Conditional Information Acquisition

Boone Bowles

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Finance.

Chapel Hill
2020

Approved by:

Adam Reed

Robert Connolly

Jesse Davis

Christian Lundblad

John Hand

ABSTRACT

BOONE BOWLES: Conditional Information Acquisition.
(Under the direction of Adam Reed)

When a portion of institutional investors are prohibited from short selling, news that generates differences of opinions also affects information acquisition. Investors facing a short-sale prohibition (e.g., mutual funds) acquire less information when the sentiment of news is positive, as positive sentiment increases the likelihood that they will be unable to trade. Also, prices are more informative following news with negative sentiment than news with positive sentiment. These novel predictions are tested empirically using new measures of information acquisition derived from a hand-collected sample of mutual fund and hedge fund IP addresses. When the sentiment of recent news has been negative instead of positive, information acquisition by mutual funds increases by 16% relative to hedge funds, and prices are up to 14% more informative.

To Shawnee. This is a product of your very hard work. Thank you.

ACKNOWLEDGMENTS

I am extremely grateful to Bob Connolly, Jesse Davis, Christian Lundblad, Adam Reed, and John Hand for their guidance and support. I also thank David Dicks, Paolo Fulghieri, Camelia Kuhnen, Jacob Sagi, and Gil Segal for their valuable comments and feedback.

I also wish to thank seminar participants at The University of North Carolina (Kenan-Flagler Business School), Brigham Young University (Marriott), Texas A&M University (Mays), Michigan State University (Broad), The University of Virginia (Darden), Florida State University, Virginia Tech (Pamblin), Lehigh University (Perella), and The University of Georgia (Terry).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Investors acquire information hoping to profitably trade upon it. However, in many settings, investors face constraints that inhibit their ability to trade on the information they acquire. One constraint commonly faced by institutional investors is a prohibition on short sales: up to 73% of mutual funds face such a restriction.[1] While it is known that short-sale prohibitions generally reduce incentives to acquire information, the conditions under which this prohibition is most likely to bind, and therefore attenuate information acquisition, are less understood.[2] Further, unconstrained investors potentially modify their information acquisition due to short-sale prohibitions faced by others, though this relationship is also unclear. This paper expands our understanding of information acquisition by highlighting an important interaction between short-sale prohibitions and recent news. In particular, information acquisition is conditional on the sentiment of recent news when it generates disagreement amongst investors.

To demonstrate this relationship, this paper extends the classic setting of Grossman and Stiglitz (1980) by (*i*) imposing a short-sale prohibition on a subset of potentially informed investors, and (*ii*) adding an initial stage in which the sentiment of news generates disagreement. From the perspective of potentially informed investors, this sentiment leads prices to predictably diverge from their estimate of the asset's fundamental value. When the sentiment of news leads constrained investors to be relatively pessimistic, they expect to be bound by their short-sale prohibition. Anticipating that they may be unable to trade, fewer

[1] Almazan, Brown, Carlson and Chapman (2004). Using Yahoo's mutual fund screener and accounting for mutual funds with at least $100 million in net assets shows that there are up to 50 times as many mutual funds in the large, mid, and small stock categories as there are in the long-short category. When accounting for funds having at least $500 million or $1 billion in assets, the ratio is as high as 86 and 140.

[2] Nezafat, Schroder and Wang (2017) shows that short-sale constraints adversely affect private information acquisition.

constrained investors become informed. In response, more unconstrained investors acquire information. In summary, the model predicts that news with positive sentiment will induce relatively less information acquisition from mutual funds (the constrained investors) and more from hedge funds (the unconstrained). The model also predicts that prices will be less informative when investors expect to be constrained as aggregate information acquisition is reduced.

To test the predictions from the model, this paper uses a hand-collected dataset that measures information acquisition by mutual funds and hedge funds. This dataset was created using records of activity on the EDGAR filing system, the SEC's online repository for public information. Using a proprietary sample of unmasked IP addresses, I refine the EDGAR data in order to produce measures of information acquisition that distinguish between mutual funds and hedge funds. This provides a unique environment in which to test the predictions from the model. Empirical analysis verifies the model's main predictions by showing that relative to hedge funds, mutual funds request less information from EDGAR when the sentiment of recent news has been positive. Additionally, when more IP addresses from mutual funds and hedge funds are acquiring information, prices more closely reflect fundamental value.

An overview of the underlying theory is useful to develop intuition for these results. A single risky asset exists in random supply. There are three types of investors in the market: mutual funds, hedge funds, and retail. Mutual funds and hedge funds (together referred to as institutional investors) can acquire costly private information prior to trading. Retail investors cannot. On the other hand, mutual funds cannot short sell, while hedge funds and retail investors are unconstrained. All agents maximize expected profits, subject to a quadratic inventory cost, and trade by submitting limit orders in a competitive market.

There are three dates. At time zero, all investors observe a public signal that is orthogonal to the asset's payoff. Retail investors erroneously consider the signal to be informative and use it to update their beliefs about the payoff.[3] This creates a difference of opinions between institutional and retail investors. Subsequent to observing the initial signal, institutional investors endogenously determine whether to acquire costly private information. Finally, investors trade, with institutional investors rationally updating given the price.[4]

Institutional investors consider the impact the initial signal will have on retail investors when deciding whether to become informed. They do this because the signal provides information regarding the likelihood that mutual funds will be bound by the short-sale prohibition. For example, when the initial signal is positive, retail investors believe the payoff will be higher than institutional investors believe it will be. From the perspective of institutional investors, demand from retail investors will, in expectation, push the price of the asset above its fundamental value. Thus, it is likely that mutual funds will be bound by their short-sale prohibition. Faced with the prospect of acquiring information they cannot utilize, fewer mutual funds learn. In response to this, more hedge funds will choose to become informed due to the substitutability of information acquisition. This example highlights the key channel at work in the model: institutional investors condition their information acquisition on the initial signal since it drives disagreement with retail investors, thereby providing valuable information as to whether mutual funds will be bound by their short-sale prohibition.

The model predicts that, relative to hedge funds, fewer mutual funds will acquire information when retail investors are relatively optimistic. Additionally, the model highlights an asymmetry in the substitutability of information acquisition. When deciding whether to acquire information, the chief concern of mutual funds is whether the initial signal will trigger their short-sale prohibition. This relationship holds regardless of the existence or information acquisition of hedge funds. On the other hand, when mutual funds are excluded from the model, information acquisition by hedge funds is invariant to the initial signal. Hedge funds

[3] Similar specifications for retail investors are used in Hirshleifer, Subrahmanyam and Titman (2006), Mendel and Shleifer (2012), Banerjee and Green (2015), and Crouzet, Dew-Becker and Nathanson (2018).

[4] Institutional investors update using insights from Breon-Drish (2015) since the price is non-linear. Consistent with differences of opinions models, the retail investors do not update on the price. This is equivalent to them assuming that institutional investors receive signals that are just noise.

are primarily concerned with the signal to the extent that it influences the information acquisition of mutual funds. Thus, the model predicts that hedge fund information acquisition is more sensitive to that of mutual funds than vice versa.

Since the initial signal (i) provides information about whether mutual funds will be able to trade and (ii) influences both mutual fund and hedge fund information acquisition, the initial signal also influences the information content of prices. At first glance, the relationship between price informativeness and the initial signal is unclear. While a positive signal generates less information acquisition from mutual funds, it simultaneously increases the