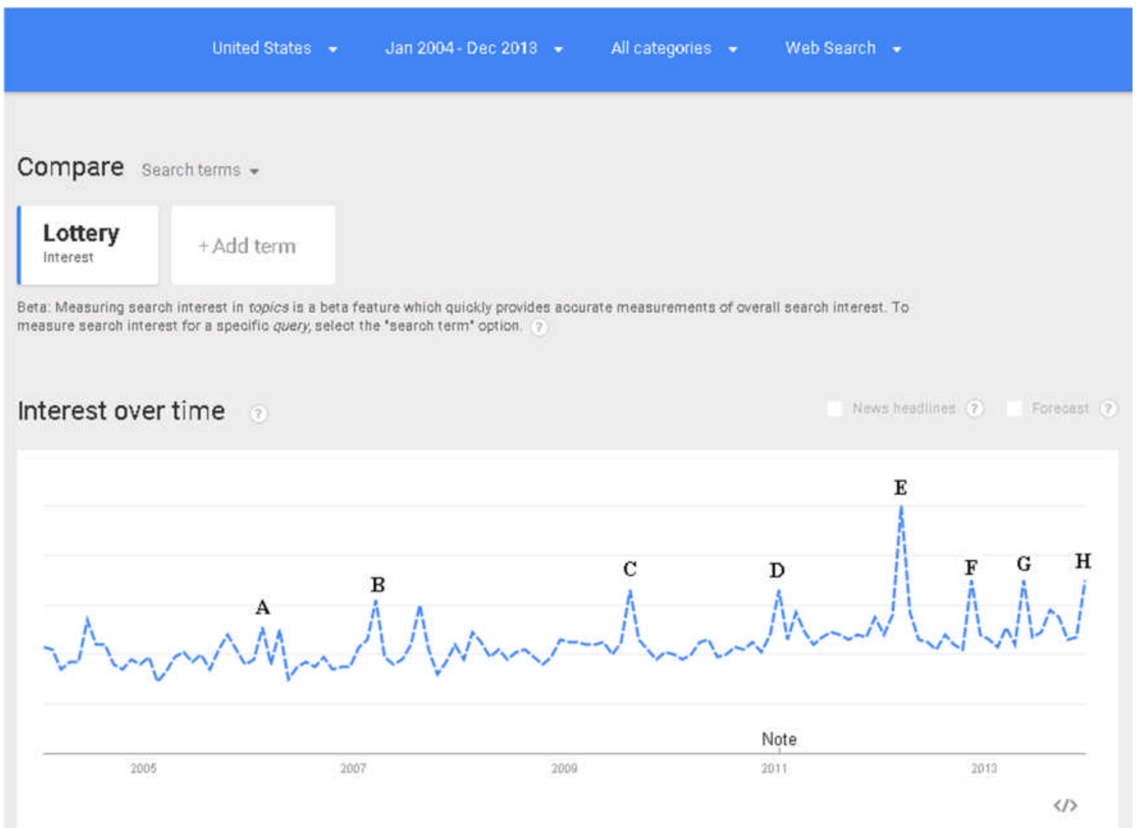


Figure 2.1 (Cont'd)

This panel plots the search volume intensity (*SVI*) for the topic "lottery" for three U.S. states: Florida, Utah, and Texas. Points A and B correspond to jackpots of single state lotto games, while points C and D correspond to jackpots of multi-state lotto games. Source: Google Trends.

Panel B: State-level search volume intensity for "lottery"

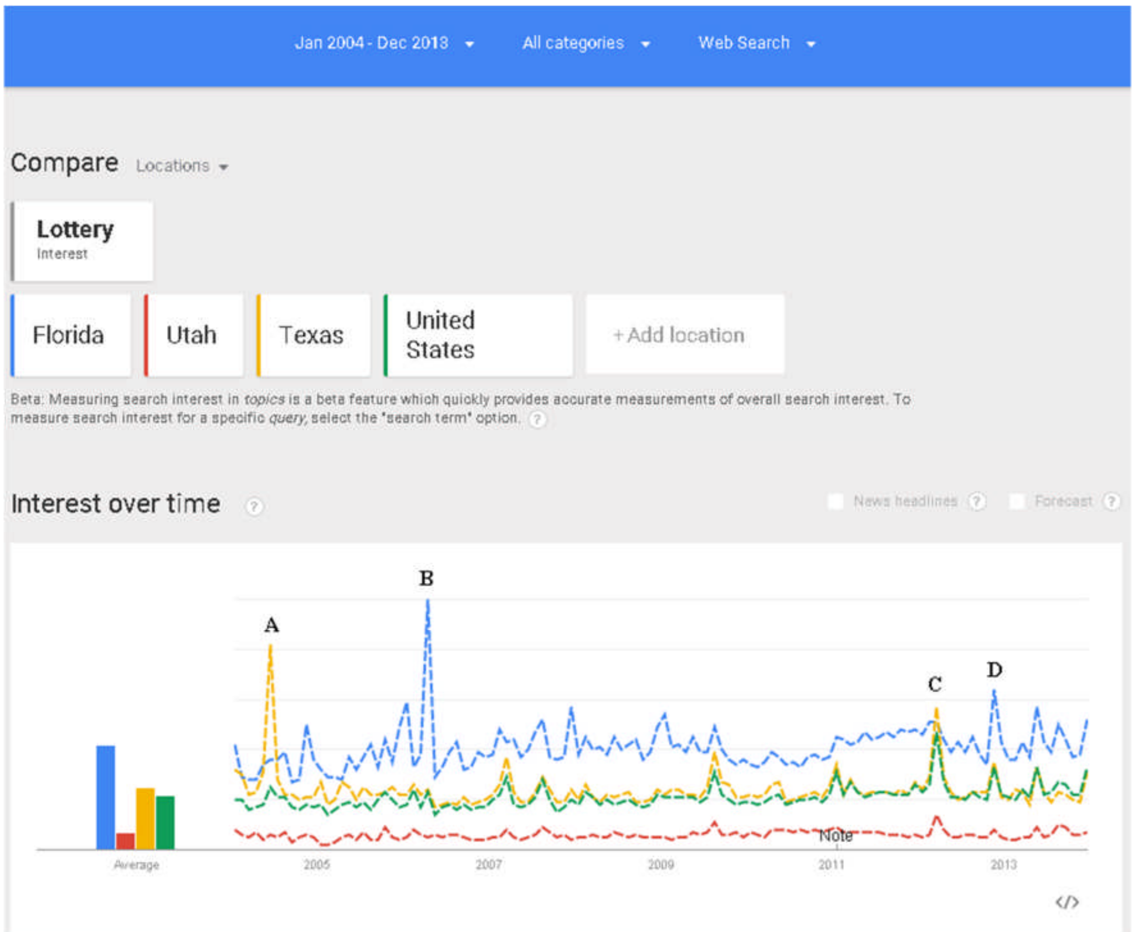
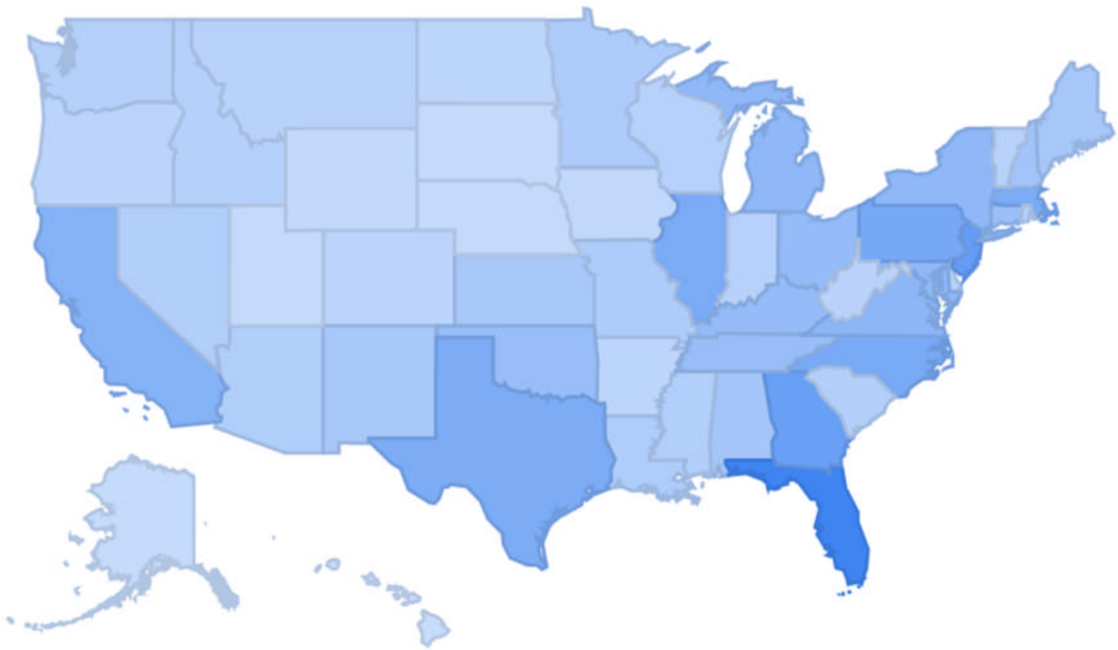


Figure 2.1 (Cont'd)

This panel shows the geographical distribution of the search volume intensity (*SVI*) for the topic "lottery". Darker color indicates stronger search volume intensity. The intensity is calculated based on the average SVI during the 2004-2013 period. Source: Google Trends.

Panel C: Geographical distribution of search volume intensity for "lottery"



Appendix

Table 2.1A: Attention-grabbing jackpots

This table provides details about the eight attention-grabbing jackpots in our sample. *ID* corresponds to data points shown in Panel A of Figure 2.1. *Jackpot date* is the final drawing day of the jackpot. *First date* is the first drawing day of the jackpot. *Value* is the prize of winning the jackpot in million dollars. *Game* is the corresponding lotto game of a jackpot. *Note* indicates whether the jackpot is the all-time largest, record-breaking, or attention-grabbing. We define the attention-grabbing jackpots in three different ways. First, the *all-time largest* jackpot is the $656 million jackpot announced on March 30, 2012, which is the largest jackpot in the U.S. history. Second, *record-breaking* jackpots are the jackpots that break national record at the time. They include the *all-time largest* jackpot and two other jackpots announced on February 18, 2006 and March 6, 2007. Third, we use a broader definition of *attention-grabbing* jackpots. In addition to the above three record-breaking jackpots, we include five near-record jackpots that are either the second largest jackpot at the time or the largest jackpot over the past 24 months. Source: Mega Millions, Powerball.

ID	*Jackpot date*	*First date*	*Value ($ m)*	*Game*	*Note*
A	Feb 18, 2006	Dec 17, 2005	365	Powerball	Record-breaking
B	Mar 6, 2007	Jan 12, 2007	390	Mega Millions	Record-breaking
C	Aug 28, 2009	Jul 10, 2009	336	Mega Millions	Near-record
D	Jan 4, 2011	Nov 12, 2010	380	Mega Millions	Near-record
E	Mar 30, 2012	Jan 27, 2012	656	Mega Millions	All-time largest
F	Nov 28, 2012	Oct 6, 2012	588	Powerball	Near-record
G	May 18, 2013	Apr 3, 2013	591	Powerball	Near-record
H	Dec 17, 2013	Oct 4, 2013	648	Mega Millions	Near-record

Chapter 3

Social Sensitivity and Stock Returns

3.1. Introduction

Corporate social responsibility (CSR) has attracted increasing attention among academics as well as practitioners. According to the U.S. Forum for Sustainable and Responsible Investment (USSIF), the total U.S. domiciled assets under management using socially responsible investing (SRI) strategies have reached $6.57 trillion in 2014.[29] In addition, survey evidence also suggests that investors care about social attributes. According to Kitzmueller and Shimshack (2012), 52% of respondents are actively seeking firm-level CSR information. Consequently, investors are likely to update their beliefs about firm-level social attributes regularly, which in turn could affect their investment choices.

In this paper, we propose a novel method for identifying stocks that are likely to be influenced by investors' recent perception on stock-level social attributes. Specifically, we estimate the return sensitivity of industries and firms with respect to Google search volume of CSR-related keywords on an 18-month rolling basis. These social sensitivity estimates capture the social attributes perceived by the market in recent periods. In particular, according to our social sensitivity estimate, investors perceive a firm or an industry as having good social attributes if it earns higher average returns during periods with high CSR search volume in the recent past. We find that this identification strategy is consistent with exiting low frequency CSR ratings and investors' preconception on industry-level CSR records, but provides a higher frequency (i.e., monthly) time-varying measure for investors' perception about social attributes.

When investors update their beliefs about social attributes, they are likely to rebalance their portfolios and tilt investments toward firms with good perceived social attributes. This trading behavior could generate predictable return patterns among firms with greater

[29] Source: USSIF 2014 Trends Report available at: http://www.ussif.org/Files/Publications/SIF_Trends_14.F.ES.pdf.

social sensitivity. Specifically, we conjecture that there will be predictable return patterns of socially sensitive firms and industries that can be identified *ex-ante*. Industries and stocks with good perceived social attributes are likely to be underpriced so these stocks are likely to have better future return.

With our social sensitivity estimates, we find that returns of stocks and industries with greater return sensitivity to investors' CSR-related searches are predictable. In particular, a trading strategy that goes long in a value-weighted industry portfolio with the most positive social sensitivity and goes short in a value-weighted industry portfolio with the most negative social sensitivity generates a characteristic-adjusted return of 1.17% per month from 2005 to 2013, or 14.02% per year. This return predictability covers 15-36% of total market capitalization, which is economically meaningful.

In addition, our results remain robust at the stock level. Specifically, a value-weighted portfolio of stocks with the most positive social sensitivity earns a seven-factor adjusted alpha of 9.5% per year. Further, our results are robust to various unconditional and conditional factor models as well as three alternative social sensitivity measures.

Beyond these return-based tests, we further investigate the potential economic channel for our findings. In particular, we examine whether perceived social attributes affect the investment choices of institutional investors using institutional trading data. This test is motivated by the evidence in recent studies which shows that institutional investors are more norm-constrained and care more about CSR (e.g., Hong and Kacperczyk, 2009; Kumar, Page, and Spalt, 2011) and have better trading skills (Puckett and Yan, 2011). In addition, our double sorted portfolio results also suggest that the predictability results are stronger for firms that are more likely to be held by institutional investors (i.e., large firms and firms with high institutional ownership). If institutional investors frequently update their portfolio holdings to include stocks with good social attributes (i.e., which are underpriced in the recent period), their trading activity could help to correct the underpricing of stocks with good perceived social attributes. Consistent with our hypothesis, we find that institutional investors have 1.3% higher net demand per month on firms with good perceived social attributes in the recent period.

In the last set of tests, we examine the longevity of our return predictability results. We find that the predictive power of our social sensitivity estimates declines as we increase the gap between social sensitivity estimation month and portfolio formation month. This finding further supports our conjecture that investors update their perceptions about social attributes and rebalance their investment frequently. Since institutional investors update their portfolio holdings frequently and correct the underpricing of stocks with good perceived social attributes, the alpha estimate of a long short strategy that based on stale information becomes statistically insignificant after 3 months. This evidence suggests that the underpricing of perceived social attributes is likely to be corrected in about 3 months.

Taken together, our findings contribute to several strands of literature. First, we contribute to the emerging literature on corporate social responsibility. As suggested by Liang and Renneboog (2016), the existing literature has typically focused on one perspective on CSR such as Employee satisfaction and workplace safety (e.g., Edmans, 2011; Cohn and Wardlaw, 2016; Li, Edmans, and Zhang, 2016), environmental protection (e.g., Dowell, Hart, and Yeung, 2000; Konar and Cohen, 2001; Chava, 2014), corporate philanthropy (e.g., Masulis and Reza, 2015), customer satisfaction (e.g., Luo and Bhattacharya, 2006; Servaes and Tamayo, 2013), or corporate governance (e.g., Dimson, Karakas, and Li, 2015; Cheng, Hong, and Shue, 2016; Ferrell, Liang, and Renneboog, 2016). We extend the previous literature by focusing on investors' perception about firm-level social attributes. Our key innovation is to identify firms and industries that are likely to be perceived as having good social attributes by the market. In addition, we identify a profitable SRI trading strategy based on social sensitivity.

Beyond the literature on CSR, our paper also contributes to the literature on return predictability. For example, Cohen and Frazzini (2008) show that consumer-supplier links can be used to identify predictable return patterns. In addition, Korniotis and Kumar (2013) find that local economic conditions predict local stocks returns. Similarly, Addoum and Kumar (2016) show that political sensitivity could also be used to identify predictable patterns in stock returns. Our paper provides evidence of return predictability along a new dimension, i.e., social responsibility dimension.

More broadly, we contribute to the literature on the value of intangible assets. The previous literature suggests that firms with high research and development (Eberhart, Maxwell, and Siddique, 2004), advertisement (Chan, Lakonishok, and Sougiannis, 2001), patent citations (Hirshleifer, Hsu, and Li, 2013), and software development costs (Aboody and Lev, 1998) are likely to have better future returns. We show that good perceived social attributes, which could be treated as a type of intangible assets, could also lead to better stock returns.

This paper is organized as follows. Section 3.2 presents the literature review and hypothesis development. Section 3.3 discusses the data and methodology. Section 3.4 presents the empirical results. Section 3.5 concludes.

3.2. Literature review and hypothesis development

3.2.1. Three views on the corporate social responsibility

The existing literature on corporate social responsibility mainly examines the relation between corporate goodness and firm value. Theoretical work, as reviewed by Benabou and Tirole (2010), has three different views for the impact of CSR on firm value. In particular, the first view posits a win-win relation (i.e., "doing well by doing good"). This view suggests that CSR is about taking a long-term perspective to maximize intertemporal profits and correct short-termism of managers. Therefore, firms that are more socially responsible should have better financial performance. Empirically, consistent with this prediction, Edmans (2011) shows that employee satisfaction is positively associated with firm value. A value-weighted portfolio of the '100 best companies to work for in America' earned an annual four-factor adjusted alpha of 3.5% during the 1984 – 2009 period. In addition, Deng, Kang, and Low (2013) find that acquirers with good social records realize higher merger announcement returns. These mergers take less time and are less likely to fail. Further, Dimson, Karakas, and Li (2015) find that successful shareholder engagements for addressing environmental, social, and governance concerns are followed by positive abnormal returns.

Similar to the first view, the second view conjectures that stakeholders delegate firms to do good on their behalf. The costs of socially responsible behavior are then pass through to stakeholders at their demand. Empirically, among others, Hong and Kacperczyk (2009) find that although sin stocks (i.e., publicly traded firms involved in producing alcohol, tobacco, and gambling) have higher expected returns than stocks that are otherwise comparable, they are less held by norm-constrained investors, and receive less analysis coverages. Further, the second view also suggest that CSR is consistent with profit maximization in the long-term. For example, El Ghoul, Guedhami, Kwok, and Mishra (2011) find that firms with better CSR scores has cheaper equity financing while firms involving in sin industries have substantially higher cost of equity.

In contrast to the first two views, the third view perceives CSR as an insider initiated corporate philanthropy. In this scenario, CSR activities are value destroying. Among others, Hong, Kubik, and Scheinkman (2012) find that firms will only do good when they have free cash flows. In addition, Cheng, Hong, and Shue (2016) and Masulis and Reza (2015) show that increased insider ownership or corporate governance reduces corporate goodness. Further, Krüger (2015) show that investors respond negatively to positive CSR news, which suggests that investors perceive goodness spending as an agency problem.

Overall, the effect of CSR on firm value remains inconclusive. In a survey of all CSR literature during the 1972 – 2007 period, Margolis, Elfenbein, and Walsh (2009) find that most studies report a non-significant relation while the overall effect of CSR on firm value is positive but small. In addition, a review by Renneboog, Ter Horst, and Zhang (2008) concludes that existing studies hint but do not clearly demonstrate that funds with CSR screenings underperform conventional funds.

3.2.2. Main hypothesis

In this paper, we examine the relation between CSR and firm value from a new perspective: perceived social attribute. The existing literature suggests that retail-based industries are commonly perceived as socially responsible. These industry firms put a lot of efforts in image building since a socially responsible corporate image could boost consumer demand, generate customer loyalty, support premium pricing, and serve as an alternative way to assure product quality (e.g., Besley and Ghatak, 2007; Castaldo, Perrini, Misani, and Tencati, 2009; Elfenbein, Fisman, and Mcmanus, 2012; and Albuquerque, Durnev, and Koskinen, 2016).

In contrast, some industries have controversial business and are commonly viewed as socially irresponsible. For example, industries involving in fossil fuel (i.e., coal, oil, and natural gas) and other natural resources (e.g., mining, precious metal) are commonly screened by socially responsible funds. In addition, the Forum for Sustainable and Responsible Investment (i.e., USSIF) also encourages all types of investors to divest from these industries to address climate changes risks. Motivated by the above evidence, we conjecture that industry and stock level social attributes could influence the investment preferences of investors.

Since investors frequently update their beliefs on firm level CSR reputation, they are likely to alert their investments based on an industry's perceived social attributes in the recent past, which in turn could impact stock prices. To summarize, the key hypothesis of this paper is that perceived social attributes affect stock returns. Firms with good perceived social attributes in the recent past are likely to have better future returns.

3.3. Data and methodology

3.3.1. Main datasets

We collect data from various sources. We obtain market excess return (*MKTRF*), the size factor (*SMB*), the value factor (*HML*), the momentum factor (*UMD*), the short term reversal (*STR*) and long-term reversal (*LTR*) factors, and value-weighted returns of the 48 Fama and French (1997) industry portfolios at daily and monthly frequencies from Kenneth French's website.[30] We obtain the liquidity factor (*LIQ*) from Lubos Pastor's website,[31] the Lettau and Ludvigson (2001) *cay* measure from Sydney Ludvigson's website, U.S. business cycle data from National Bureau of Economic Research (NBER), and Business Wire CSR newsletters from Factiva.

We obtain daily and monthly stock prices, stock returns, and shares outstanding from Center for Research on Security Prices (CRSP). We focus on all common stocks (i.e., share code equals to 10 or 11) in the CRSP universe and obtain relevant accounting information from the CRSP-Compustat Merged dataset. We use the historical Standard Industry Classification (SIC) codes from Compustat to assign all stocks into the 48 Fama and French (1997) industries. If the historical SIC code is not available, we use the SIC code from CRSP. We calculate book-to-market ratio for each firm using data from Compustat. Specifically, book-to-market ratio is calculated as the ratio of year-end stockholders' equity plus deferred taxes and investment tax credit minus preferred stocks to year-end market equity, as in Daniel and Titman (1997).

[30] Available at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
[31] The liquidity factor is the main variable in Pastor and Stambaugh (2003) and is available at: http://faculty.chicagobooth.edu/lubos.pastor/research/.

We obtain the Daniel, Grinblatt, Titman, and Wermers (DGTW, 1997) characteristic-adjustment stock assignments and benchmark portfolio returns from Russ Wermers' website. Since the benchmark returns are available until 2011, we use the Daniel, Grinblatt, Titman, and Wermes (1997) method to generate DGTW returns for the 2012 to 2013 period.[32] We estimate the DGTW return for each of the 48 Fama and French industry portfolio by value weighting stock-level DGTW returns.[33]

To ensure our return-based social sensitivity measure captures industry- or firm-level social attributes, we obtain stock-level ratings of corporate social responsibility from KLD. This database covers the Russell 3000 stocks during the 2005 to 2013 period and is widely used in the CSR literature. We focus on all the seven dimensions of corporate goodness rated by KLD: Community, Diversity, Corporate Governance, Employee Relations, Environment, Human Rights, and Product. KLD reports for each firm, its number of strengths and concerns across all these seven dimensions. Since the number of strength and concern indicators for most dimensions varies considerably each year, we use the Deng, Kang, and Low (2013) method to calculate stock-level KLD score so that it is comparable across years and dimensions. Specifically, we divide the aggregated strength (concern) of each stock by the total number of strengths (concerns) in each year to construct adjusted strength (concern). KLD score is the difference between adjusted strength and adjusted concern. Further, as KLD ratings are updated annually, we use the equal-weighted lagged KLD scores of industry firms in the Russell 3000 universe as proxies for the scores of the corresponding 48 Fama and French industry portfolios.[34]

---

[32] We verify the accuracy of our generated DGTW returns over the 2005 to 2011 period using the data from Russ Wermers' web site: http://www.smith.umd.edu/faculty/rwermers/ftpsite/Dgtw/coverpage.htm.

[33] Following Daniel, Grinblatt, Titman, and Wermers (1997) DGTW returns are calculated as follows: First, we rank all stocks listed on NYSE, AMEX, or Nasdaq, at the end of June, by their market capitalization and form quintile portfolios using NYSE quintile size breakpoints. We then further divide each quintile portfolio into book-to-market quintiles based on their most recently available book-to-market ratio as of the end of the December immediately prior to the ranking year. Finally, each of the resulting 25 portfolios are further subdivided into quintiles based on the return in the past 12 months through the end of May of the ranking year. This procedure forms 125 portfolios with each having a distinct combination of size, book-to-market, and momentum characteristics. We reconstruct the 125 portfolios at the end of each June. We calculate value-weighted returns for each of the 125 portfolios. DGTW adjusted return is defined as the return difference between the stock and the corresponding portfolio of which that stock is a member.

[34] We require all stocks to have valid adjusted strength and concern scores in all the seven dimensions to ensure comparability. This filter does not affect the KLD coverage during the 2005 – 2012 period. However, due to the significant criteria changes, only 300 firms have relevant data in 2013. Therefore, our sample for KLD scores ends in 2012.

To examine the institutional trading, we obtain transaction-level data of institutional investors for the 2005 to 2010 period from ANcerno Ltd. This dataset reports execution price, execution volume, side (i.e., buy or sell), and CUSIP for each transaction. As suggested by Puckett and Yan (2011), ANcerno institutions are larger than the typical institutions in the 13F universe and account for about 10% of all institutional trading volume. In addition, characteristics of stocks held and traded by ANcerno institutions are similar to those held by 13F institutions. We obtain institutional ownership data provided by Ferreira and Matos (2008) from FactSet. We measure institutional ownership of a firm using its average quarterly total institutional ownership in the previous year.

3.3.2. CSR attention measure

Following Da, Engelberg, and Gao (2011, 2015), we use the search volume intensity (*SVI*) reported by Google Trends to directly capture investors' attention to CSR. Specifically, we use the *SVI* for the *topic* "corporate social responsibility". This time-series measurement aggregates online search queries in different languages and different keywords if they are related to CSR. The topic feature in Google Trends is able to identify CSR-related searches even when a query does not explicitly contain the keyword "CSR".[35] As Google accounts for approximately 67 percent of all search queries in the U.S.,[36] high CSR attention reflects a market-level increase in social awareness. We restrict the search location to U.S. and measure the abnormal change in *SVI* (i.e., *ASVI*) as the log difference in *SVIs* between month t and month $t\text{-}1$, as in Da, Engelberg, and Gao (2011).[37] The time-series of *ASVI* starts from 2004.

Figure 3.1 plots the time-series of *SVI* of the topic "corporate social responsibility". We find that the attention to CSR is time-varying. Before presenting our social sensitivity estimation methods, we first validate our Google measure. In particular, we examine whether CSR related news draws investors' attention. This test is motivated by the "news headlines" feature reported by Google Trends, which implies that the peaks of the *SVI*

[35] For example, if you input "capital of Japan", Google Trends will aggregate your search into the topic "Tokyo". Google Trends is available at: http://www.google.com/trends/.
[36] The market share of Google is measured as of July, 2013. Source: comScore qSearch.
[37] The time-series is in weekly frequency. Since CSR literature that focusing on the asset pricing implications commonly examines monthly returns (e.g., Hong and Kacperczyk, 2009; Edmans, 2011), we aggregate the time-series to monthly frequency by linear interpolation. In addition, to be consistent with the literature using Google measure, we do not construct *ASVI* using error terms of time-series regressions.

time-series coincide with CSR news headlines. In addition, Barber and Odean (2008) show that news could generate attention-driven purchases.

To ensure the selection criteria of CSR news are objective and consistent over time, we hand collect CSR newsletters reported by Business Wire. Since March 6, 2008, Business Wire started to publish CSR newsletters called “corporate social responsibility weekly recap” on a weekly basis. Each newsletter summaries CSR news around the world during the previous week and covers various aspects of CSR: CSR awards, corporate philanthropy, corporate policies, employee satisfaction, technological breakthrough, non-profit organizations, and international environmental agreements. For each news headline, Business Wire reports the location, summary, publishing date, and information source. A sample newsletter is exhibited in Figure 3.2.[38]

Using these news headlines, we construct the monthly CSR news volume as follows. We first obtain all newsletters (i.e., 292 newsletters) from Factiva for the March 2008 to December 2013 period. We then assign each news headline into a corresponding calendar month using its publishing date. Further, during the March 2008 to December 2013 period, there are five weeks without CSR newsletters. No evidence suggests that the absence of CSR newsletters is caused by the lack of CSR news. Therefore, to ensure the comparability of CSR news volume over time, we delete months with missing newsletters because the total news volume of these months is lower than they should be and the CSR news volume in these months are likely to be biased downward.[39] Finally, we define log CSR news volume as the natural log of the total number of news headlines in each month.

Figure 3.3 plots the standardized log *SVI* of CSR and standardized log CSR news volume. We find that the two time-series share similar patterns. The pairwise correlation test shows that the correlation between log *SVI* and CSR news volume is 0.56, significant at the 1% level. Overall, we find that shifts in CSR attention could be associated with the time-variation of CSR news volume.

[38] The newsletter is available to the public from the Business Wire website: https://www.businesswire.com/portal/site/home/distribution/csr/

[39] These five weeks are: 30 April – 6 May 2009, 22 – 28 March 2012, 25 – 31 October 2012, 21 – 27 February 2013, and 22 – 28 August 2013. Our results are quantitatively similar if we include months with missing CSR newsletters.

3.3.3. Social sensitivity estimation and portfolio construction

Using Google search volume to capture investor attention, we estimate the return-based social sensitivity for each of the 48 Fama and French industry portfolios. This industry-level social sensitivity measure is motivated by the specifications used in Santa-Clara and Valkanov (2003) and Addoum and Kumar (2016). For example, Kumar (2016) shows that industry-level return sensitivity to the political party in power is able to capture the market's attitude toward an industry in the recent period, which in turn could identify industries that are favored by investors in the current political climate. Similarly, we use return sensitivity of industries to the attention to CSR to identify industries that are favored by the socially responsible investors in the recent period ex-ante. Specifically, in each month and for each industry portfolio, we regress the excess value-weighted industry returns during the past 18 months on the excess market return and a social sentiment indicator.[40] In particular, we estimate the following time-series regression:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_i (r_{mkt} - r_f)_t + \theta_i D_{social,t} + \epsilon_{i,t}. \qquad (3.1)$$

In equation (3.1), the CSR attention indicator variable (D_{social}) equals to one when *ASVI* of the topic "corporate social responsibility" in the previous month is above the time-series median, or zero otherwise. This definition of D_{social} gives us a longer sample period but raises the concern of potential look-ahead bias. In Section 3.7, we re-define D_{social} using only past information, and our results remain unchanged.

Our focus is on the θ_i estimate, which captures the social sensitivity of an industry. A positive θ_i estimate suggests that the industry earns higher average returns during high CSR attention periods. In contrast, a negative θ_i suggest that the industry earns higher average returns during low CSR attention periods. To allow for time-variation in both the magnitude and direction of the social sensitivity estimates, we estimate θ_i with rolling window regressions. The estimation period for θ_i is from April 2004 to November 2013 for our baseline specifications.

Using these social sensitivity estimates, we define five social sensitivity based industry portfolios as follows: in each month, we sort the 48 Fama and French (1997) industries by θ_i in descending order. We use the top five industries to form the Top portfolio and use

[40] We use 18 months as the estimation window to ensure that we have a longer sample period. Our results remain quantitatively similar if we use 24 or 36 months as the estimation window.

the bottom five industries to form the Bottom portfolio. The Top portfolio includes industries that are most favored by the market during high CSR attention periods, while the Bottom portfolio includes industries that are least favored by the market during the same periods. We assign the remaining 38 industries into portfolios 2, 3, and 4, with approximately 13 industries in each portfolio.[41] Portfolio returns are value-weighted by industry-level market capitalization in the previous month. We update industry sorting and portfolio construction on a monthly basis.

3.3.4. Characteristics of portfolios sorted by social sensitivity

Table 3.1 reports the characteristics of the five portfolios sorted by the social sensitivity measure. Panel A reports the mean social sensitivity, size (log market capitalization), book-to-market ratio, return over the past six months, and the average KLD score. We calculate portfolio-level KLD score by value weighting industry-level scores using industry market capitalizations in the previous month.

We find that social sensitivity and past returns increase monotonically from the Bottom to the Top portfolio while the five portfolios have similar book-to-market ratios. Further, as the Bottom and Top portfolios contain fewer industries, they are smaller than portfolios 2 to 4.

If our social sensitivity estimates are positively correlated with the level of corporate goodness, then industries in the Top portfolio are expected to have better CSR rating than the Bottom portfolio. The average KLD scores reported in Panel A support this conjecture. We find that the Top portfolio has higher average KLD score than the Bottom portfolio, which is significant at the 5% level.

Next, we examine the most prevalent industries in the Top and Bottom portfolios, respectively. Prevalence is measured by the number or percentage of months an industry is classified into the Top or Bottom portfolio. The existing literature on finance and economics suggests that investors have preconceptions about industry-level social attributes. For example, retail-based industries are commonly perceived as socially responsible. Firms in these industries invest extensively in CSR since good CSR

[41] During our sample period, we always have 48 industry portfolios in each month. Portfolio 2 contains 12 industries while portfolios 3 and 4 have 13 industries, respectively.

reputation could boost consumer demand, generate customer loyalty, support premium pricing, and assure product quality (e.g., Besley and Ghatak, 2007; Castaldo, Perrini, Misani, and Tencati, 2009; Elfenbein, Fisman, and Mcmanus, 2012; and Albuquerque, Durnev, and Koskinen, 2016).

In contrast, industries involving in fossil fuel (i.e., coal, oil, and natural gas) and other natural resources (e.g., mining, precious metal) are commonly screened by SRI investors (Geczy, Levin, and Stambaugh, 2005). Further, the USSIF also encourages all types of investors to divest from these industries to address climate changes risks.[42]

Panel B of Table 3.1 reports the five most prevalent industries in the Top and Bottom portfolios, respectively. We find that prevalent industries in the Top and Bottom portfolios are consistent with investors' preconception about social attributes. Specifically, retail-based industries are more prevalent in the Top portfolio while industries with controversial business operations are more prevalent in the Bottom portfolio. Further, prevalent industries remain similar when we assign more industries to the Top and Bottom portfolios, as reported in Panel C of Table 3.1. Overall, these findings suggest that our return-based social sensitivity estimates are consistent with conventional CSR ratings.

3.4. Empirical results

3.4.1. Univariate sorting results

To investigate whether social attributes affect stock returns, we first examine the performance of industry portfolios sorted by social sensitivity. Specifically, we examine the performance of the following portfolios: (i) the Bottom portfolio, which is a value-weighted portfolio of the five industries with the most negative social sensitivity during the past 18 months, (ii) the Top portfolio, which is a value-weighted portfolio of the five industries with the most positive social sensitivity in the past 18 months, (iii) the return differences between the Top and Bottom portfolios, and (iv-vi) portfolios 2-4, which are value-weighted portfolios of remaining Fama and French (1997) industries sorted into terciles based on social sensitivity in the previous 18 months.

[42] Source: http://www.ussif.org/climatereinvestment.

Table 3.2 presents the portfolio performance estimates. In Panel A, we report the raw and DGTW returns for the full sample period from October 2005 to December 2013. For robustness, we also report the portfolio return estimates by varying the number of industries in the Top and Bottom portfolios. Specifically, we use three alternative specifications to form the Top and Bottom portfolios and the new extreme portfolios contain three, seven, or ten industries, respectively. Standard errors reported in parentheses are adjusted for autocorrelation using the method in Newey and West (1987).

We find that portfolio raw returns increase monotonically from the Bottom portfolio to the Top portfolio. Industries in the Bottom portfolio earn an average raw monthly return of 0.02%, while industries in the Top portfolio earn an average raw monthly return of 1.58%. The monthly raw return difference between the Top and Bottom portfolios is 1.56% and is significant at the 1% level. Further, the return pattern remains similar when we use DGTW returns to measure performance. After adjusting for size, book-to-market, and past performance, we find that industries in the Bottom portfolio earn a significantly negative return of -0.56%. The DGTW return difference between the Top and Bottom portfolios is 1.17%, which translates into an economically meaningful annualized return of 14.01%.

Columns (3) to (8) reports the performance estimates when we vary the number of industries in the Top and Bottom portfolios. We find that even when we have ten industries in the two extreme portfolios, the return difference still translates into an annualized return of 0.613%×12=7.36%, which is both statistically and economically significant.

Next, we explicitly examine the riskiness of our social sensitivity based industry portfolios. In Panel B of Table 3.2, we find that the Top and Bottom portfolios have higher return standard deviations than the remaining portfolios. However, the Sharpe ratio increases monotonically from the Bottom portfolio to the Top Portfolio. The pattern remains similar when we vary the number of industries in the extreme portfolios.

In addition, we test whether our social sensitivity measure covers an economically meaningful segment of the market by examining the market share of the Top and Bottom portfolios in the CRSP universe. In Panel C of Table 3.2, we report the average monthly market shares for social sensitivity sorted portfolios. Top-Bottom reports the combined market share of the Top and Bottom portfolios. We find that for both raw and DGTW return measures, the Top-Bottom strategy covers about 15% of the market share in the CRSP universe. When we use ten portfolios to form the extreme portfolios, the combined market share of the Top and Bottom portfolios increases to 36%. Therefore, our social sensitivity based trading strategy covers an economically meaningful segment of the market.

To further ensure the robustness of our estimation results, Figure 3.4 plots the monthly DGTW return difference between the Top and Bottom portfolios for the full sample period. The bar chart shows that Top industries outperform Bottom industries in 60% of the sample period. In addition, the average monthly outperformance magnitude (i.e., 3.3%) is larger than the average underperformance magnitude (i.e., -1.9%). Further, we find that the outperformance of industries in the Top portfolio becomes stronger during the recent financial crisis. This is expected since firms with good social attributes are less likely to experience negative outcomes in social and environmental areas such as lawsuits and negative news related to environment and employee workplace safety. In addition, investors pay greater attention to corporate behaviour during poor economic conditions (e.g., Hirshleifer, 2008; Shefrin and Statman, 1993). Therefore, these firms may hold up better during financial crisis. This pattern is also consistent with the findings in Nofsinger and Varma (2014), who show that investors have higher demand for CSR during economic downturns.[43] Overall, these findings suggest that social sensitivity is positively correlated with future stock returns.

3.4.2. Factor model estimates

Our results based on raw and DGTW returns suggest that social sensitivity is positively correlated with future stock returns. In this section, we use various unconditional and conditional factor models to control for additional factors and to allow for a time-varying factor sensitivity. Specifically, our unconditional factor models include different combinations of the following factors: the market excess return (*MKTRF*), the size factor

[43] Due to the data limitation of Google Trends, this statement is based on a short sample period. However, arguably speaking, our definition of economic recession is based on the NBER business cycle definition. In addition, our sample covers months before 2006, a period that is commonly regarded as an economic expansion. Further, our sample also includes the post 2007 financial crisis period. Therefore, even though we only have one crisis during our sample periods, our sample also has two economic upturns.

(*SMB*), the value factor (*HML*), the momentum factor (*UMD*), short-term reversal (*STR*) and long-term reversal (*LTR*) factors, and the liquidity factor (*LIQ*). The estimation period is from October 2005 to December 2013.

Table 3.3 reports the unconditional factor model estimation results. We find that our results remain robust across all specifications. In particular, even after including seven risk factors, the monthly alpha for the Top and Bottom portfolios are 79 and -91 basis points, respectively, both significant at the 1% level. The performance difference between the Top and the Bottom portfolios converts into an annualized risk-adjusted return of 1.704%×12=20.45%.

Next, to ensure our results are not driven by improper adjustment for time-varying exposures to systematic risks, we use conditional factor models to address portfolio risks. In particular, we interact factors in unconditional models with macroeconomic variables to account for the U.S. business cycle.

Specifically, we include the interaction with the following macroeconomic variables: (i) NBER recession indicator (REC), (ii) the Lettau and Ludvigson (2001) *cay* measure, defined as the difference between current consumption and its long-run value based on assets and income, (iii) the yield on the 90-day T-bill (YLD), (iv) the dividend yield of the value-weighted CRSP index (DIV), (v) the term spread (TERM), defined as the difference between the yields of a constant maturity 10-year Treasury bond and a 90-day T-bill, and (vi) the default spread (DEF), defined as the difference between Moody's Baa-rated and Aaa-rated corporate bond yields.

Table 3.4 reports the conditional factor model estimation results. Our focus is the alpha estimate of a trading strategy that goes long in the Top portfolio and goes short in the Bottom portfolio. The interaction variable for each conditional factor model is indicated at the top of each column. We find that the alpha estimates remain economically significant (1.19% - 1.53%) when we use conditional factor model. These findings show that our results are robust to accounting for changes in business cycle over time.

3.4.3. Fama-MacBeth regression estimates

In the last sets of our baseline tests, we estimate Fama and MacBeth (1973) regressions. The dependent variable is the monthly value-weighted returns of the 48 Fama and French (1997) industry portfolios. The main explanatory variable is the lagged social sensitivity estimate θ_i. We also include the following explanatory variables that are commonly used to predict cross-sectional returns: the factor loadings of the Fama and French (1992) three-factor model estimated by daily industry returns in the previous month (*Beta MKTRF*, *Beta SMB*, *Beta HML*), return over the past six months (*Lag 6m Return*), value weighted log market capitalization of industry firms in the previous month (*Size*), and value-weighted book-to-market ratio of industry firms by using book-to-market-ratio available in the previous year (book-to-market). The sample period is from October 2005 to December 2013. We report the time-series averages of monthly cross-sectional regression coefficients, Standard errors are adjusted using the Newey and West (1987) method.

Table 3.5 reports the results. We find that industries with higher social sensitivity earn higher returns even after controlling for all commonly used factors in the literature. In economic terms, a one standard deviation increase in social sensitivity is associated with an additional return of 0.118×0.446=5% in the following month. Our results show that social sensitivity of industries is an important factor in explaining the cross-sectional variation in returns and this effect is different from firm characteristics that are known to predict cross-sectional returns. This evidence further supports our main hypothesis.

3.4.4. Factor model estimates: stock level

Our results so far are based on industry-level analyses. In this section, we examine whether the positive relation between social sensitivity and stock returns remains robust at the stock level. In particular, we focus on all common stocks in the CRSP universe (i.e., share code = 10 or 11). In each month, we estimate social sensitivity for each stock and use these estimates to sort all stocks in descending order. We form the Top (Bottom) portfolio by value weighting stocks located in the top (bottom) quintile.

Table 3.6 reports the stock-level unconditional factor model estimation results. Since the social sensitivity estimates at stock level are more volatile than the industry level estimates, component stocks in stock quintile portfolios are likely to vary a lot. Therefore, we expect a smaller alpha estimates for stock-level analysis. Consistent with our expectation, we find that the magnitude for alpha estimates reduced by more than 50%. Nevertheless, the performance difference between the Top and Bottom portfolios still translates into an annualized seven-factor adjusted alpha of 0.655×12=7.9%, which is economically meaningful.

3.4.5. Institutional trading

Our baseline results demonstrate that perceived social attributes, as measured by social sensitivity, predict stock returns. In this section, we investigate a potential economic channel for this return predictability.

Recent literature suggests that with norm constraints, institutional investors are less likely to hold sin stocks (i.e., publicly traded firms involved in producing alcohol, tobacco, and gaming, Hong and Kacperczyk, 2009) and stocks with lottery-like characteristics (Kumar, Page, and Spalt, 2011). In addition, institutional investors are sophisticated and have better investment skills (e.g., Puckett and Yan, 2011). Therefore, they are likely to alter their portfolio holdings to include stocks with good perceived social attributes in the recent period.

We start by examining the return performance of double-sorted portfolios. In particular, if social attributes generate institutional trading, then the return predictability patterns are likely to be stronger among firms with high institutional ownership. In addition, the existing literature also demonstrates that institutions prefer large stocks (e.g., Lakonishock, Shleifer, and Vishny, 1992; Gompers and Metrick, 2001). Therefore, we construct portfolios sorted by firm size and social sensitivity, or by institutional ownership and social sensitivity. We focus on all common stocks in the CRSP universe. In each month, we classify stocks into large or small size group (high or low institutional ownership group) if its market capitalization (institutional ownership) is above or below the median value across all firms in that month. Within each size (institutional ownership) category, we further partition stocks in to high (low) social sensitivity group if the firm's social sensitivity is above (below) the top (bottom) tercile across firms.

Table 3.7 presents the value weighted seven-factor adjusted alpha for the double-sorted portfolios. Standard errors reported in the parentheses are adjusted for autocorrelation using the method in Newey and West (1987). Consistent with the view that social attributes trigger institutional trading, we find that the alpha estimates are stronger among large firms (Panel A) and firms with high institutional ownership (Panel B).

Next, we directly test whether perceived social attributes generate institutional trading. In particular, we use the actual transactions of institutional investors during the 2005 to 2010 period. Following Kumar and Lee (2006), we measure the aggregate demand for top social sensitivity stocks as the excess buy-sell imbalance (*EBSI*) defined as the difference in buy-sell imbalance between top and bottom social sensitivity stocks.[44] This measure captures the changes in net demand for top social sensitivity stocks relative to bottom social sensitivity stocks. For robustness, we also examine the *EBSI* between top social sensitivity stocks and remaining stocks.

Table 3.8 presents the results. Consistent with our expectation, we find that the average *EBSI* between top and bottom social sensitivity stocks is 1.3% per month, significant at the 5% level. This evidence suggests that institutions have 1.3% more net purchases of top social sensitivity stocks relative to bottom social sensitivity stocks during our sample period. In addition, we find that the net purchase of top social sensitivity stocks is higher than that of bottom social sensitivity stocks in 60% of the sample. Similarly, institutions also have 1.8% net purchase of top social sensitivity stocks relative to all the remaining stocks, and the net purchase is positive in two-thirds of our sample. Overall, these findings suggest that institutional investors regularly rebalance their portfolios to include stocks with good perceived social attributes in the recent period.

3.4.6. Longevity of return predictability

In this section, we study the longevity of the predictive power of our social sensitivity estimates. In particular, if it is driven by mispricing, then one might expect the perdition power of our social sensitivity estimates to decline if the gap between social sensitivity estimation month and portfolio formation month is widened. Since perceived social attribute is not a permanent characteristic, investors are likely to update their perception on social attributes when new information draws their attention. For example, investors immediately adjust their perception about Volkswagen (which was widely regarded as a pioneer in clean technology) when the diesel emission scandal occurs. In addition, our institutional trading results also suggest that institutional investors are likely to identify

[44] The buy-sell imbalance (*BSI*) of portfolio p in month t is defined as $BSI_{pt} = \frac{100}{N_{pt}}\sum_{i=1}^{N_{pt}} BSI_{it}$, where the *BSI* for stock i in month t is defined as $BSI_{it} = \frac{\sum_{j=1}^{D_t}(VB_{ijt}-VS_{ijt})}{\sum_{j=1}^{D_t}(VB_{ijt}+VS_{ijt})}$. Here, D_t is the number of days in month t. VB_{ijt} (VS_{ijt}) is the dollar buying (selling) volume of stock i on day j in month t, and N_{pt} is the number of traded stocks in portfolio p in month t.

stocks with good perceived social attributes in the recent past and help to correct the underpricing of these stocks.

Table 3.9 reports the unconditional seven-factor adjusted alpha of a trading strategy that goes long in the Top portfolio and goes short in the Bottom portfolio when we vary the portfolio formation periods. Consistent with the mispricing hypothesis, we find that the magnitude of seven-factor alpha gradually declines as we increase the formation period. It is not significantly different from zero beyond three months. This evidence suggests that institutional investors are likely to correct the underpricing of perceived social attributes in about 3 months.[45]

3.4.7. Robustness checks

In this section, we conduct several robustness checks for our baseline results. First, Table 3.1 shows that past returns increase monotonically with social sensitivity, which raises the concern that our CSR sensitivity measure merely captures momentum return. However, results in Table 3.3 suggest that momentum factor does not explain our results. For further robustness, we use the Carhart (1997) four-factor model to re-estimate industry-level social sensitivity as follows:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_i\left(r_{mkt} - r_f\right)_t + SMB_t + HML_t + UMD_t + \theta_i D_{social,t} + \epsilon_{i,t}. \tag{3.2}$$

This alternative social sensitivity measure explicitly accounts for the momentum return. The long-run abnormal return test results using social sensitivity estimated by equation (3.2) are reported in Columns 1 to 3 of Table 3.10. We find that the seven-factor alphas for the Top, Bottom, and the Top-Bottom portfolios are quantitatively similar after adding the momentum factor. Therefore, this evidence suggests that our return predictability results are not explained by the momentum factor.

Second, we address the potential look-ahead bias in our social sensitivity estimates. In our baseline specification, to ensure a longer sample, we use the median of the time-series as the cutoff point to define D_{social}. For robustness, we use expanding window medians as an alternative method to define D_{social}. In particular, $D_{social,\,t}$ equals to 1 if $ASVI_{t-1}$ is larger

[45] This finding also suggests that our return predictability results are not likely to be explained by an overpricing story since we do not observe price reversal after three months.

than the median of all previous observations, or zero otherwise. We use the median of *ASVIs* in 2004 to define the first D_{social} in January 2005.

Columns 4 to 6 of Table 3.10 reports the seven-factor alpha results using expanding-window social sensitivity. We find that our results remain quantitatively similar. Further, we also define social sensitivity using equation (3.2) and expanding window median values to account for momentum return and potential look-ahead bias at the same time. Columns 7 to 9 of Table 3.10 reports the results. We find that the seven-factor alphas remain unchanged. Overall, evidence in Table 3.10 suggests that the outperformance of the Top portfolio is not explained by the momentum factor or look-ahead bias.

Third, we also re-estimate conditional factor model and Fama-MacBeth type regressions using the three alternative social sensitivity definitions. The estimation results for conditional factor models and Fama-MacBeth type regressions are reported in Table 3.11 and Table 3.12, respectively. Again, the estimation results remain quantitatively similar.

3.5. Summary and conclusion

Survey evidence suggests that investors are actively updating their beliefs about firm-level CSR records. Therefore, their recent perception on firm-level social attributes could affect their investment decisions. In particular, investors are likely to invest more in firms with good perceived social attributes. In this paper, we propose a novel measure to identify firms that are likely to be perceived as having good social attributes by the market. Specifically, we use social sensitivity, defined as the return sensitivity to the aggregate attention to CSR, to capture perceived social attributes.

We show that social sensitivity is positively correlated with industry- or stock-level CSR records. Using social sensitivity estimates, we find that returns of market segments with high social sensitivity are predictable. A trading strategy that goes long in stocks with good perceived social attributes and goes short in stocks with bad perceived social attributes generates a monthly DGTW return of 1.17%. Our return predictability evidence remains robust after controlling for a broad set of factors or observable characteristics.

Further, by investigating institutional trading, we demonstrate that social attributes trigger institutional demand. Overall, our results suggest that perceived social attributes affect stock returns. In future work, it would be interesting to examine whether perceived social attributes affect mutual fund flows since institutional investors have exhibited a clear preference for firms with good social attributes.

Table 3.1: Characteristics of social sensitivity based industry portfolios

This table reports characteristics of portfolios defined by social sensitivity. We focus on 48 Fama and French (1997) industries. The Top (Bottom) portfolio is a value-weighted portfolio of the five industries with the most positive (negative) social sensitivity in the past 18 months. Portfolios 2-4 are value-weighted portfolios of the remaining industries sorted into terciles based on social sensitivity during the same estimation window. Panel A reports the mean social sensitivity, size (log market capitalization), book-to-market ratio, six months' cumulated return with a one-month lag, and KLD score. KLD score is estimated at stock-level by using the Deng, Kang, and Low (2013) method and then valued weighted to portfolio-level by using stock-level market capitalization in the previous month. Panel B (C) reports the five (ten) most prevalent industries in the Top and Bottom portfolios when five (ten) industries are used to defined the Top and Bottom portfolios. Prevalence is measured by the number (i.e., *N*) or percentage (i.e., *% months*) of months an industry is included into a given portfolio during the estimation period. The estimation period is from April 2004 to November 2013. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Portfolio characteristics

Portfolio	Social sensitivity	Size	BM	Lag 6m Return	KLD score
1 (Bottom)	-3.262	13.885	0.600	2.647	-0.375***
2	-1.088	15.113	0.531	4.101	-0.320***
3	0.342	15.219	0.471	5.917	-0.258***
4	1.548	15.026	0.491	7.400	-0.250***
5 (Top)	3.246	13.182	0.483	7.557	-0.253***
Top - Bottom					0.121**

Panel B: Five prevalent industries in the Top and Bottom portfolios

Industry (Top)	N	% months	Industry (Bottom)	N	% months
Fabricated Products	61	61.616	Agriculture	56	56.566
Textiles	37	37.374	Precious Metals	38	38.384
Recreation	30	30.303	Mining	33	33.333
Apparel	29	29.293	Coal	33	33.333
Automobiles	26	26.263	Banking	29	29.293

Panel C: Ten prevalent industries in the Top and Bottom portfolios

Industry (Top)	N	% months	Industry (Bottom)	N	% months
Fabricated Products	80	80.808	Agriculture	74	74.747
Apparel	53	53.535	Construction	48	48.485
Textiles	46	46.465	Precious Metals	48	48.485
Recreation	45	45.455	Defense	45	45.455
Transportation	40	40.404	Medical Equipment	45	45.455
Rubber and Plastic Products	36	36.364	Pharmaceutical Products	44	44.444
Healthcare	35	35.354	Mining	44	44.444
Computers	34	34.343	Banking	42	42.424
Construction Materials	32	32.323	Petroleum and Natural Gas	41	41.414
Personal Services	30	30.303	Other	38	38.384

Table 3.2: Performance of social sensitivity based industry portfolios

This table reports the performance of the five portfolios sorted by social sensitivity. We focus on 48 Fama and French (1997) industries. The Top (Bottom) portfolio is a value-weighted portfolio of the five industries with the most positive (negative) social sensitivity in the past 18 months. Portfolios 2-4 are value-weighted portfolios of the remaining industries sorted into terciles based on social sensitivity during the same estimation window. Top-Bottom reports the performance difference between the Top and Bottom portfolios. Panel A reports the raw and DGTW returns during our sample period from October 2005 to December 2013. DGTW returns are calculated using the Daniel, Grinblatt, Titman, and Wermers (1997) method. Columns (1) and (2) reports the raw and DGTW returns when we use five industries to construct the Top and Bottom portfolios. For robustness, we also vary the number of industries in the Top and Bottom portfolios. In Columns (3) to (8), we report the raw and DGTW returns when we use three, seven, and ten industries to form the Top and Bottom portfolios. Panel B reports the standard deviation of portfolio excess returns and the Sharpe ratio. Panel C reports the average monthly market share for the raw and DGTW return portfolios. We also report the total market share covered by the Top-Bottom trading strategy. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Portfolio performance estimates

	Baseline		3 industries		7 industries		10 industries	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Portfolio	Raw return	DGTW return	Raw return	DGTW return	Raw return	DGTW return	Raw return	DGTW return
1 (Bottom)	0.016	-0.564**	0.294	-0.587	0.268	-0.297*	0.482	-0.203
	(0.816)	(0.272)	(0.856)	(0.411)	(0.674)	(0.169)	(0.596)	(0.168)
2	0.557	-0.141	0.415	-0.220**	0.496	-0.172	0.452	-0.163
	(0.585)	(0.094)	(0.600)	(0.095)	(0.586)	(0.107)	(0.600)	(0.106)
3	0.743	-0.076	0.744	-0.073	0.774	-0.079	0.824	-0.059
	(0.594)	(0.117)	(0.583)	(0.112)	(0.599)	(0.122)	(0.604)	(0.123)
4	1.254**	0.241**	1.306**	0.272**	1.264**	0.267**	1.114**	0.146
	(0.540)	(0.100)	(0.533)	(0.110)	(0.530)	(0.111)	(0.553)	(0.119)
5 (Top)	1.575***	0.604**	2.131***	1.007***	1.499***	0.516***	1.474***	0.410***
	(0.544)	(0.238)	(0.568)	(0.330)	(0.562)	(0.183)	(0.516)	(0.135)
Top - Bottom	1.559***	1.168***	1.838***	1.595***	1.230***	0.813***	0.992***	0.613**
	(0.509)	(0.443)	(0.592)	(0.550)	(0.350)	(0.271)	(0.338)	(0.252)
N months	99	99	99	99	99	99	99	99

Table 3.2 (Cont'd)

Panel B: Portfolio performance characteristics

	Baseline		3 industries		7 industries		10 industries	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Portfolio	Std Dev	Sharpe ratio	Std Dev	Sharpe ratio	Std Dev	Sharpe ratio	Std Dev	Sharpe ratio
1 (Bottom)	6.270	-0.017	6.618	0.026	5.649	0.026	5.399	0.067
2	5.011	0.087	5.068	0.058	4.975	0.075	4.927	0.067
3	4.793	0.130	4.771	0.130	4.805	0.136	4.842	0.145
4	4.923	0.230	4.908	0.241	4.948	0.231	4.982	0.199
5 (Top)	5.504	0.264	6.008	0.334	5.320	0.259	4.901	0.276
Top - Bottom	4.773	0.327	5.549	0.331	3.760	0.327	3.392	0.292

Panel C: Average monthly portfolio market share

	Baseline		3 industries		7 industries		10 industries	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Portfolio	Raw return	DGTW return	Raw return	DGTW return	Raw return	DGTW return	Raw return	DGTW return
1 (Bottom)	10.098	9.894	5.669	5.453	14.594	14.470	21.397	21.357
2	28.158	28.415	32.587	32.856	25.996	26.266	21.569	21.887
3	30.520	30.405	32.614	32.376	28.186	27.979	23.649	23.379
4	25.802	25.982	26.272	26.494	22.600	22.797	19.078	19.021
5 (Top)	5.423	5.304	2.858	2.820	8.624	8.488	14.307	14.356
Top - Bottom	15.521	15.198	8.527	8.273	23.218	22.959	35.704	35.713

Table 3.3: Factor model estimation

This table reports factor model performance estimation of portfolios sorted by social sensitivity. We focus on the 48 Fama and French (1997) industries. The Top (Bottom) portfolio is a value-weighted portfolio of the five industries with the most positive (negative) social sensitivity in the past 18 months. Top-Bottom reports the performance difference between the Top and Bottom portfolios. The factor models include the following factors: the market excess return (MKTRF), the size factor (SMB), the value factor (HML), the momentum factor (UMD), short-term reversal (STR) and long-term reversal (LTR) factors, and the liquidity factor (LIQ). The sample period is from October 2005 to December 2013. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Factor	Top	Bottom	Top - Bottom	Top	Bottom	Top - Bottom	Top	Bottom	Top - Bottom
Alpha	0.815***	-0.822**	1.637***	0.808***	-0.869***	1.677***	0.793***	-0.911***	1.704***
	(0.273)	(0.329)	(0.519)	(0.254)	(0.323)	(0.481)	(0.252)	(0.311)	(0.474)
MKTRF	0.985***	1.205***	-0.220	0.995***	1.256***	-0.261	0.989***	1.241***	-0.252
	(0.121)	(0.112)	(0.220)	(0.111)	(0.103)	(0.190)	(0.112)	(0.100)	(0.191)
SMB	0.101	-0.190	0.291	0.072	-0.064	0.136	0.031	-0.179	0.210
	(0.117)	(0.214)	(0.294)	(0.121)	(0.222)	(0.316)	(0.122)	(0.197)	(0.301)
HML	0.182	-0.089	0.271	0.109	0.128	-0.019	0.152	0.248	-0.097
	(0.196)	(0.184)	(0.317)	(0.166)	(0.207)	(0.283)	(0.142)	(0.218)	(0.315)
UMD	-0.002	-0.014	0.013	-0.006	0.004	-0.010	-0.016	-0.025	0.009
	(0.077)	(0.069)	(0.105)	(0.078)	(0.069)	(0.122)	(0.076)	(0.065)	(0.124)
STR				-0.047	-0.125	0.078	-0.072	-0.195*	0.123
				(0.084)	(0.106)	(0.138)	(0.083)	(0.106)	(0.146)
LTR				0.122	-0.356*	0.478	0.177	-0.202	0.378
				(0.180)	(0.200)	(0.343)	(0.198)	(0.180)	(0.347)
LIQ							0.084	0.238**	-0.154
							(0.070)	(0.093)	(0.127)
Adj R2	0.782	0.732	0.003	0.780	0.747	0.033	0.781	0.764	0.037
N months	99	99	99	99	99	99	99	99	99

Table 3.4: Conditional factor model: Top-Bottom

This table reports conditional factor model performance estimation of portfolios sorted by social sensitivity. We focus on the 48 Fama and French (1997) industries. The factor models include the following factors: the market excess return (MKTRF), the size factor (SMB), the value factor (HML), the momentum factor (UMD), short-term reversal (STR) factor, and long-term reversal (LTR) factor. We interact each factor with one of the following interaction variables (INT): the recession indicator from the National Bureau of Economic Research (REC), the Lettau and Ludvigson (2004) *cay* measure, the dividend yield of the value-weighted CRSP index (DIV), the yield on the 90-day T-bill (YLD), the term spread (TERM), and the default spread (DEF). Columns (1) to (6) report the performance difference between the Top and Bottom portfolios. The interaction variable used in each regression is indicated at the top of each column. The Top (Bottom) portfolio is a value-weighted portfolio of the five industries with the most positive (negative) social sensitivity in the past 18 months. The sample period is from October 2005 to December 2013. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Interaction variable (INT)	REC	cay	DIV	YLD	TERM	DEF
Factor	(1)	(2)	(3)	(4)	(5)	(6)
Alpha	1.366***	1.190***	1.483***	1.528***	1.480***	1.381***
	(0.488)	(0.403)	(0.470)	(0.527)	(0.482)	(0.493)
MKTRF	-0.125	-0.313**	-0.186	-0.245	-0.389	0.273
	(0.194)	(0.134)	(0.429)	(0.268)	(0.236)	(0.337)
SMB	-0.118	0.771	-1.251	0.579*	-0.438	0.310
	(0.401)	(0.496)	(1.265)	(0.307)	(0.542)	(0.819)
HML	-0.188	0.273	-1.880**	0.161	-1.179**	-0.409
	(0.446)	(0.273)	(0.816)	(0.293)	(0.570)	(0.765)
UMD	0.190	-0.065	0.632	-0.002	0.341	0.747*
	(0.208)	(0.275)	(0.435)	(0.105)	(0.450)	(0.382)
STR	0.192	0.010	0.955	-0.223	1.615***	0.181
	(0.274)	(0.172)	(0.659)	(0.169)	(0.422)	(0.481)
LTR	0.519	0.163	1.604*	0.450	-0.252	0.456
	(0.408)	(0.260)	(0.861)	(0.323)	(0.722)	(0.791)
MKTRF×INT	-0.507	-0.260**	-0.650	-0.048	0.058	-0.472**
	(0.395)	(0.124)	(2.576)	(0.080)	(0.111)	(0.218)
SMB×INT	0.614	0.479	8.351	-0.273**	0.273	-0.334
	(0.584)	(0.410)	(6.229)	(0.131)	(0.223)	(0.527)
HML×INT	0.718	0.278	9.918***	-0.304**	0.463*	0.364
	(0.667)	(0.243)	(3.583)	(0.152)	(0.238)	(0.373)
UMD×INT	-0.288	-0.034	-3.232	0.134	-0.112	-0.424**
	(0.246)	(0.192)	(2.019)	(0.105)	(0.153)	(0.197)
STR×INT	-0.130	-0.339*	-5.359	0.465***	-0.636***	0.128
	(0.347)	(0.193)	(3.585)	(0.133)	(0.166)	(0.298)
LTR×INT	-0.403	0.161	-5.547	-0.108	0.249	0.031
	(0.614)	(0.257)	(3.823)	(0.187)	(0.255)	(0.398)
Adj R^2	0.031	0.136	0.059	0.157	0.168	0.038
N months	99	99	99	99	99	99

Table 3.5: Social sensitivity and expected returns

This table reports estimates from Fama Macbeth (1973) regressions. We focus on returns of value-weighted Fama and French (1997) 48 industry portfolios. The dependent variable is monthly industry portfolio return. Regressors include lagged social sensitivity loading, industry-level Fama and French (1992) three-factor loadings estimated by daily returns over the previous month, cumulated industry return over the past six months (*Lag 6m Return*), lagged value-weighted log market capitalization of industry-firms (*Size*), value-weighted book-to-market ratio of industry-firms using book-to-market ratio in the previous year (*Book-to-market*). We report the time-series averages of monthly cross-sectional regression coefficients. The sample period is from October 2005 to December 2013. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Factor	(1)	(2)	(3)	(4)
Social sensitivity	0.133**	0.126**	0.121***	0.118**
	(0.057)	(0.048)	(0.044)	(0.045)
Beta MKTRF	0.104	-0.463	-0.588	-0.522
	(0.533)	(0.502)	(0.463)	(0.477)
Beta SMB	-0.266	-0.309	-0.392	-0.373
	(0.242)	(0.251)	(0.267)	(0.268)
Beta HML	0.047	-0.017	-0.025	0.028
	(0.148)	(0.142)	(0.146)	(0.152)
Lag 6m Return		0.010	0.009	0.009
		(0.013)	(0.013)	(0.013)
Size			-0.027	-0.048
			(0.086)	(0.084)
Book-to-market				-0.424
				(0.360)
Constant	0.750*	0.875**	1.254	1.580
	(0.389)	(0.433)	(1.038)	(0.958)
Average R^2	0.294	0.345	0.378	0.395
Observations	4,752	4,752	4,752	4,752

Table 3.6: Factor model estimation: stock level

This table reports factor model performance estimation of portfolios sorted by social sensitivity. We focus on stock-level social sensitivity by using all common stocks (share code equals 10 or 11) in the CRSP universe. Stocks are sorted into quintiles based on social sensitivity. The Top (Bottom) portfolio is a value-weighted portfolio of stocks in the top (bottom) social sensitivity quintile. Top -Bottom reports the performance difference between the Top and Bottom portfolios. The factor models include the following factors: the market excess return (MKTRF), the size factor (SMB), the value factor (HML), the momentum factor (UMD), short-term reversal (STR) and long-term reversal (LTR) factors, and the liquidity factor (LIQ). The sample period is from October 2005 to December 2013. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Factor	Top	Bottom	Top - Bottom	Top	Bottom	Top - Bottom	Top	Bottom	Top - Bottom
Alpha	0.262*	-0.328	0.590*	0.273**	-0.371	0.644**	0.267**	-0.388	0.655**
	(0.150)	(0.244)	(0.326)	(0.134)	(0.245)	(0.303)	(0.134)	(0.240)	(0.297)
MKTRF	1.074***	1.226***	-0.152	1.062***	1.272***	-0.210**	1.060***	1.266***	-0.206**
	(0.048)	(0.090)	(0.130)	(0.035)	(0.080)	(0.102)	(0.036)	(0.079)	(0.103)
SMB	0.285***	0.110	0.175	0.233**	0.192	0.041	0.216**	0.144	0.072
	(0.085)	(0.148)	(0.202)	(0.091)	(0.149)	(0.209)	(0.098)	(0.133)	(0.196)
HML	0.044	0.036	0.008	-0.055	0.164	-0.219	-0.037	0.215	-0.252
	(0.121)	(0.127)	(0.213)	(0.107)	(0.129)	(0.183)	(0.107)	(0.134)	(0.211)
UMD	0.059*	-0.227***	0.287***	0.052	-0.216***	0.268***	0.047	-0.228***	0.275***
	(0.036)	(0.056)	(0.075)	(0.037)	(0.039)	(0.057)	(0.038)	(0.036)	(0.056)
STR				0.016	-0.128*	0.145	0.006	-0.158**	0.163
				(0.050)	(0.069)	(0.090)	(0.053)	(0.071)	(0.099)
LTR				0.165*	-0.208	0.373**	0.188*	-0.143	0.331*
				(0.097)	(0.131)	(0.174)	(0.105)	(0.120)	(0.179)
LIQ							0.036	0.100*	-0.064
							(0.044)	(0.060)	(0.087)
Adj R^2	0.920	0.895	0.234	0.922	0.902	0.298	0.922	0.903	0.295
N months	99	99	99	99	99	99	99	99	99

Table 3.7: Performance of double-sorted portfolios

This table reports performance estimates of double sorted portfolios defined using social sensitivity and an additional firm characteristic. We use market capitalization or institutional ownership (IO) as an additional firm characteristic. Component returns are those of all common stocks (share code=10 or 11) in the CRSP universe. In each month, stocks are classified into large or small size groups (high or low institutional ownership group) if its market capitalization (institutional ownership) is above or below the sample median. In addition, within a given size (institutional ownership) group, a stock is further classified as being in the high (low) return-based social sensitivity (RBSS) category if its RBSS above (below) the top (bottom) tercile. We report the alpha estimates using the seven-factor unconditional model. Panel A reports the alpha estimates of size and social sensitivity sorted portfolios while Panel B reports the alpha estimates of institutional ownership and social sensitivity sorted portfolios. The sample period is from October 2005 to December 2013. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Portfolios sorted by size and social sensitivity

Portfolio	1 (Small size)	2 (Large size)	Large - Small
1 (Low RBSS)	-0.309	-0.286**	0.023
	(0.210)	(0.127)	(0.205)
2	-0.062	0.018	0.080
	(0.126)	(0.055)	(0.137)
3 (High RBSS)	-0.084	0.286**	0.370**
	(0.136)	(0.120)	(0.167)
High - Low RBSS	0.225	0.572**	
	(0.216)	(0.230)	

Panel B: Portfolios sorted by IO and social sensitivity

Portfolio	1 (Low IO)	2 (High IO)	High - Low IO
1 (Low RBSS)	-0.462*	-0.366**	0.097
	(0.249)	(0.162)	(0.195)
2	0.094	0.026	-0.068
	(0.117)	(0.047)	(0.125)
3 (High RBSS)	0.185	0.238*	0.053
	(0.198)	(0.131)	(0.226)
High - Low RBSS	0.647*	0.604**	
	(0.356)	(0.274)	

Table 3.8: Institutional trading of top social sensitivity stocks

This table reports the excess buy-sell imbalance (*EBSI*) of top social sensitivity stocks. *EBSI* is the monthly difference in buy-sell imbalance between top and bottom social sensitivity stocks (Top-Bottom), or between top social sensitivity stocks and all the remaining common stocks (share code=10 or 11) in the CRSP universe (Top-Remaining). Stocks are sorted into quintiles based on social sensitivity. We define top (bottom) social sensitivity stocks as stocks in the top (bottom) social sensitivity quintile. Remaining stocks are stocks that are not classified into the top social sensitivity quintile. We also report the percentage of months when *EBSI* is positive. The sample period is from October 2005 to December 2013. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Top-Bottom	Top-Remaining
EBSI	1.327**	1.796***
	(0.565)	(0.497)
% of positive EBSI	60.32%	66.67%

Table 3.9: Longevity of return predictability

This table reports the effect of varying portfolio formation periods on average monthly seven-factor adjusted abnormal return to of portfolios sorted by social sensitivity. We focus on the performance difference between the Top and Bottom portfolios. The Top (Bottom) portfolio is a value-weighted portfolio of the five industries with the most positive (negative) social sensitivity in the past 18 months. We report the monthly seven-factor alpha when we positively shift the portfolio formation period by 1-6 months. A positive shift in portfolio formation period corresponds to delayed formation of the Top and Bottom portfolios. A shift of zero is equivalent to the baseline portfolio formation procedure and the coefficients are the same as those reported in Column (9) of Table 3.3. The sample period is from October 2005 to December 2013. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Baseline	Shift 1 month	Shift 2 month	Shift 3 month	Shift 4 month	Shift 5 month	Shift 6 month
Alpha	1.704***	1.131***	0.849***	0.610*	0.093	0.185	0.184
	(0.474)	(0.404)	(0.303)	(0.346)	(0.384)	(0.502)	(0.461)
N months	99	99	99	99	99	99	99

Table 3.10: Factor model estimation: robustness

This table reports factor model performance estimation of industry portfolios sorted by social sensitivity. We estimate social sensitivity in three alternative ways: (i) in Columns 1 to 3, we estimate social sensitivity by equation (3.2) to account for the Carhart (1997) four factors, (ii) in Columns 4 to 6, we define D_{social} using expanding window medians to address look-ahead bias, and (iii) in Columns 7 to 9, we combine methods use in (i) and (ii) to account for more factors and look-ahead bias at the same time. We focus on the 48 Fama and French (1997) industries. The Top (Bottom) portfolio is a value-weighted portfolio of the five industries with the most positive (negative) social sensitivity in the past 18 months. Top-Bottom reports the performance difference between the Top and Bottom portfolios. The factor models include the following factors: the market excess return (MKTRF), the size factor (SMB), the value factor (HML), the momentum factor (UMD), short-term reversal (STR) and long-term reversal (LTR) factors, and the liquidity factor (LIQ). The sample period is from October 2005 to December 2013. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Multi-factor sensitivity			Expanding window sensitivity			Combined		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Factor	Top	Bottom	Top - Bottom	Top	Bottom	Top - Bottom	Top	Bottom	Top - Bottom
Alpha	0.790***	-0.355	1.145***	0.683***	-0.979***	1.662***	0.994***	-0.399	1.393***
	(0.211)	(0.254)	(0.325)	(0.250)	(0.344)	(0.448)	(0.235)	(0.317)	(0.434)
MKTRF	0.948***	1.091***	-0.142	1.029***	1.227***	-0.198	0.903***	1.118***	-0.215
	(0.066)	(0.084)	(0.115)	(0.082)	(0.119)	(0.157)	(0.074)	(0.106)	(0.161)
SMB	0.175	-0.073	0.248	0.199*	-0.411***	0.610**	0.263**	-0.207	0.469**
	(0.120)	(0.187)	(0.257)	(0.117)	(0.142)	(0.235)	(0.129)	(0.179)	(0.229)
HML	0.399***	-0.284	0.683**	0.202	0.131	0.071	0.472***	-0.289	0.760**
	(0.133)	(0.226)	(0.317)	(0.154)	(0.198)	(0.292)	(0.142)	(0.259)	(0.357)
UMD	0.080*	0.020	0.059	-0.036	-0.087	0.051	0.083*	-0.007	0.090
	(0.045)	(0.067)	(0.090)	(0.062)	(0.061)	(0.098)	(0.046)	(0.081)	(0.104)
STR	-0.203**	0.090	-0.294	0.044	-0.072	0.115	-0.151**	0.253	-0.404**
	(0.082)	(0.178)	(0.202)	(0.093)	(0.100)	(0.143)	(0.064)	(0.163)	(0.185)
LTR	-0.128	0.013	-0.140	-0.025	-0.169	0.143	-0.123	-0.221	0.097
	(0.101)	(0.178)	(0.208)	(0.128)	(0.202)	(0.271)	(0.112)	(0.212)	(0.288)
LIQ	-0.195***	0.168*	-0.363***	0.085	0.197**	-0.112	-0.203***	0.140	-0.343**
	(0.070)	(0.095)	(0.123)	(0.076)	(0.087)	(0.119)	(0.072)	(0.098)	(0.132)
Adj R^2	0.795	0.732	0.262	0.838	0.808	0.071	0.806	0.749	0.344
N months	99	99	99	90	90	90	90	90	90

Table 3.11: Conditional factor model estimation: robustness

This table reports conditional factor model performance estimation of portfolios sorted by social sensitivity. We estimate social sensitivity in three alternative ways: (i) in Columns 1 to 3, we estimate social sensitivity by equation (3.2) to account for the Carhart (1997) four factors, (ii) in Columns 4 to 6, we define D_{social} using expanding window medians to address look-ahead bias, and (iii) in Columns 7 to 9, we combine methods use in (i) and (ii) to account for more factors and look-ahead bias at the same time. We focus on the 48 Fama and French (1997) industries. The factor models include the following factors: the market excess return (MKTRF), the size factor (SMB), the value factor (HML), the momentum factor (UMD), short-term reversal (STR) factor, and long-term reversal (LTR) factor. We interact each factor with one of the following interaction variables (INT): the recession indicator from the National Bureau of Economic Research (REC), the Lettau and Ludvigson (2004) *cay* measure, the dividend yield of the value-weighted CRSP index (DIV), the yield on the 90-day T-bill (YLD), the term spread (TERM), and the default spread (DEF). Columns (1) to (6) report the performance difference between the Top and Bottom portfolios. The interaction variable used in each regression is indicated at the top of each column. The Top (Bottom) portfolio is a value-weighted portfolio of the five industries with the most positive (negative) social sensitivity in the past 18 months. The sample period is from October 2005 to December 2013. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Multi-factor sensitivity

Interaction variable (INT)	REC (1)	cay (2)	DIV (3)	YLD (4)	TERM (5)	DEF (6)
Alpha	1.041***	1.097***	0.988**	1.040***	1.029***	1.185***
	(0.372)	(0.350)	(0.430)	(0.357)	(0.344)	(0.380)
Adj R^2	0.254	0.349	0.304	0.265	0.253	0.301
N months	99	99	99	99	99	99

Panel B: Expanding window sensitivity

Interaction variable (INT)	REC (1)	cay (2)	DIV (3)	YLD (4)	TERM (5)	DEF (6)
Alpha	1.125**	1.008**	1.504***	1.609***	1.602***	1.165**
	(0.462)	(0.408)	(0.426)	(0.434)	(0.391)	(0.452)
Adj R^2	0.111	0.139	0.090	0.148	0.130	0.129
N months	90	90	90	90	90	90

Panel C: Combined

Interaction variable (INT)	REC (1)	cay (2)	DIV (3)	YLD (4)	TERM (5)	DEF (6)
Alpha	0.982***	0.786**	1.422**	1.346***	1.255**	1.078**
	(0.370)	(0.387)	(0.543)	(0.478)	(0.497)	(0.482)
Adj R^2	0.447	0.471	0.367	0.253	0.291	0.409
N months	90	90	90	90	90	90

Table 3.12: Social sensitivity and expected returns: robustness

This table reports estimates from Fama Macbeth (1973) regressions. We focus on returns of value-weighted Fama and French (1997) 48 industry portfolios. The dependent variable is monthly industry portfolio return. Regressors include lagged social sensitivity loading, industry-level Fama and French (1992) three-factor loadings estimated by daily returns over the previous month, cumulated industry return over the past six months (*Lag 6m Return*), lagged value-weighted log market capitalization of industry-firms (*Size*), value-weighted book-to-market ratio of industry-firms using book-to-market ratio in the previous year (*Book-to-market*). We estimate social sensitivity in three alternative ways: (i) in Columns 1 to 3, we estimate social sensitivity by equation (3.2) to account for the Carhart (1997) four factors, (ii) in Columns 4 to 6, we define D_{social} using expanding window medians to address look-ahead bias, and (iii) in Columns 7 to 9, we combine methods use in (i) and (ii) to account for more factors and look-ahead bias at the same time. We report the time-series averages of monthly cross-sectional regression coefficients. The sample period is from October 2005 to December 2013. Standard errors (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Multi-factor sensitivity (1)	Expanding window sensitivity (2)	Combined (3)
Social sensitivity	0.086**	0.131***	0.072*
	(0.040)	(0.047)	(0.043)
Beta MKTRF	-0.539	-0.549	-0.567
	(0.480)	(0.528)	(0.534)
Beta SMB	-0.307	-0.459	-0.391
	(0.265)	(0.289)	(0.282)
Beta HML	0.001	0.018	-0.033
	(0.152)	(0.165)	(0.164)
Lag 6m Return	0.009	0.007	0.006
	(0.013)	(0.014)	(0.014)
Size	-0.066	-0.068	-0.099
	(0.086)	(0.088)	(0.089)
Book-to-market	-0.477	-0.622	-0.643*
	(0.355)	(0.380)	(0.360)
Constant	1.847*	1.893*	2.271**
	(0.998)	(1.008)	(1.030)
Average R^2	0.396	0.391	0.393
Observations	4,752	4,320	4,320

Figure 3.1: Search volume intensity for CSR

This figure plots the time-series search volume intensity (*SVI*) for the topic "corporate social responsibility" in the U.S. region from January 2004 to December 2013. Source: Google Trends.

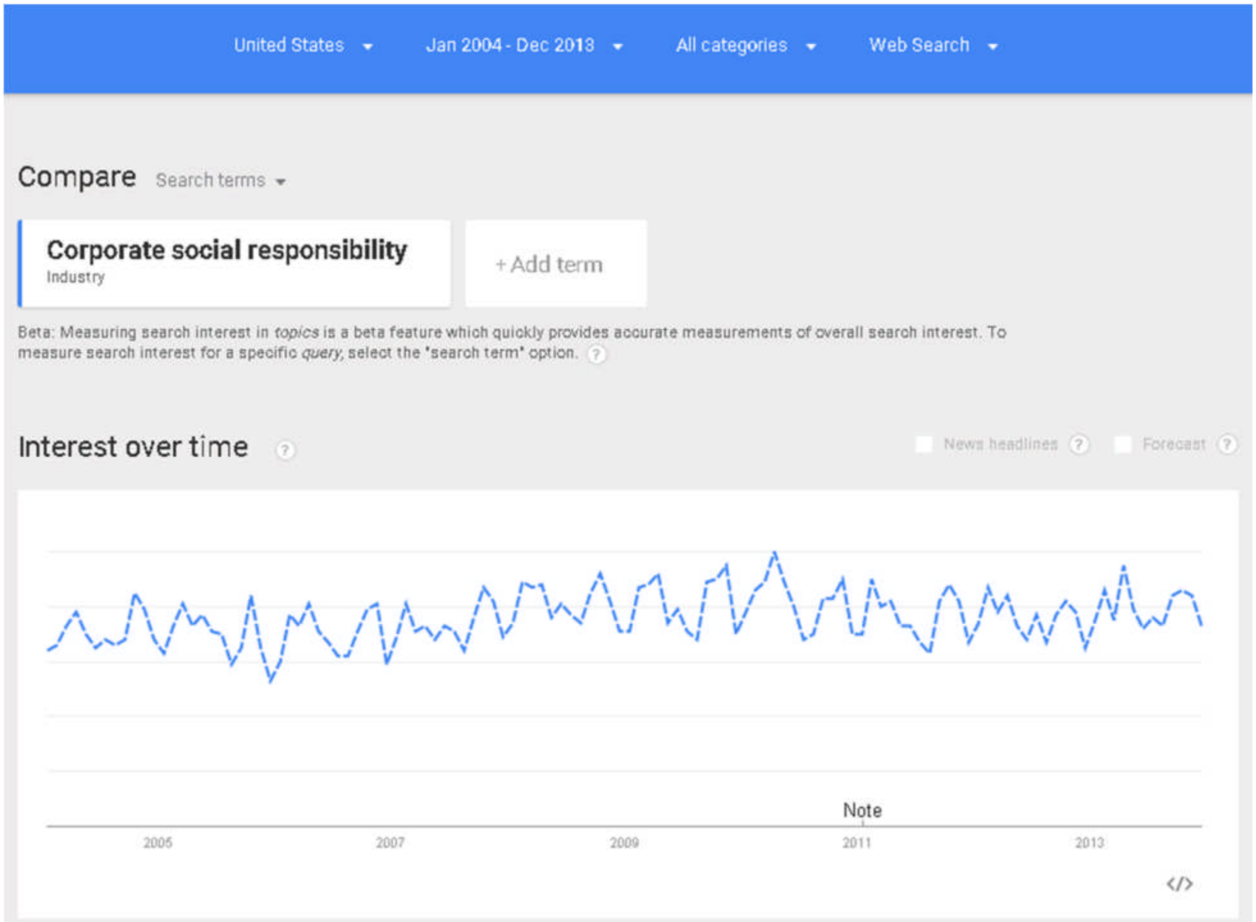


Figure 3.2: Sample CSR newsletter

This figure reports a sample "corporate social responsibility weekly recap" newsletter from Business Wire. Source: Factiva.

Corporate Social Responsibility Weekly Recap (October 27 -- November 3, 2010)

828 words
4 November 2010
09:08
Business Wire
BWR
English

NEW YORK--(BUSINESS WIRE)--November 04, 2010--

The following releases focus on Corporate Social Responsibility and moved during the week of October 27 -- November 3, 2010.

Follow the CSR Circuit newsfeed on www.twitter.com/BWCSRNews

ALL TIME-OFFS ARE IN EASTERN TIME

ABU DHABI, UAE -- 2011 $2.2 Million (USD) Zayed Future Energy Prize Receives Record Number of Submissions (November 03, 2010 09:10 AM) Source: CNBC

NEW YORK -- Alcoa's Vanessa Lau Honored with Woman of Distinction "Badge" from Girl Scout Council of New York (November 03, 2010 09:00 AM) Source: Alcoa Inc.

EL SEGUNDO, Calif. -- Ban on Petroleum-Based Plastic Bags Driving Rapid Growth for Cereplast Bioplastic Resin in Europe (November 03, 2010 08:30 AM) Source: Cereplast, Inc.

BEIJING -- Bank of America Charitable Foundation Provides $2 Million to Support Forest Conservation and Address Global Climate Change (November 03, 2010 05:00 AM) Source: Bank of America

EL SEGUNDO, Calif. -- Cereplast, Inc. to Present at the Stifel Nicolaus Cleantech Conference in New York City on November 9, 2010 (November 02, 2010 03:52 PM) Source: Cereplast, Inc.

EL SEGUNDO, Calif. -- Millions of Children in Need Positively Impacted around the World, Mattel Philanthropic Efforts Continue to Make a Meaningful Difference (November 01, 2010 03:04 PM) Source: Mattel, Inc.

CHARLESTON, S.C. -- Plug In Carolina Selects Eaton to Supply and Support Electric Car "Re-fueling" Stations Throughout South Carolina (November 01, 2010 02:21 PM) Source: Eaton Corporation

BOCA RATON, Fla. -- Office Depot Named America's Greenest Large Retailer in Newsweek Magazine's Annual Green Ranking (November 1, 2010 08:45 AM) Source: Office Depot

BEDFORD, Mass. -- Spire Named New England's 5th Largest Cleantech Employer (October 29, 2010 11:00 AM) Source: Spire Corporation

MINNEAPOLIS -- MoneyGram International Donates $200,000 for Pakistan Disaster Relief (October 29, 2010 05:00 AM) Source: MoneyGram

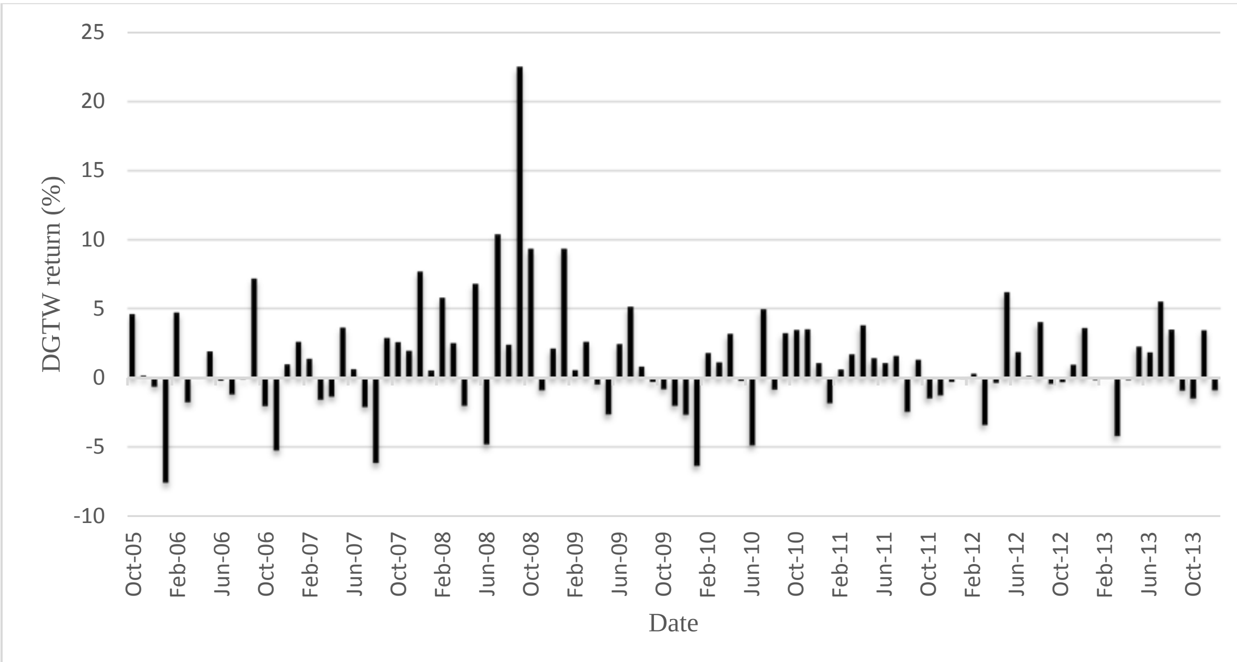
NEW YORK -- Ethisphere Announces 'Attorneys Who Matter' for 2010 (October 28, 2010 03:18 PM) Source: Ethisphere Institute

Figure 3.3: Relation between CSR search volume and CSR news volume

This figure plots the standardized log volume for CSR related news (blue line) and standardized logsvi for the topic “corporate social responsibility” (red line) during the March 2008 to December 2013 period.

Figure 3.4: Social sensitivity based industry portfolios: DGTW returns

This figure plots the DGTW returns of the social sensitivity based Top-Bottom portfolio formed using value-weighted Fama and French (1997) 48 industry portfolios. The sample period is from October 2005 to December 2013.



Chapter 4

Socially Sensitive Fund Flows

4.1. Introduction

Early studies on mutual funds suggest that investors chase performance. Cash flows into and out of mutual funds are strongly correlated with past returns (e.g., Chevalier and Ellison, 1997; Sirri and Tufano, 1998). In addition, Berk and Green (2004) predict that rational investors are likely to use past fund performance to update their beliefs about fund managers' ability to generate excess returns, which also implies a positive relation between fund flows and past performance.

However, recent studies show that investors also care about mutual fund attributes. Among others, Kumar, Niessen-Ruenzi, and Spalt (2015) document that investors are likely to avoid funds managed by individuals with foreign-sounding names even if these funds do not have inferior return performance. In addition, to pursue social objectives, mutual fund investors are willing to accept lower risk-adjusted returns (e.g., Renneboog, Ter Horst, and Zhang, 2008). Further, fund attributes could also influence the flow-return sensitivity of mutual funds (e.g., Bollen, 2007; Renneboog, Ter Horst, and Zhang, 2011).

Motivated by these studies, in this paper, we investigate whether potential stereotypes associated with a fund's social attributes affect the investment decision of mutual fund investors. Since socially responsible investing (SRI) has become an important criterion to mutual fund investors, we ask whether investors are more likely to invest in funds that are perceived to have good social attributes. In addition, we also examine whether perceived social attributes influence flow-return sensitivity. Our key conjecture is that funds perceived to have good social attributes attract higher flows even if these funds do not have superior return performance. In addition, we posit that these funds are likely to be rewarded more following good performance and punished less following the bad performance.

We test our conjectures using a novel method to identify the social attributes of each fund. We are not interested in name-based or social screen-based SRI funds since these funds only account for a very small market segment of the U.S. equity fund universe.[46] Instead, we want to capture the social attributes perceived by the market. As investors care about social attributes (e.g., Bollen, 2007; Hong and Kacperczyk, 2009; Benabou and Tirole, 2010), their perception of fund-level social attributes might affect their investment choices.

Specifically, we estimate the social sensitivity of each fund with respect to the shifts in social sentiment on a 24 months rolling basis. We measure the market-level social sentiment using Google's search volume intensity (*SVI*) for the *topic* corporate social responsibility (CSR). These social sensitivity estimates capture the perceived social attributes of each fund. Funds with positive social sensitivity estimates are likely to be perceived to have good social attributes. This identification strategy is motivated by recent studies which show that Google search could reveal market-level sentiment (Da, Engelberg, and Gao, 2015), and return sensitivity could help to identify stocks and industries that are in line with the market sentiment (Addoum and Kumar, 2016).

The main finding of this paper is that monthly flows are around 0.1 percent higher for funds in the top (i.e., most positive) social sensitivity quintile. In addition, compared to equity funds with comparable characteristics, top socially sensitive funds experience 0.19 percentage lower outflows when their recent performance is in the bottom decile of all mutual funds and 0.46 percentage higher inflows per month when their recent performance is in the top decile. We observe these flow differences even if top socially sensitive funds and other funds have similar performance.

In additional, the test results of fund flows during the recent financial crisis also support the conjecture that perceived social attributes affect fund flows. As investors have stronger social sentiment during economic downturns, the financial crisis could negatively impact the cash flows of mutual funds with poor perceived social attributes. We find that funds in the bottom social sensitivity quintile experience a 0.17% to 0.25% drop in flows relative to funds with moderate social sensitivity during the recent financial crisis.

Overall, these finding suggest that investors prefer funds with good perceived social attributes. They invest more money when these funds have good performance and

[46] For example, the widely used SRI fund list published by the U.S. Forum for Sustainable and Responsible Investment (USSIF) only includes about 40 U.S. equity funds.

withdraw less money when these funds have poor performance. Our evidence is consistent with the conjecture that the perceived social attributes of mutual funds could be an important driver of differences in flow patterns.

Taken together, our findings contribute the growing finance literature that examines investor behavior and financial performance related to socially responsible investing. In the spirit of Markowitz (1952), early literature on SRI has typically focus on comparing return performance of SRI and conventional portfolios (e.g., Hamilton, Jo, Statman, 1993; Statman, 2000; 2005; Bauer, Koedijk, and Otten, 2005; Dwewall, Guenster, Bauer, and Koedijk, 2005; 2011; and Geczy, Stambaugh, and Levin, 2005). We extend the evidence from the previous literature by mainly focusing on whether social attributes influence the investment choices of mutual fund investors. We show that investors are more likely to invest in funds with good perceived social attributes in the U.S. market.

More recently, Bollen (2007) and Renneboog, Ter Horst, and Zhang (2011) show that flows of SRI funds are less sensitive to poor past performance. Rather than focusing on SRI funds, we extend their findings by examining the impact of perceived social attributes on all U.S. equity funds. We show that funds with good social attributes are not only punished less following the bad performance but also on average attract more flows even if these funds do not have superior return performance.

Beyond the literature on SRI, our results also contribute to a broader literature on mutual funds. Previous studies have shown that fund attributes such as fund names and manager names could influence the investment decisions of mutual fund investors. For example, investors are more likely to invest in mutual funds with hot investment style names (Cooper, Gulen, and Rau, 2006) but less likely to invest in mutual funds that are managed by individuals with foreign sounding names (Kumar, Niessen-Ruenzi, and Spalt, 2015). We contribute to this literature by relating the investment decision of mutual fund investors to social attributes.

This paper is organized as follows. Section 4.2 reviews related literature and develops our key hypotheses. Section 4.3 discusses our datasets and methodology. Section 4.4 presents the empirical results, and Section 4.5 concludes.

4.2. Hypothesis development

Mutual fund literature suggests that investors chase past performance. Cash flows into and out of mutual funds are strongly correlated with past returns. Among others, Chevalier and Ellison (1997) find that investors react strongly to risk adjusted returns and fund managers frequently adjust the risk exposure to ensure the attractiveness of their funds. In addition, Sirri and Tufano (1998) show that consumers base their investment decisions on past performance information but do so asymmetrically. They tend to invest disproportionally more in funds with good performance and are less sensitive to poor past performance. Further, Lynch and Musto (2003) predict that investors are less sensitive to poor past performance because funds with poor past performance are likely to discard strategies that underperform. They show that strategy changes only occur after bad performance. Collectively, early studies on mutual funds show little evidence that investors are likely to pay attention to fund attributes that are unrelated to performance.

In contrast, recent studies show that investors' social objectives could override their financial objectives (e.g., Hong and Kacperczyk, 2009). Specifically, socially responsible investing (SRI) has become an important attribute for mutual fund investors when considering new investment opportunities. According to the U.S. Forum for Sustainable and Responsible Investment (USSIF), mutual funds are one of the fastest growing SRI segments in the U.S. For example, the total net asset of SRI funds in the U.S. has increased from $641 billion to $1.93 trillion from 2012 to 2014. SRI is an investment discipline that focuses on environmental, social, and corporate governance (ESG) criteria to generate positive social impact. Therefore, financial performance could be less important for SRI investors because they derive nonfinancial attributes from their investment.

Consistent with this view, Bollen (2007) shows that the behavior of SRI fund investors is different from that of conventional funds. As SRI investors derive utility from investing in stocks that are consistent with their personal values or address social concerns, they are likely to care less about the financial performance. He shows that SRI investors are more sensitive to past positive returns. In addition, Renneboog, Ter Horst, and Zhang (2011) also find similar results using international data.

Despite the fact that SRI funds has accounted for more than 12% of the total net assets of all mutual funds in the U.S., the existing literature has typically focused on a small segment of the mutual fund universe.[47] For example, the widely used USSIF list of SRI funds only contains about 40 equity funds, which accounts for less than 2% of the U.S. equity fund universe.

To generalize the flow-return sensitivity results documented in the relatively small SRI fund universe, in this paper, we propose a novel measure to identify fund-level social attributes for all U.S. equity funds. This identification strategy would allow us to examine the behavior of SRI investors in a much broader setting. In particular, we assign all U.S. equity funds to a corresponding quintile based on their return sensitivity to the attention of corporate social responsibility.

If investors value social attributes, they are likely to find mutual funds that are perceived to have good social attributes in the recent past attractive. Therefore, these funds are likely to attract more flows after controlling for other fund characteristics that are known to explain fund flows. In addition, since these investors care less about financial performance, funds with good perceived social attributes are likely to be punished less following poor performance. Meanwhile, compared to funds that are otherwise similar, funds with good perceived social attributes are likely to be rewarded more following good performance since investors derive both financial and social utilities from investing in these funds.

Further, the recent literature suggests that mutual fund investors value social attributes more during financial crisis. Among others, Nofsinger and Varma (2014) find that SRI funds outperform conventional funds during financial crisis since SRI funds are less likely to suffer from litigation costs related to environmental pollution, product quality, and employee safety. Such litigation costs are more likely to occur among conventional funds and are more costly during financial crisis. Therefore, we conjecture that mutual fund investors are likely to avoid funds with poor social attributes during crisis periods.

To summarize, this paper has the follow three testable hypotheses:

H1: Mutual funds in the top perceived social attributes quintile have higher flows than funds that are otherwise similar.

[47] According to the 2015 investment company fact book, the total net asset of all mutual funds in the U.S. is $15.9 trillion at the end of 2014. Source: https://www.ici.org/pdf/2015_factbook.pdf.

H2: Investors are likely to reward mutual funds with good perceived social attributes more following good performance, but punish these funds less following poor performance.

H3: Mutual funds with poor perceived social attributes are likely to experience lower flows during financial crisis.

4.3. Data and method

To test our hypotheses, we collect data from various sources. In this section, we describe our main data sets and also describe our measure of the social sensitivity of mutual funds.

4.3.1. Mutual fund data

Our main data source for mutual fund is the CRSP Survivor Bias-Free Mutual Fund Database. We obtain fund returns, expenses, loads, total net asset (*TNA*), investment objectives, and other fund characteristics from this dataset. We combine funds with different share classes into a single fund using the MFLINKS files from the Wharton Research Data Services. The main unit of fund analysis is the "wficn" identifier from MFLINKS. Since CRSP reports each share class as a separate series (i.e., "fundno"), we aggregate funds with multiple fundnos into a single wficn. Specifically, we calculate fund-level *TNA* as the sum of *TNAs* of its individual share classes and calculate fund age and objective using the characteristics of the oldest share class. For all other characteristics, we take the weighted average using the one-month lagged *TNAs* of individual fundnos as the weights.

We restrict our sample to U.S. domestic equity mutual funds. We use the investment objective code provided by CRSP (i.e., crsp_obj_cd) to define our sample and fund segments, as in Doshi, Elkamhi, and Simutin (2015). In particular, we screen investment objective codes and fund names to exclude international, balanced, sector, bond, money market, hedge, and index funds. Following Elton, Gruber, and Blake (2001), we exclude funds with less than $15 million *TNA* because returns on small funds tend to be biased upward in the CRSP database. We also remove the first eighteen months of returns on each fund to mitigate the effect of incubator bias documented by Evans (2010). Finally, we exclude funds with missing names in CRSP, as in Cremers and Petajisto (2009) and Amihud and Goyenko (2013). Our final sample contains 1,858 distinct funds during the 2004 to 2015 period, and our sample also includes all U.S. SRI equity funds.

Following Sirri and Tufano (1998), we define fund flow, the main variable of interest in our paper, as follows:

$$Flow_{i,t} = (TNA_{i,t} - TNA_{i,t-1})/TNA_{i,t-1} - r_{i,t}. \tag{4.1}$$

In particular, $Flow_{i,t}$ denotes money flow of fund *i* at the end of month *t*. $TNA_{i,t}$ denotes fund *i*'s total net asset at the end of month *t* and $r_{i,t}$ denotes fund *i*'s return (net of fees) in month *t*, as reported in CRSP. To eliminate the effect of outliers, we winsorize flow observations by 0.5% at each tail. Our results remain similar if we winsorize fund flows by 1% at each tail.

4.3.2. Estimating social sensitivity of mutual funds

We estimate the social sensitivity of each mutual fund. The estimation of fund-level social sensitivity is motivated by the specifications used in Santa-Clara and Valkanov (2003) and Addoum and Kumar (2016). Specifically, in each month, for each mutual fund, we regress the excess fund return during the past two years on the market risk premium and a Corporate Social Responsibility (CSR) attention indicator. Specifically, we estimate the following time-series regression:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_i (r_{mkt} - r_f)_t + \theta_i D_{social,t} + \epsilon_{i,t}. \tag{4.2}$$

In this equation, we define the CSR attention indicator variable (D_{social}) using the search volume intensity (*SVI*) reported by Google Trends. Specifically, we use the *SVI* in the U.S. region for the topic "Corporate Social Responsibility" to measure the overall CSR attention in the country. We use the log difference between *SVIs* in the contemporaneous and previous months to capture shifts in overall CSR attention, as in Da, Engelberg, and Gao (2011, 2015). The CSR attention indicator variable (D_{social}) is equal to one if the log difference of *SVIs* in the previous month is above the 75th percentile of the time-series during the 2004 to 2015 period, or zero otherwise.[48]

[48] We choose this particular model to ensure the consistency with the return sensitivity model used in Chapter 3 since Chapters 3 and 4 are related research.

The θ_i estimate captures the fund-level sensitivity to CSR attention. In particular, a positive θ_i suggests that the fund earns higher average returns in months with high CSR attention relative to its return in months with low CSR attention. In contrast, a negative θ_i estimate indicates that the fund has better return performance during months with low CSR attention. Since investors tend to be net buyers of attention grabbing stocks and such net purchase could drive up stock prices (Barber and Odean, 2008; Da, Engelberg, and Gao, 2011), mutual funds with better returns during high attention periods are likely to hold stocks that are perceived to be socially responsible by investors.

We estimate θ_i by rolling window regressions to allow the magnitude and direction of θ_i to vary over time. Using θ_i estimates, in each month, we sort mutual funds in descending order and create funds quintiles. The top social sensitivity quintile includes funds that are most favored by the market in months with high CSR attention, while bottom social sensitivity quintile includes funds that are least favored by the market during the same periods. We re-estimate θ_i and re-construct mutual fund quintiles on a monthly basis.[49]

4.3.3. Attributes of social sensitivity sorted mutual funds

Table 4.1 reports the summary statistics of mutual fund quintiles defined by our social sensitivity estimates. The most striking feature of this table is that flows are 0.2% higher for funds with top social sensitivity when compared to those of funds with bottom social sensitivity. However, the higher flows into the top socially sensitive funds cannot be explained by higher flows into the same segment as the average segment flow is actually lower for top socially sensitive funds compared to that of funds in the bottom social sensitivity quintile. While flows are higher for funds with top social sensitivity, there is no economically or statistically significant difference among other fund characteristics that are known to explain fund flows, including fund risk, expense ratio, management fee, fund age, or the load versus no load feature. Overall, results in Table 4.1 show that investors have a preference for mutual funds with good perceived social attributes in the recent past.

Figure 4.1 plots the monthly flow difference between funds in the top and bottom quintiles. Panel A plots the equal-weighted flow difference. We find that funds in the top

[49] Our results remain similar if we form social sensitivity based terciles rather than quintiles.

social sensitivity quintile have higher flows in 65% of the sample period. In addition, the flow difference is widened in the recent financial crisis period from December 2007 to June 2009, as defined by the National Bureau of Economic Research (NBER). This evidence is consistent with the conjecture that investors have higher CSR demand during the financial crisis which in turn could enlarge the flow difference between funds with good and bad perceived social attributes. Our results are not sensitive to different weighting methods as the value-weighted flow difference in Panel B are similar to that in Panel A. Further, our findings also remain similar when we examine the flow difference between funds in the top social sensitivity quintile and all the remaining funds in the CRSP universe. For brevity, we do not report those results.

Table 4.1 also reports the name attributes of the social sensitivity sorted funds. We examine the proportion of social-style names in the five quintiles relative to the total number social style names in the CRSP universe. In particular, we screen the following keywords in mutual fund names: “Calvert”, “Sentinel”, “Pax”, “Sustainable”, Responsible, “Social”, “Socially responsive”, “Ethical”, “ESG”, “Women”, “Environmental”, “Water”, “Energy solution”, or “Tobacco free”). In particular, Calvert, Sentinel, and Pax are mutual fund companies that only have socially responsible funds. Other keywords are selected by screening through all equity fund names in the CRSP mutual fund database as well as the names of SRI funds listed by USSIF. We find that the proportion of social-style names increases monotonically from the bottom (14.45%) to the top (29.32%) quintiles. This evidence suggests that the perceived social attributes of mutual funds captured by our social sensitivity measure are consistent with the conventional measure of SRI funds.

4.4. Empirical results

4.4.1. Fund flow

In this section, we present the first set of our main results. Our main objective is to examine whether investors’ social sentiment affect flows of mutual funds with different social sensitivity. Our key conjecture is that fund characteristics, such as social attributes, are likely to affect investment decisions of mutual fund investors. If investors value CSR and derive utility from social attributes, then they are likely to skew their investments more toward funds in the top social sensitivity quintile, which are perceived as being socially responsible during the recent period. To test this conjecture, we start by examining the relation between social sensitivity and mutual fund flows.

Specifically, we estimate flow regressions using the monthly fund flows as the dependent variable. The main independent variable is the D_{top} dummy variable, which equals to one when a fund is located in the top social sensitivity quintile, or zero otherwise. Likewise, we also include the D_{bottom} dummy variable, which indicates whether a fund is located in the bottom social sensitivity quintile.

Our primary set of control variables includes (i) fund size (the natural logarithm of the fund's *TNA* in million USD in the previous month), (ii) fund risk (the standard deviation of monthly returns in the past twelve months), (iii) fund age (the natural logarithm of fund age in years), (iv) segment flow, defined as the growth rate of fund *i*'s market segment (i.e., funds with the same CRSP investment objective code) due to flows in month *t*, excluding flows in fund *i*, (v) family flow, defined as the growth rate of fund *i*'s fund family due to flows in month *t*, excluding flows in fund *i*, and (vi) a loaded fund dummy variable, which indicates whether a fund has front or rear load. In addition, we also account for expense ratio, management fee and turnover in previous month and control for fund returns using performance rank and squared performance rank measures. Performance rank is defined as the rank of the fund in the previous twelve months relative to all other funds in the same market segment.[50] We estimate the flow regressions as a panel and include combinations of year, segment, segment × year, and fund family fixed effects. For robustness, we also estimate Fama and MacBeth (1973) regressions. We cluster standard errors by fund and by date. The specification used is indicated at the top of each column. FMB refers to Fama MacBeth regression.

Table 4.2 presents the flow regression estimates. Consistent with our conjecture, we find that the D_{Top} variable is significantly positively correlated with fund flows for all panel regression specifications. All else equal, fund flows are around 0.095% to 0.131% higher for funds in the top social sensitivity quintile. In addition, the Fama and MacBeth regression yields similar results. Compared to the monthly average flow of -0.473% during our sample period, the magnitude of these estimates are economically meaningful.

[50] Our results remain quantitatively similar if we use raw return to account for past performance.

In contrast, flows of funds in the bottom social sensitivity quintile do not have significantly lower flows when compared to the middle three quintiles. This finding suggests that investors do not perceived funds in the bottom quintile and those in the three middle quintiles to have significantly different social attributes.

In particular, specifications (1) and (3) include year and segment fixed effects. In addition, after accounting for segment × year fixed effects in specification (5), we find that even within the same segment and year, funds with top social sensitivity receive more money flows. This evidence suggests that the higher flows into top social sensitivity funds are not driven by unobservable differences at the segment level. Further, we also control for fund family fixed effects in specification (6). We find that the results remain similar, which suggest that our findings are not explained by the time-invariant unknown heterogeneity at the fund family level. Collectively, results in Table 4.2 show that investors have a preference for funds with good social reputation and these funds attract more flows than funds that are otherwise comparable.

4.4.2. Performance of funds sorted by social sensitivity

Results in Table 4.2 suggest that funds with better social reputation attract more flows. One potential explanation for the higher flow among top social sensitivity funds could be that these funds have better performance during our sample period. To rule out this alternative explanation, we investigate the return performance of our social sensitivity sorted quintile portfolios. In particular, we examine both value-weighted and equal-weighted returns of mutual fund portfolios.

Panel A of Table 4.3 reports the raw portfolio return, return standard deviation, and the Sharpe ratio for the five quintile portfolios sorted by social sensitivity. Columns (1) to (3) report the equal-weighted results. We find that raw returns are similar among the five quintile portfolios. The return difference between the top and bottom quintiles is statistically insignificantly. In addition, the Sharpe ratios are also similar across the five portfolios. Further, Columns (4) to (6) show that the results remain quantitatively similar using value-weighted returns. Figure 4.2 plots the monthly raw return difference between the top and bottom social sensitivity quintiles. Consistent with results in Panel A of Table 4.3, we find that the return difference is evenly distributed around zero.

Our results so far are based on raw returns. In the next set of tests, we use factor models to account for risks. In particular, we include the following factors: the market (*MKTRF*), size (*SMB*), and value (*HML*) factors in Fama and French (1992), the momentum factor (*MOM*) in Carhart (1997), the short-term and long-term reversal factors (*STR*, *LTR*), and the liquidity factor (*LIQ*). We obtain the liquidity factor from Lubos Pastor's website [51] and obtain other factors from Kenneth French's website.[52] Standard errors are calculated using the Newey and West (1987) method.

Panel B of Table 4.3 reports the alpha estimates of factor model regressions. We focus on the difference between the top and bottom quintile portfolios. Consistent with Panel A, we find that there is no significant performance difference in any of the specifications considered in Panel B. Further, in unreported tests, we also examine the alpha estimate of a strategy that goes long in the top quintile and goes short in the remaining four quintiles. The results remain quantitatively similar.

Collectively, results in Table 4.3 suggest that a more positive fund-level social sensitivity is not associated with better risk-adjusted return. Therefore, if investors prefer funds with top social sensitivity, such preference is unlikely to be driven by the difference in fund performance.

4.4.3. Flow-return sensitivity

In the next set of tests, we examine the flow-return relation among top social sensitivity funds. In particular, we test a more specific conjecture, which posits that funds with top social sensitivity are likely to be "rewarded" more after a good performance and "punished" less after a bad performance. This conjecture is motivated by the existing literature which suggests that investors with fund attribute preferences react more strongly to extreme past returns (both positive and negative) than to average past returns (e.g., Kumar, Niessen-Ruenzi, and Spalt, 2015) while SRI investors care less about past performance (e.g., Bollen, 2007; Renneboog, Ter Horst, and Zhang, 2011).

To test this conjecture, we include several interactions in our baseline flow regressions. Specifically, we interact D_{top} and D_{bottom} with performance rank variables (i.e., PRank and

---

[51] The liquidity factor is the main variable in Pastor and Stambaugh (2003) and is available at: http://faculty.chicagobooth.edu/lubos.pastor/research/.

[52] Available at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Prank^2), respectively. In addition, we also interact performance rank variables with fund size to account for the differences in the flow-return relation driven by fund size.

The flow-return regression results are reported in Table 4.4. In specifications (1) and (2), we interact D_{top} and D_{bottom} with PRank to examine the linear flow-return relation. We find that the interaction term between D_{top} and PRank is significantly positive, regardless of the inclusion of lagged fund flow as an additional control. These results suggest that compared to other domestic equity funds, funds in the top social sensitivity quintile attract more cash inflows following the good performance.

In addition, the mutual fund literature suggests that the flow-return relation is likely to be nonlinear (e.g., Sirri and Tufano, 1998). Therefore, we include squared performance measure PRank^2 to capture the nonlinear effect. The results are presented in Column (3) of Table 4.4. Consistent with our conjecture, we find that the flow-return relation is more salient for funds with extreme return ranks.

To better illustrate the nonlinearity, Figure 4.3 plots the difference in fund flows between funds in the top quintile and those in the three middle quintiles, conditional on performance ranks implied by the estimates in specification (3) of Table 4.4. The figure displays a U shape. This evidence is consistent with our conjecture that, following the good performance, investors reward top social sensitivity funds more by investing more capital, but punish these funds less following bad performance through withdrawing less capital. In economic terms, the monthly flow difference due to fund-level social sensitivity is 0.19% for the worst performing funds and 0.46% for the best performing funds.

Columns (4) to (6) of Table 4.4 present the results for three additional robustness checks. First, we examine the flow-return relation using Fama and MacBeth (1973) regression. As implied by the estimates in specification (4), the monthly flow difference is 0.16% for funds with the worst performance and 0.50% for funds with the best performance. In specification (5), to ensure that our results are not driven by segment-level flow dynamics, we replace the segment and year fixed effects with segment × year fixed effects. Further, in specification (6), we account for fund family fixed effects. Our results remain quantitatively similar across all the three alternative specifications. Overall, evidence in Table 4.4 suggests that investors have a more tolerant attitude toward funds with top social sensitivity.

4.4.4. Fund flow during financial crisis

Next, we investigate whether business cycles affect the money flows of mutual funds with different social sensitivity.[53] The recent literature suggests that investors have higher CSR demand during the financial crisis (e.g., Nofsinger and Varma, 2014)). In addition, negative corporate events attract more attention during the financial crisis (Hirshleifer, 2008). Motivated by the fact that investors are likely to divest from funds with poor social attributes during economic downturns, we examine whether funds with bottom social sensitivity are likely to have lower flows during the financial crisis.

Table 4.5 reports the results. The main variables of interest are the interaction term between D_{top} and a financial crisis indicator D_{crisis}, as well as the interaction term between D_{bottom} and D_{crisis}. According to the definition of NBER, D_{crisis} is equal to one during the December 2007 to June 2009 period, or zero otherwise. We use the same set of control variables as in Table 4.2. Consistent with our conjecture, we find that funds in the bottom social sensitivity quintile experience significant reductions in flows during the recent financial crisis. The monthly average flows of these funds reduce by 0.165% to 0.252%, depending on the specification. As investors pay more attention to negative corporate events during financial crisis (e.g., Hirshleifer, 2008; Shefrin and Statman, 1993) and the occurrence of negative corporate events are costlier for these investors during crisis periods, they are likely to avoid funds in the bottom social sensitivity quintile during economic downturns to achieve better risk-adjusted returns.

In contrast, D_{top} is significantly positive in all specifications while the interaction variable between D_{top} and D_{crisis} is statistically insignificant. This is expected since investors with social norms derive social utility from investing in funds with good social attributes. Therefore, their investment decisions are less affected by economic conditions.

[53] One potential limitation of this empirical analysis is that we have a short sample period. However, our definition of economic recession is based on the NBER business cycle definition. In addition, our sample covers months before 2006, a period that is commonly regarded as an economic expansion. Further, our sample also includes the post 2007 financial crisis period. Therefore, even though we only have one crisis during our sample periods, our sample also has two economic upturns.

Overall, our evidence in Table 4.5 is consistent with the view that investors with increased CSR demand are more likely to avoid funds with poor social attributes.

4.4.5. Robustness checks

In this section, we address the potential look-ahead bias in our baseline specification. In particular, to ensure a longer sample, we use the 75th percentile of *ASVI* of the full sample period to define D_{social} in our baseline regressions, which raises the concern of potential look-ahead bias in our social sensitivity measure. We address the potential look-ahead bias using an alternative definition of D_{social}. In particular, we re-define D_{social} using expanding-window method and re-estimate equation 4.2. Under this alternative measure, D_{social} equals to one if *ASVI* in the previous month is above the 75th percentile of all previous observations, or zero otherwise. For example, we use the 75th percentile of all *ASVI* observations in 2004 to define our first D_{social} in January 2005 so that D_{social} only relies on past information. We re-estimate our flow regressions (i.e., Table 4.2) and flow return sensitivity regressions (i.e., Table 4.4) using this alternative social sensitivity estimate.

Table 4.6 reports the flow regression results. We find that even with a shorter sample, the regression estimates of D_{social} remain quantitatively similar for both panel and Fama and MacBeth (1973) regressions. Therefore, results in Table 4.6 suggest that the higher flows into top social sensitivity funds are not explained by look-ahead bias. Mutual funds attract more flows if they are perceived to have good social attributes in the recent past.

Table 4.7 presents the flow-return sensitivity results using the alternative social sensitivity estimates. Again, we find that our results are quantitatively similar. As implied by the estimates using Fama and MacBeth (1973) regression, the monthly flow difference is 0.19% for the worst performers and 0.46% for the best performers. In addition, panel regressions also yield similar results. Therefore, results in Table 4.6 also suggest that investors tend to reward funds with good perceived social attributes more following good return performance but punish these funds less following poor return performance.

4.5. Summary and Conclusion

This study investigates how perceived social attributes affect mutual fund flows. Using social sensitivity as a novel measure to proxy for social attributes of all U.S. equity funds, we show that investors are more likely to invest in funds with top social sensitivity even though these funds do not have better return performance. This evidence suggests that social attributes provide additional utility to mutual fund investors.

In addition, consistent with the flow-return relation documented among SRI funds, we demonstrate that investors reward top social sensitivity funds more following good performance and punish these funds less following the bad performance, compared to funds that are otherwise similar. These results remain robust after accounting for fund-level characteristics, segment-level flow dynamics, and fund family fixed effects. Taken together, our findings show that besides financial performance, mutual fund investors also care about social attributes of mutual funds. In addition, their preference for funds with good perceived social attributes becomes stronger during periods with the high social sentiment.

Table 4.1: Social sensitivity and fund characteristics

This table reports the monthly mean fund characteristics sorted by our social sensitivity estimates θ_i, with Column (1) (Column (5)) reports the characteristics of funds in the bottom (top) social sensitivity quintile. The θ_i estimate captures the return sensitivity of each fund to overall CSR attention in the U.S. *Social sensitivity* reports the mean of θ_i estimates. *Fund flow* is the net change in fund asset beyond asset appreciation in the current month defined as $(TNA_{i,t} – TNA_{i,t-1})/TNA_{i,t-1} – r_{i,t}$, where $TNA_{i,t}$ denotes fund *i*'s total net asset in month *t*, and $r_{i,t}$ denotes fund *i*'s return in month *t* as reported in CRSP. *PRank* is the performance rank of the fund in the previous twelve months relative to all other funds with the same CRSP investment objective code, scaled to lie between zero (worst performance) and one (best performance). *Fund size* is the natural logarithm of the fund's size in million USD in the previous month. *Fund risk* is the standard deviation of monthly returns in the previous twelve months. *Expense ratio*, *management fee*, and *turnover* are fund *i*'s expense ratio, management fee, and turnover rate in the previous month. *Fund age* is the natural logarithm of fund age in years in month *t*. *Segment flow* is the growth rate of fund *i*'s market segment (i.e., funds with the same CRSP investment objective code) due to flows in month *t*, excluding flows in fund *i*. *Family flow* is the growth rate of fund *i*'s fund family due to flows in month *t*, excluding flows in fund *i*. *Loaded fund* is a dummy variable that equals to one when the fund has front load or rear load, or zero otherwise. *No. of funds* reports the monthly average number of funds in each quintile. We also report the name attributes of quintile funds. *Social* indicates the percentage of social-style names in a quintile relative to all social-style names in the sample while social-style names contain at least one of the following keywords: "Calvert", "Sentinel", "Pax", "Sustainable", "Responsible", "Social", "Socially responsive", "Ethical", "ESG", "Women", "Environment", "Water", "Energy solution", or "Tobacco free". The sample period is from April 2006 to December 2015.

Variable	Bottom (1)	2 (2)	3 (3)	4 (4)	Top (5)
Fund attributes					
Social sensitivity	-1.055	-0.450	-0.085	0.287	0.855
Fund flow	-0.580	-0.465	-0.481	-0.462	-0.375
PRank	0.469	0.495	0.501	0.513	0.522
PRank^2	0.314	0.324	0.330	0.344	0.360
Fund size	6.057	6.087	6.095	6.025	5.928
Fund risk	4.697	4.439	4.377	4.436	4.602
Expense ratio	0.096	0.092	0.092	0.094	0.098
Management fee	0.060	0.056	0.056	0.056	0.058
Turnover	6.035	6.153	6.172	6.636	6.853
Fund age	2.548	2.550	2.577	2.600	2.597
Segment flow	-0.188	-0.185	-0.232	-0.272	-0.339
Family flow	-0.389	-0.350	-0.295	-0.307	-0.316
Loaded fund	0.709	0.702	0.705	0.701	0.713
No. of funds	251	252	252	252	251
Fund name					
Social	14.452	15.851	16.534	23.839	29.324

Table 4.2: Flow regression estimates

This table shows the estimates of percentage fund flows regressed on the top and bottom social sensitivity dummy variables D_{top} and D_{bottom} and various control variables. *Fund flow* is the net change in fund asset beyond asset appreciation in the current month defined as $(TNA_{i,t} - TNA_{i,t-1})/TNA_{i,t-1} - r_{i,t}$, where $TNA_{i,t}$ denotes fund i's total net asset in month t, and $r_{i,t}$ denotes fund i's return in month t as reported in CRSP. D_{top} (D_{bottom}) is equal to one if a fund has the most positive (negative) return sensitivity to CSR attention during the past two years, or zero otherwise. All control variables, except for segment and family flows are lagged by one month and have been defined in Table 4.1. Specification (4) applies Fama and MacBeth (1973) estimation method with Newey and West (1987) standard errors. All other specifications use pooled OLS regressions with standard error clustered by fund and by date. Specifications (1) to (3) include segment and year fixed effects while specification (6) also includes fund family fixed effects. In specification (5), we include segment × year fixed effects to account for segment-level flow dynamics. The sample period is from April 2006 to December 2015. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) OLS	(2) OLS	(3) OLS	(4) FMB	(5) OLS	(6) OLS
D_{top}	0.131***	0.103***	0.095**	0.117***	0.110***	0.106***
	(0.050)	(0.037)	(0.037)	(0.028)	(0.039)	(0.038)
D_{bottom}	-0.007	0.004	-0.013	0.018	-0.023	-0.015
	(0.050)	(0.036)	(0.037)	(0.032)	(0.038)	(0.037)
PRank	2.913***	2.109***	1.003***	0.923***	0.998***	0.950***
	(0.110)	(0.078)	(0.217)	(0.176)	(0.218)	(0.213)
$PRank^2$			1.107***	1.012***	1.114***	1.193***
			(0.217)	(0.169)	(0.218)	(0.214)
Fund size	0.096***	0.039***	0.039***	0.053***	0.041***	0.009
	(0.016)	(0.011)	(0.011)	(0.008)	(0.011)	(0.015)
Fund risk	-0.097***	-0.071***	-0.076***	-0.158***	-0.079***	-0.060***
	(0.024)	(0.019)	(0.019)	(0.024)	(0.019)	(0.020)
Expense ratio	-1.887*	-1.675**	-1.897***	-0.168	-1.902***	-4.444***
	(0.969)	(0.741)	(0.734)	(0.626)	(0.737)	(0.889)
Management fee	-0.359***	-0.248***	-0.253***	-2.790***	-0.252***	-0.250***
	(0.127)	(0.090)	(0.090)	(0.609)	(0.094)	(0.085)
Turnover	-0.003	-0.003	-0.003*	-0.005*	-0.003*	-0.002
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)
Fund age	-0.563***	-0.364***	-0.365***	-0.312***	-0.370***	-0.393***
	(0.054)	(0.038)	(0.038)	(0.023)	(0.038)	(0.044)
Segment flow	0.418***	0.355***	0.356***	-31.269***	0.308***	0.376***
	(0.047)	(0.042)	(0.042)	(1.195)	(0.044)	(0.043)
Family flow	0.239***	0.212***	0.212***	0.176***	0.213***	0.181***
	(0.015)	(0.014)	(0.014)	(0.010)	(0.014)	(0.014)
Load	0.074	0.065	0.067	0.025	0.066	0.205***
	(0.061)	(0.044)	(0.044)	(0.024)	(0.044)	(0.067)
Lagged fund flow		0.291***	0.291***	0.266***	0.289***	0.278***
		(0.010)	(0.010)	(0.007)	(0.010)	(0.010)
Year FE	Yes	Yes	Yes	No	No	Yes
Segment FE	Yes	Yes	Yes	Yes	No	Yes
Segment × year FE	No	No	No	No	Yes	No
Fund family FE	No	No	No	No	No	Yes
Adj/ave. R^2	0.080	0.159	0.159	0.240	0.160	0.166
Observations	145,762	145,762	145,762	145,762	145,762	145,762

Table 4.3: Performance of funds sorted by social sensitivity estimates

This table presents the fund performance estimates. In Panel A, we report equal-weighted and value-weighted raw return (in percentage) of mutual fund quintiles sorted by social sensitivity, the standard deviation of raw return (*Std. Dev*), and the Sharpe ratio. In Panel B, we regress the monthly difference in the returns between top and bottom social sensitivity quintile portfolios on the market, *SMB*, *HML*, *UMD*, *STR*, *LTR* and *LIQ* factors. Standard error (reported in parentheses) are adjusted for auto-correlation using the Newey and West (1987) method. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Raw returns of mutual funds sorted by social sensitivity

	Equal-weighted (Columns 1-3)			Value-weighted (Columns 4-6)		
	(1)	(2)	(3)	(4)	(5)	(6)
Quintile	Return	Std. Dev	Sharpe ratio	Return	Std. Dev	Sharpe ratio
Bottom	0.536	5.072	0.089	0.568	5.019	0.096
	(0.536)			(0.536)		
2	0.596	4.823	0.106	0.653	4.785	0.118
	(0.506)			(0.506)		
3	0.595	4.697	0.108	0.623	4.570	0.117
	(0.492)			(0.484)		
4	0.610	4.673	0.112	0.648	4.603	0.122
	(0.484)			(0.483)		
Top	0.588	4.689	0.107	0.596	4.642	0.110
	(0.489)			(0.492)		
5 - 1	0.052	1.286	-0.027	0.028	1.349	-0.043
	(0.123)			(0.126)		
N months						107

Table 4.3 (Cont'd)

Panel B: Factor model estimates: Top – bottom

	Equal-weighted (Columns 1 – 4)				Value-weighted (Columns 5 – 8)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	CAPM	3-Factor	4-Factor	7-Factor	CAPM	3-Factor	4-Factor	7-Factor
Alpha	0.099	0.050	0.046	0.029	0.078	0.017	0.011	0.000
	(0.115)	(0.120)	(0.119)	(0.121)	(0.116)	(0.117)	(0.113)	(0.113)
MKTRF	-0.080***	-0.021	-0.014	-0.009	-0.085***	-0.014	-0.002	-0.003
	(0.030)	(0.029)	(0.028)	(0.034)	(0.028)	(0.027)	(0.026)	(0.033)
SMB		-0.171**	-0.172**	-0.187**		-0.192***	-0.193***	-0.198***
		(0.082)	(0.082)	(0.087)		(0.070)	(0.070)	(0.074)
HML		-0.159***	-0.140**	-0.102		-0.205***	-0.175***	-0.121
		(0.048)	(0.054)	(0.070)		(0.048)	(0.053)	(0.075)
MOM			0.026	0.022			0.042	0.040
			(0.027)	(0.028)			(0.030)	(0.030)
STR				-0.022				0.006
				(0.039)				(0.039)
LTR				0.001				-0.045
				(0.057)				(0.059)
LIQ				0.050**				0.038
				(0.024)				(0.023)
Observations	117	117	117	117	117	117	117	117
Adj. R^2	0.071	0.231	0.232	0.233	0.072	0.289	0.302	0.303

Table 4.4: Flow-return relation of top and bottom socially sensitive funds

This table reports the estimates of percentage fund flows regressed on the top and bottom social sensitivity dummy variables D_{top} and D_{bottom} interacted with lagged performance measures. We use the same specifications as in Table 4.2 and add interactions terms of D_{top} and D_{bottom} with performance variables. *Fund flow* is the net change in fund asset beyond asset appreciation in the current month defined as $(TNA_{i,t} - TNA_{i,t-1})/TNA_{i,t-1} - r_{i,t}$, where $TNA_{i,t}$ denotes fund *i*'s total net asset in month *t*, and $r_{i,t}$ denotes fund *i*'s return in month *t* as reported in CRSP. D_{top} (D_{bottom}) is equal to one if a fund has the most positive (negative) return sensitivity to CSR attention during the past two years, or zero otherwise. All control variables, except for segment and family flows are lagged by one month and have been defined in Table 4.1. Specification (4) applies Fama and MacBeth (1973) estimation method with Newey and West (1987) standard errors. All other specifications use pooled OLS regressions with standard error clustered by fund and by date. Specifications (1) to (3) include segment and year fixed effects while specification (6) also includes fund family fixed effects. In specification (5), we include segment × year fixed effects to account for segment-level flow dynamics. The sample period is from April 2006 to December 2015. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) OLS	(2) OLS	(3) OLS	(4) FMB	(5) OLS	(6) OLS
D_{top}	-0.098	-0.057	0.185*	0.163*	0.195**	0.163*
	(0.092)	(0.069)	(0.095)	(0.084)	(0.097)	(0.097)
D_{bottom}	-0.119	-0.075	-0.077	-0.034	-0.080	-0.072
	(0.078)	(0.058)	(0.082)	(0.079)	(0.082)	(0.080)
D_{top} × *PRank*	0.444**	0.311**	-1.148***	-1.028***	-1.105**	-0.987**
	(0.174)	(0.130)	(0.443)	(0.388)	(0.442)	(0.443)
D_{bottom} × *PRank*	0.231	0.162	0.039	0.110	0.033	0.025
	(0.159)	(0.122)	(0.404)	(0.376)	(0.409)	(0.405)
D_{top} × $PRank^2$			1.422***	1.360***	1.371***	1.284***
			(0.450)	(0.387)	(0.449)	(0.447)
D_{bottom} × $PRank^2$			0.155	0.002	0.144	0.149
			(0.432)	(0.375)	(0.436)	(0.437)
$PRank^2$			2.504***	2.655***	2.536***	2.452***
			(0.771)	(0.607)	(0.778)	(0.762)
PRank	3.380***	2.508***	0.104	1.094*	0.072	0.234
	(0.303)	(0.222)	(0.773)	(0.628)	(0.778)	(0.765)
Fund size	0.146***	0.080***	0.042*	0.165***	0.043*	0.020
	(0.026)	(0.019)	(0.025)	(0.021)	(0.025)	(0.027)
PRank × fund size	-0.102**	-0.083**	0.181	-0.029	0.185*	0.146
	(0.044)	(0.032)	(0.111)	(0.087)	(0.112)	(0.112)
$PRank^2$ × *fund size*			-0.280**	-0.293***	-0.283**	-0.252**
			(0.111)	(0.084)	(0.112)	(0.112)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	No	No	Yes
Segment FE	Yes	Yes	Yes	Yes	No	Yes
Segment × year FE	No	No	No	No	Yes	No
Fund family FE	No	No	No	No	No	Yes
Adj/ave. R^2	0.080	0.159	0.159	0.248	0.160	0.166
Observations	145,762	145,762	145,762	145,762	145,762	145,762

Table 4.5: Flow regression estimates: financial crisis

This table shows the estimates of fund flow regressions around the 2007 financial crisis. The financial crisis dummy variable D_{crisis} is equal to one during the December 2007 to June 2009 period. The dependent variable and other independent variables are the same as in Table 4.2. All specifications use pooled OLS regressions with standard error (reported in parentheses) clustered by fund and by date. Specifications (1) to (3) include segment and year fixed effects while specification (5) also includes fund family fixed effects. In specification (4), we include segment × year fixed effects to account for segment-level flow dynamics. The sample period is from April 2006 to December 2015. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
D_{top}	0.112**	0.095**	0.088**	0.098**	0.103**
	(0.055)	(0.041)	(0.041)	(0.043)	(0.042)
D_{bottom}	0.036	0.032	0.018	0.011	0.015
	(0.054)	(0.039)	(0.039)	(0.040)	(0.040)
$D_{top} \times D_{crisis}$	0.107	0.051	0.039	0.067	0.020
	(0.103)	(0.080)	(0.080)	(0.086)	(0.081)
$D_{bottom} \times D_{crisis}$	-0.252**	-0.165**	-0.180**	-0.189**	-0.171**
	(0.108)	(0.084)	(0.084)	(0.085)	(0.085)
D_{crisis}	-0.018	0.033	0.038	0.021	0.028
	(0.125)	(0.122)	(0.123)	(0.131)	(0.122)
PRank	2.902***	2.102***	0.989***	0.984***	0.937***
	(0.109)	(0.078)	(0.217)	(0.217)	(0.213)
$PRank^2$			1.113***	1.120***	1.200***
			(0.217)	(0.218)	(0.213)
Fund size	0.095***	0.039***	0.039***	0.041***	0.009
	(0.016)	(0.011)	(0.011)	(0.011)	(0.015)
Fund risk	-0.096***	-0.070***	-0.075***	-0.079***	-0.060***
	(0.024)	(0.019)	(0.019)	(0.019)	(0.020)
Expense ratio	-1.884*	-1.670**	-1.892***	-1.896**	-4.425***
	(0.970)	(0.741)	(0.734)	(0.737)	(0.889)
Management fee	-0.360***	-0.248***	-0.252***	-0.252***	-0.250***
	(0.128)	(0.091)	(0.091)	(0.095)	(0.085)
Turnover	-0.003	-0.003	-0.003*	-0.003*	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Fund age	-0.563***	-0.365***	-0.365***	-0.370***	-0.394***
	(0.054)	(0.038)	(0.038)	(0.038)	(0.044)
Segment flow	0.417***	0.355***	0.356***	0.308***	0.375***
	(0.047)	(0.041)	(0.042)	(0.044)	(0.043)
Family flow	0.238***	0.212***	0.212***	0.213***	0.181***
	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)
Load	0.074	0.065	0.068	0.067	0.206***
	(0.060)	(0.044)	(0.044)	(0.044)	(0.067)
Lagged fund flow		0.291***	0.291***	0.289***	0.278***
		(0.010)	(0.010)	(0.010)	(0.010)
Year FE	Yes	Yes	Yes	No	Yes
Segment FE	Yes	Yes	Yes	No	Yes
Segment × year FE	No	No	No	Yes	No
Fund family FE	No	No	No	No	Yes
Adj. R^2	0.080	0.159	0.159	0.160	0.166
Observations	145,762	145,762	145,762	145,762	145,762

Table 4.6: Flow regression estimates: alternative specification

This table reports the estimates of percentage fund flows regressed on the top and bottom social sensitivity dummy variables D_{top} and D_{bottom} and various control variables. We focus on the last four specifications in Table 4.2 and use an alternative definition for social sensitivity to address look ahead bias. Specifically, we use the expanding-window median to re-define D_{social} and re-estimate social sensitivity for each fund. *Fund flow* is the net change in fund asset beyond asset appreciation in the current month defined as $(TNA_{i,t} - TNA_{i,t-1})/TNA_{i,t-1} - r_{i,t}$, where $TNA_{i,t}$ denotes fund i's total net asset in month t, and $r_{i,t}$ denotes fund i's return in month t as reported in CRSP. D_{top} (D_{bottom}) is equal to one if a fund has the most positive (negative) return sensitivity to CSR attention during the past two years, or zero otherwise. All control variables, except for segment and family flows are lagged by one month and have been defined in Table 4.1. Specification (2) applies Fama and MacBeth (1973) estimation method with Newey and West (1987) standard errors. All other specifications use pooled OLS regressions with standard error clustered by fund and by date. Specification (1) includes segment and year fixed effects while specification (4) also includes fund family fixed effects. In specification (3), we include segment × year fixed effects to account for segment-level flow dynamics. The sample period is from January 2007 to December 2015. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) OLS	(2) FMB	(3) OLS	(4) OLS
D_{top}	0.075*	0.113***	0.100**	0.087**
	(0.039)	(0.030)	(0.041)	(0.040)
D_{bottom}	-0.045	-0.008	-0.065	-0.033
	(0.038)	(0.031)	(0.040)	(0.039)
PRank	1.040***	0.927***	1.032***	1.007***
	(0.227)	(0.184)	(0.228)	(0.224)
$PRank^2$	1.052***	1.007***	1.060***	1.122***
	(0.224)	(0.177)	(0.225)	(0.221)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes
Segment FE	Yes	Yes	No	Yes
Segment × year FE	No	No	Yes	No
Fund family FE	No	No	No	Yes
Adj/ave. R^2	0.157	0.240	0.158	0.165
Observations	133,928	133,928	133,928	133,928

Table 4.7: Flow-return relation: alternative specification

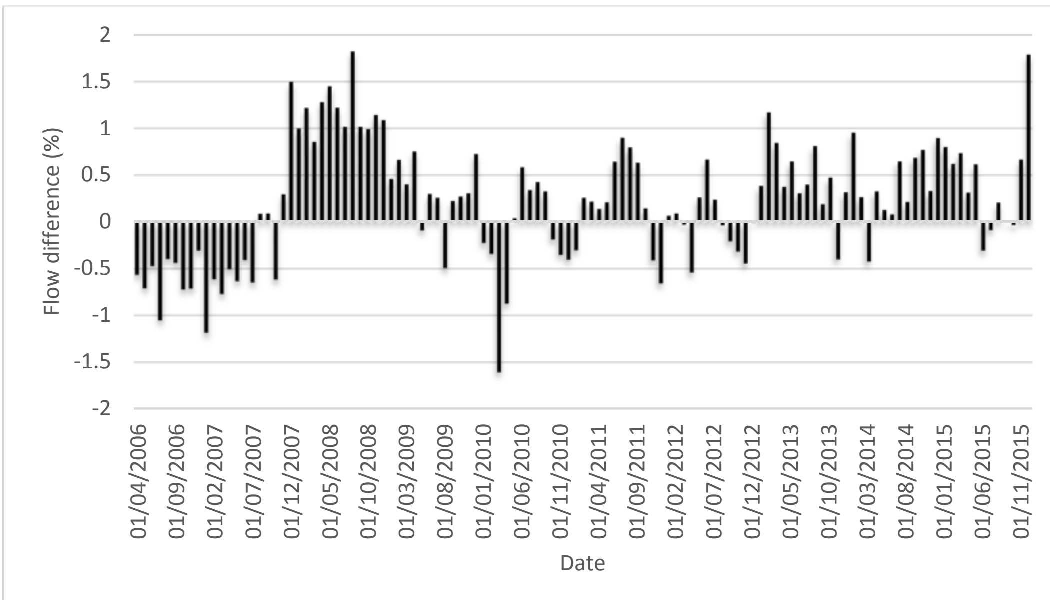
This table reports the estimates of percentage fund flows regressed on the top and bottom social sensitivity dummy variables D_{top} and D_{bottom} interacted with lagged performance measures. We focus on the last four specifications in Table 4.5 and use an alternative definition for social sensitivity to address look ahead bias. Specifically, we use the expanding-window median to re-define D_{social} and re-estimate social sensitivity for each fund. *Fund flow* is the net change in fund asset beyond asset appreciation in the current month defined as $(TNA_{i,t} - TNA_{i,t-1})/TNA_{i,t-1} - r_{i,t}$, where $TNA_{i,t}$ denotes fund i's total net asset in month t, and $r_{i,t}$ denotes fund i's return in month t as reported in CRSP. D_{top} (D_{bottom}) is equal to one if a fund has the most positive (negative) return sensitivity to CSR attention during the past two years, or zero otherwise. All control variables, except for segment and family flows are lagged by one month and have been defined in Table 4.1. Specification (2) applies Fama and MacBeth (1973) estimation method with Newey and West (1987) standard errors. All other specifications use pooled OLS regressions with standard error clustered by fund and by date. Specification (1) includes segment and year fixed effects while specification (4) also includes fund family fixed effects. In specification (3), we include segment × year fixed effects to account for segment-level flow dynamics. The sample period is from January 2007 to December 2015. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1) OLS	(2) FMB	(3) OLS	(4) OLS
D_{top}	0.243**	0.185**	0.264***	0.235**
	(0.097)	(0.086)	(0.099)	(0.100)
D_{bottom}	-0.038	0.001	-0.044	-0.017
	(0.090)	(0.086)	(0.090)	(0.090)
$D_{top} \times PRank$	-1.364***	-1.117***	-1.321***	-1.267***
	(0.440)	(0.372)	(0.437)	(0.439)
$D_{bottom} \times PRank$	-0.234	-0.055	-0.265	-0.262
	(0.440)	(0.414)	(0.443)	(0.446)
$D_{top} \times PRank^2$	1.516***	1.395***	1.463***	1.434***
	(0.439)	(0.359)	(0.436)	(0.434)
$D_{bottom} \times PRank^2$	0.340	0.006	0.344	0.354
	(0.477)	(0.415)	(0.479)	(0.486)
$PRank^2$	2.358***	2.499***	2.386***	2.265***
	(0.801)	(0.622)	(0.808)	(0.794)
PRank	0.295	1.276*	0.269	0.469
	(0.800)	(0.655)	(0.804)	(0.792)
Fund size	0.043	0.165***	0.045*	0.017
	(0.026)	(0.022)	(0.026)	(0.029)
PRank × fund size	0.174	-0.042	0.177	0.137
	(0.116)	(0.091)	(0.116)	(0.117)
$PRank^2$ *× fund size*	-0.276**	-0.278***	-0.278**	-0.246**
	(0.116)	(0.087)	(0.117)	(0.118)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes
Segment FE	Yes	Yes	No	Yes
Segment × year FE	No	No	Yes	No
Fund family FE	No	No	No	Yes
Adj/ave. R^2	0.157	0.247	0.158	0.165
Observations	133928	133928	133928	133928

Figure 4.1: Monthly flow difference between top and bottom socially sensitive funds

This figure shows the monthly average flow difference between the top and bottom socially sensitive funds. *Fund flow* is calculated as: *Fund flow*$_t$ = $(TNA_t - TNA_{t-1})/TNA_{t-1} - r_t$. Top (bottom) socially sensitive funds are funds in the top (bottom) social sensitivity quintile. Panel A plots the equal-weighted flow difference and Panel B plots the value-weighted flow difference, respectively.

Panel A: Equal-weighted monthly flow difference



Panel B: Value weighted monthly flow difference

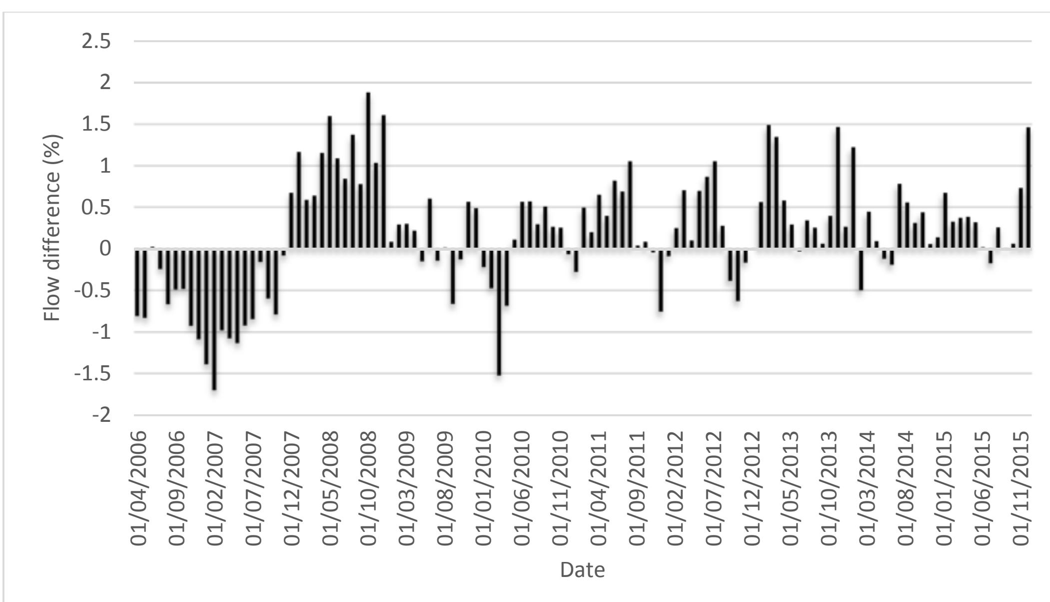
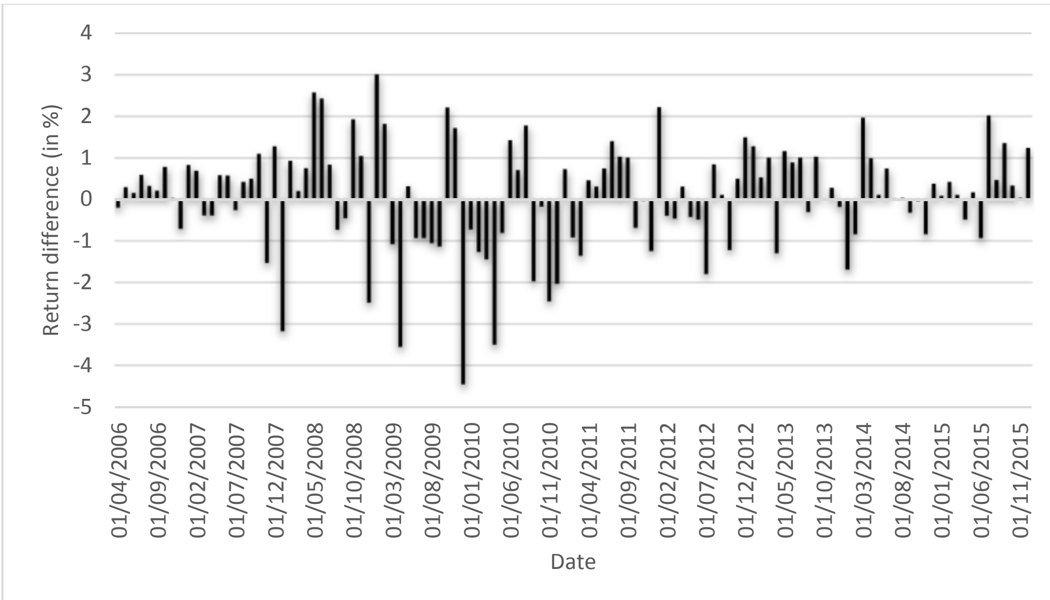


Figure 4.2: Monthly return difference between top and bottom socially sensitive funds

This figure shows the monthly average raw return difference between the top and bottom socially sensitive funds. Top (bottom) socially sensitive funds are funds in the top (bottom) social sensitivity quintile. Panel A plots the equal-weighted return difference and Panel B plots the value-weighted return difference, respectively.

Panel A: Equal-weighted monthly return difference



Panel B: Value-weighted monthly return difference

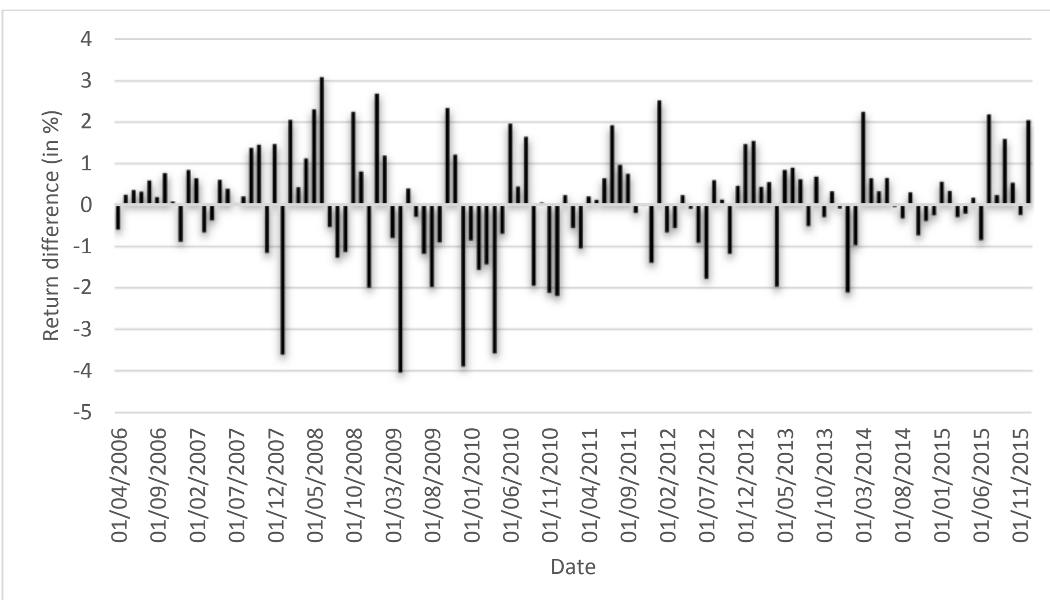
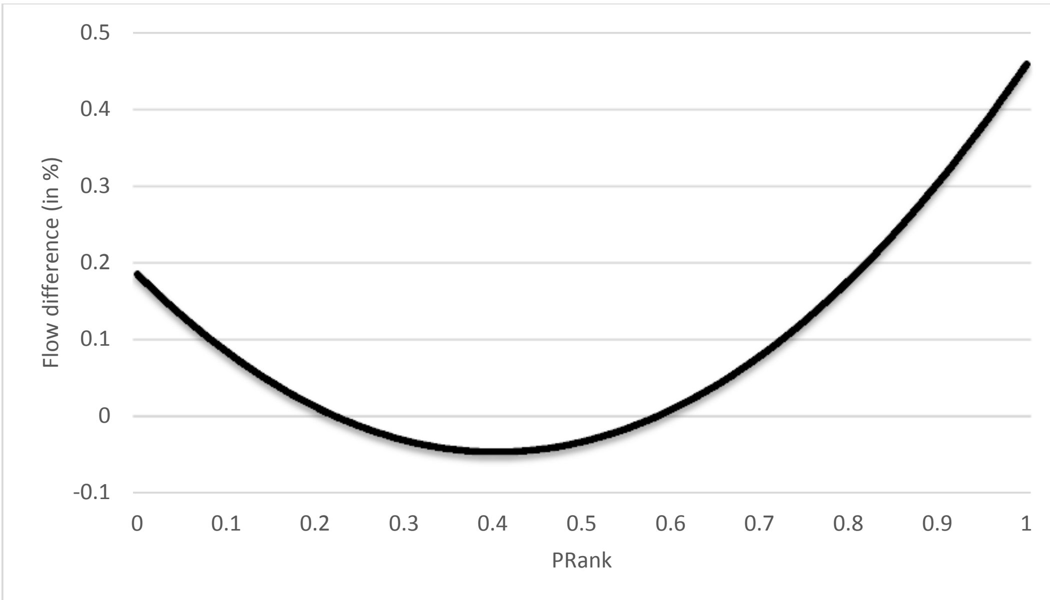


Figure 4.3: Differences in fund flow conditional on performance: Top funds

This figure shows the predicted difference (in percentage points) in fund flow conditional on fund performance in the past twelve months between funds with top social sensitivity and those in the three middle social sensitivity quintiles. The graph is based on specification (3) in Table 4.4. The positive values indicate higher flows into top socially sensitive funds.



Chapter 5

Conclusion

This thesis comprises three essays that investigate the impact of investor attention on financial market outcomes. Chapter 2 investigates how changes in overall attitudes toward gambling affect financial market outcomes. Using a novel measure of gambling sentiment based on lottery-related Internet search volume, we show that the time-variation in gambling attitudes predicts the returns of lottery-like stocks. Further, using attention-grabbing lottery jackpots as exogenous shocks to gambling sentiment, we show that our results do not reflect potential reverse causality. We find that large lottery jackpots not only increase people's participation in lotteries, but also enhance investors' propensity to purchase stocks with lottery-like characteristics. Analyzing trades of retail investors from a major U.S. discount brokerage firm, we also show directly that investors increase aggregate demand for lottery-like stocks around large jackpots and large drawings.

The time-variation in gambling attitudes also affects corporate financial decisions. Specifically, firms with high nominal share prices are more likely to split their shares when investors' gambling sentiment becomes stronger. Stronger gambling sentiment is also associated with higher first-day returns of initial public offerings. Collectively, these results suggest that shifts in overall gambling attitudes have spillover effects on financial markets.

These findings contribute to the growing finance literature that examines the role of gambling in financial markets. Our paper adds a new dimension to this literature by demonstrating that time-variation in gambling attitudes generates short-term mispricing and also affect corporate decisions. Our results also relate to the public economics. Since state lottery is an important source for public funding, practitioners could study the behavior of lottery players using our Google measure and develop their games accordingly to increase lottery sales.

In Chapter 3, we propose a novel measure to identify firms that are likely to be perceived as having good social attributes by the market. Specifically, we use social sensitivity, defined as the return sensitivity to the aggregate attention to CSR, to capture perceived social attributes. We show that social sensitivity is positively correlated with industry- or stock-level CSR records. Using social sensitivity estimates, we find that returns of market segments with high social sensitivity are predictable. A trading strategy that goes long in stocks with good perceived social attributes and goes short in stocks with bad perceived social attributes generates a monthly DGTW return of 1.17%. Our results have great implication for practitioners since we identify a profitable socially responsible investing strategy that could be easily implemented.

Relatedly, using social sensitivity as a novel measure to proxy for social attributes of all U.S. equity funds, in Chapter 4, we show that investors are more likely to invest in funds with top social sensitivity even though these funds do not have better return performance. This evidence suggests that social attributes provide additional utility to mutual fund investors.

In addition, consistent with the flow-return relation documented among SRI funds, we demonstrate that investors reward top social sensitivity funds more following good performance and punish these funds less following the bad performance, compared to funds that are otherwise similar. These results remain robust after accounting for fund-level characteristics, segment-level flow dynamics, and fund family fixed effects. Our social sensitivity identification strategy could also be implemented by practitioners to identify non-SRI funds that are likely to be favored by SRI investors.

Taken together, this thesis shows that investor attention reveals market level sentiment and affects investment choices of both retail and institutional investors. Their trading activity could in turn impact stock prices and corporate decisions.

In future work, it could be interesting to investigate whether investors react differently to firms with different perceived social attributes when the earnings of these firms miss the analyst forecasts. In particular, I conjecture that firms with poor perceived social attributes would have more negative price reaction when their earnings miss analyst forecasts. In addition, another potential project is to examine whether corporate social responsibility engagements are more likely to happen following major CSR related disasters (e.g., BP oil spill) and how do these events affect the success rate of engagements and the return performance around announcement dates.

Bibliography

Aboody, D., Lev, B., 1998. The value relevance of intangibles: The case of software capitalization. *Journal of Accounting Research* 36, 161-191.

Addoum, J.M., Kumar, A., 2016. Political sentiment and predictable returns. *Review of Financial Studies*, forthcoming.

Albuquerque, R.A., Durnev, A., Koskinen, Y., 2016. Corporate social responsibility and firm risk: Theory and empirical evidence. *Unpublished working paper*.

Amihud, Y., Goyenko, R., 2013. Mutual fund's R2 as predictor of performance. *Review of Financial Studies* 26, 667-694.

Angel, J.J., 1997. Tick size, share prices, and stock splits. *Journal of Finance* 52, 655-681.

Antoniou, C., Kumar, A., Maligkris, A., 2016. Terrorist attacks, analyst sentiment, and earnings forecasts. *Unpublished working paper*.

Baker, H.K., Gallagher, P.L., 1980. Management's view of stock splits. *Financial Management* 9, 73-77.

Baker, H.K., Powell, G.E., 1992. Why companies issue stock splits. *Financial Management* 21, 11-11.

Baker, H.K., Powell, G.E., 1993. Further evidence on managerial motives for stock splits. *Quarterly Journal of Business and Economics* 32, 20-31.

Baker, M., Greenwood, R., Wurgler, J., 2009. Catering through nominal share prices. *Journal of Finance* 64, 2559-2590.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. *Journal of Finance* 61, 1645-1680.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. *Journal of Economic Perspectives* 21, 129-152.

Bali, T.G., Cakici, N., Whitelaw, R.F., 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99, 427-446.

Barber, B.M., Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance* 55, 773-806.

Barber, B.M., Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21, 785-818.

Barberis, N., Huang, M., 2008. Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review* 98, 2066-2100.

Bauer, R., Koedijk, K., Otten, R., 2005. International evidence on ethical mutual fund performance and investment style. *Journal of Banking and Finance* 29, 1751-1767.

Benabou, R., Tirole, J., 2010. Individual and corporate social responsibility. *Economica* 77, 1-19.

Ben-Rephael, A., Da, Z., Israelsen, R.D., 2017. It depends on where you search: A comparison of institutional and retail attention. *Review of Financial Studies*, forthcoming.

Berk, J.B., Green, R.C., 2004. Mutual fund flows and performance in rational markets. *Journal of Political Economy* 112, 1269-1295.

Besley, T., Ghatak, M., 2007. Retailing public goods: The economics of corporate social responsibility. *Journal of Public Economics* 91, 1645-1663.

Bollen, N.P.B., 2007. Mutual fund attributes and investor behavior. *Journal of Financial and Quantitative Analysis* 42, 683-708.

Brunnermeier, M.K., Gollier, C., Parker, J.A., 2007. Optimal beliefs, asset prices, and the preference for skewed returns. *American Economic Review*, 159-165.

Calcagno, P., Walker, D., Jackson, J., 2010. Determinants of the probability and timing of commercial casino legalization in the United States. *Public Choice* 142, 69-90.

Campbell, C., Wasley, C., 1996. Measuring abnormal daily trading volume for samples of NYSE/ASE and NASDAQ securities using parametric and nonparametric test statistics. *Review of Quantitative Finance and Accounting* 6, 309-326.

Campbell, J.Y., Hilscher, J., Szilagyi, J.A.N., 2008. In search of distress risk. *Journal of Finance* 63, 2899-2939.

Carhart, M.M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52, 57-82.

Castaldo, S., Perrini, F., Misani, N., Tencati, A., 2009. The missing link between corporate social responsibility and consumer trust: The case of fair trade products. *Journal of Business Ethics* 84, 1-15.

Chae, J., 2005. Trading volume, information asymmetry, and timing information. Journal of Finance 60, 413-442.

Chan, L.K.C., Lakonishok, J., Sougiannis, T., 2001. The stock market valuation of research and development expenditures. *Journal of Finance* 56, 2431-2456.

Chava, S., 2014. Environmental externalities and cost of capital. *Management Science* 60, 2223-2247.

Cheng, I.-H., Hong, H., Shue, K., 2016. Do managers do good with other people's money? *Unpublished working paper*.

Chevalier, J., Ellison, G., 1997. Risk taking by mutual funds as a response to incentives. *Journal of Political Economy* 105, 1167-1200.

Cohen, L., Frazzini, A., 2008. Economic links and predictable returns. *Journal of Finance* 63, 1977-2011.

Cohn, J.B., Wardlaw, M.I., 2016. Financing constraints and workplace safety. *Journal of Finance* 71, 2017-2058.

Cooper, M.J., Gulen, H., Rau, P.R., 2005. Changing names with style: mutual fund name changes and their effects on fund flows. *Journal of Finance* 60, 2825-2858.

Coval, J.D., Moskowitz, T.J., 1999. Home bias at home: Local equity preference in domestic portfolios. *Journal of Finance* 54, 2045-2073.

Cremers, K.J.M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies* 22, 3329-3365.

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. *Journal of Finance* 66, 1461-1499.

Da, Z., Engelberg, J., Gao, P., 2015. The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies* 28, 1-32.

Daniel, K., Grinblatt, M., Titman, S., Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52, 1035-1058.

Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance* 52, 1-33.

DellaVigna, S., Pollet, J.M., 2009. Investor inattention and Friday earnings announcements. *Journal of Finance* 64, 709-749.

Deng, X., Kang, J.-k., Low, B.S., 2013. Corporate social responsibility and stakeholder value maximization: Evidence from mergers. *Journal of Financial Economics* 110, 87-109.

Dimson, E., Karakaş, O., Li, X., 2015. Active ownership. *Review of Financial Studies* 28, 3225-3268.

Doran, J.S., Jiang, D., Peterson, D.R., 2012. Gambling preference and the New Year effect of assets with lottery features. *Review of Finance* 16, 685-731.

Dorn, A.J., Dorn, D., Sengmueller, P., 2015. Trading as gambling. *Management Science* 61, 2376-2393.

Dorn, D., Sengmueller, P., 2009. Trading as entertainment. *Management Science* 55, 591-603.

Doshi, H., Elkamhi, R., Simutin, M., 2015. Managerial activeness and mutual fund performance. *Review of Asset Pricing Studies* 5, 156-184.

Dowell, G., Hart, S., Yeung, B., 2000. Do corporate global environmental standards create or destroy market value? *Management Science* 46, 1059-1074.

Eberhart, A.C., Maxwell, W.F., Siddique, A.R., 2004. An examination of long-term abnormal stock returns and operating performance following R&D increases. *Journal of Finance* 59, 623-650.

Edmans, A., 2011. Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics* 101, 621-640.

Edmans, A., Li, L., Zhang, C., 2016. Employee satisfaction, labor market flexibility, and stock returns around the world. *Unpublished working paper*.

Edward A. Dyl, William B. Elliott, 2006. The share price puzzle. *Journal of Business* 79, 2045-2066.

El Ghoul, S., Guedhami, O., Kwok, C.C.Y., Mishra, D.R., 2011. Does corporate social responsibility affect the cost of capital? *Journal of Banking and Finance* 35, 2388-2406.

Elfenbein, D.W., Fisman, R., Mcmanus, B., 2012. Charity as a substitute for reputation: Evidence from an online marketplace. *Review of Economic Studies* 79, 1441-1468.

Elton, E.J., Gruber, M.J., Blake, C.R., 2001. A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases. *Journal of Finance* 56, 2415-2430.

Evans, R.B., 2010. Mutual fund incubation. *Journal of Finance* 65, 1581-1611.

Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. *Journal of Finance* 47, 427-465.

Fama, E.F., French, K.R., 1997. Industry costs of equity. *Journal of Financial Economics* 43, 153-193.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81, 607-636.

Fernando, C.S., Krishnamurthy, S., Spindt, P.A., 1999. Is share price related to marketability? Evidence from mutual fund share splits. *Financial Management* 28, 54-67.

Fernando, C.S., Krishnamurthy, S., Spindt, P.A., 2004. Are share price levels informative? Evidence from the ownership, pricing, turnover and performance of IPO firms. *Journal of Financial Markets* 7, 377-403.

Ferreira, M.A., Matos, P., 2008. The colors of investors' money: The role of institutional investors around the world. *Journal of Financial Economics* 88, 499-533.

Ferrell, A., Liang, H., Renneboog, L., 2016. Socially responsible firms. *Journal of Financial Economics*, forthcoming.

Gao, X., Lin, T.-C., 2015. Do individual investors treat trading as a fun and exciting gambling activity? Evidence from repeated natural experiments. *Review of Financial Studies* 28, 2128-2166.

Geczy, C., Stambaugh, R.F., Levin, D., 2005. Investing in socially responsible mutual funds. *Unpublished working paper*.

Gervais, S., Kaniel, R., Mingelgrin, D.H., 2001. The high-volume return premium. *Journal of Finance* 56, 877-919.

Ghysels, E., 1998. On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *Journal of Finance* 53, 549-573.

Gompers, P.A., Metrick, A., 2001. Institutional investors and equity prices. *Quarterly Journal of Economics* 116, 229-259.

Green, T.C., Hwang, B.-H., 2012. Initial public offerings as lotteries: Skewness preference and first-day returns. *Management Science* 58, 432-444.

Grinblatt, M., Keloharju, M., 2001. How distance, language, and culture influence stock holdings and trades. *Journal of Finance* 56, 1053-1073.

Grinblatt, M., Keloharju, M., 2009. Sensation seeking, overconfidence, and trading activity. *Journal of Finance* 64, 549-578.

Grote, K.R., Matheson, V.A., 2007. Examining the 'Halo Effect' in lotto games. *Applied Economics Letters* 14, 307-310.

Grullon, G., Kanatas, G., Weston, J.P., 2004. Advertising, breadth of ownership, and liquidity. *Review of Financial Studies* 17, 439-461.

Harvey, C.R., Siddique, A., 2000. Conditional skewness in asset pricing tests. *Journal of Finance* 55, 1263-1295.

Hirshleifer, D., 2008. Psychological bias as a driver of financial regulation. *European Financial Management* 14, 856-874.

Hirshleifer, D., Hsu, P.-H., Li, D., 2013. Innovative efficiency and stock returns. *Journal of Financial Economics* 107, 632-654.

Hirshleifer, D., Lim, S.S., Teoh, S.H., 2009. Driven to distraction: Extraneous events and underreaction to earnings news. *Journal of Finance* 64, 2289-2325.

Hong, H., Kacperczyk, M., 2009. The price of sin: The effects of social norms on markets. *Journal of Financial Economics* 93, 15-36.

Hong, H., Kubik, J.D., Scheinkman, J.A., 2012. Financial constraints on corporate goodness. *Unpublished working paper*.

Hong, H., Kubik, J.D., Stein, J.C., 2008. The only game in town: Stock-price consequences of local bias. *Journal of Financial Economics* 90, 20-37.

Hong, H., Scheinkman, J., Xiong, W., 2006. Asset float and speculative bubbles. *Journal of Finance* 61, 1073-1117.

Hou, K., Xiong, W., Peng, L., 2009. A tale of two anomalies: The implications of investor attention for price and earnings momentum. *Unpublished working paper*.

Ibbotson, R.G., Sindelar, J.L., Ritter, J.R., 1994. The market's problems with the pricing of initial public offerings. *Journal of Applied Corporate Finance* 7, 66-74.

Jeroen, D., Nadja, G., Rob, B., Koedijk, K., 2005. The eco-efficiency premium puzzle. *Financial Analysts Journal* 61, 51-63.

Jeroen, D., Nadja, G., Rob, B., Koedijk, K., 2011. The economic value of corporate eco-efficiency. *European Financial Management* 17, 679-704.

Kearney, M.S., 2005. State lotteries and consumer behavior. *Journal of Public Economics* 89, 2269-2299.

Kitzmueller, M., Shimshack, J., 2012. Economic perspectives on corporate social responsibility. *Journal of Economic Literature* 50, 51-84.

Konar, S., Cohen, M.A., 2001. Does the market value environmental performance? *Review of Economics and Statistics* 83, 281-289.

Korniotis, G.M., Kumar, A., 2013. State-level business cycles and local return predictability. *Journal of Finance* 68, 1037-1096.

Krüger, P., 2015. Corporate goodness and shareholder wealth. *Journal of Financial Economics* 115, 304-329.

Kumar, A., 2009. Who gambles in the stock market? *Journal of Finance* 64, 1889-1933.

Kumar, A., Lee, C.M.C., 2006. Retail investor sentiment and return comovements. *Journal of Finance* 61, 2451-2486.

Kumar, A., Niessen-Ruenzi, A., Spalt, O.G., 2015. What's in a name? Mutual fund flows when managers have foreign-sounding names. *Review of Financial Studies* 28, 2281-2321.

Kumar, A., Page, J.K., Spalt, O.G., 2011. Religious beliefs, gambling attitudes, and financial market outcomes. *Journal of Financial Economics* 102, 671-708.

Kumar, A., Page, J.K., Spalt, O.G., 2016. Gambling and comovement. *Journal of Financial and Quantitative Analysis* 51, 85-111.

Lakonishok, J., Lev, B., 1987. Stock splits and stock dividends: why, who, and when. *Journal of Finance* 42, 913-932.

Lakonishok, J., Shleifer, A., Vishny, R.W., 1992. The impact of institutional trading on stock prices. *Journal of Financial Economics* 32, 23-43.

Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* 56, 815-849.

Liang, H., Renneboog, L., 2016. On the Foundations of Corporate Social Responsibility. *Journal of Finance,* forthcoming.

Lin, J.-C., Singh, A.K., Yu, W., 2009. Stock splits, trading continuity, and the cost of equity capital. *Journal of Financial Economics* 93, 474-489.

Lou, D., 2014. Attracting investor attention through advertising. *Review of Financial Studies* 27, 1797-1829.

Loughran, T., Ritter, J., 2004. Why has IPO underpricing changed over time? *Financial Management* 33, 5-37.

Luo, X., Bhattacharya, C.B., 2006. Corporate social responsibility, customer satisfaction, and market Value. *Journal of Marketing* 70, 1-18.

Lynch, A.W., Musto, D.K., 2003. How Investors Interpret Past Fund Returns. *Journal of Finance* 58, 2033-2058.

Macey, J., O'Hara, M., Pompilio, D., 2008. Down and out in the stock market: The law and economics of the delisting process. *Journal of Law and Economics* 51, 683-713.

Margolis, J.D., Elfenbein, H.A., Walsh, J.P., 2009. Does it pay to be good... and does it matter? A meta-analysis of the relationship between corporate social and financial performance. *Unpublished working paper*.

Markowitz, H., 1952. Portfolio selection. *Journal of Finance* 7, 77-91.

Markowitz, H., 1952. The utility of wealth. *Journal of Political Economy* 60, 151-158.

Masulis, R.W., Reza, S.W., 2015. Agency problems of corporate philanthropy. *Review of Financial Studies* 28, 592-636.

Minnick, K., Raman, K., 2014. Why are stock splits declining? *Financial Management* 43, 29-60.

Mitton, T., Vorkink, K., 2007. Equilibrium underdiversification and the preference for Skewness. *Review of Financial Studies* 20, 1255-1288.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703-708.

Nofsinger, J., Varma, A., 2014. Socially responsible funds and market crises. *Journal of Banking and Finance* 48, 180-193.

Odean, T., 1999. Do investors trade too much? *American Economic Review* 89, 1279-1298.

Palomino, F., Renneboog, L., Zhang, C., 2009. Information salience, investor sentiment, and stock returns: The case of British soccer betting. *Journal of Corporate Finance* 15, 368-387.

Pastor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111, 642-685.

Puckett, A., Yan, X., 2011. The interim trading skills of institutional investors. *Journal of Finance* 66, 601-633.

Renneboog, L., Ter Horst, J., Zhang, C., 2008. The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. *Journal of Corporate Finance* 14, 302-322.

Renneboog, L., Ter Horst, J., Zhang, C., 2011. Is ethical money financially smart? Nonfinancial attributes and money flows of socially responsible investment funds. *Journal of Financial Intermediation* 20, 562-588.

Sally, H., Jo, H., Statman, M., 1993. Doing well while doing good? The investment performance of socially responsible mutual funds. *Financial Analysts Journal* 49, 62-66.

Santa-Clara, P., Valkanov, R., 2003. The presidential puzzle: Political cycles and the stock market. *Journal of Finance* 58, 1841-1872.

Scheinkman, J.A., Xiong, W., 2003. Overconfidence and speculative bubbles. *Journal of Political Economy* 111, 1183-1220.

Scott, F., Garen, J., 1994. Probability of purchase, amount of purchase, and the demographic incidence of the lottery tax. *Journal of Public Economics* 54, 121-143.

Seasholes, M.S., Wu, G., 2007. Predictable behavior, profits, and attention. Journal of Empirical Finance 14, 590-610.

Seasholes, M.S., Zhu, N., 2010. Individual investors and local bias. *Journal of Finance* 65, 1987-2010.

Servaes, H., Tamayo, A., 2013. The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science* 59, 1045-1061.

Shefrin, H., Statman, M., 2009. Behavioral portfolio theory. *Journal of Financial and Quantitative Analysis* 35, 127-151.

Sirri, E.R., Tufano, P., 1998. Costly search and mutual fund flows. *Journal of Finance* 53, 1589-1622.

Statman, M., 2000. Socially responsible mutual funds. *Financial Analysts Journal* 56, 30-39.

Statman, M., 2005. The religions of social responsibility. *Journal of Investing* 14, 14-21.

Weld, W.C., Michaely, R., Thaler, R.H., Benartzi, S., 2009. The nominal share price puzzle. *Journal of Economic Perspectives* 23, 121-42.

Williams, L.V., Siegel, D.S., 2014. The Oxford Handbook of the Economics of Gambling. *Oxford University Press*.

Woolridge, J.R., Chambers, D.R., 1983. Reverse splits and shareholder wealth. *Financial Management* 12, 5-15.

Yuan, Y., 2011. Attention and Trading. *Unpublished working paper*.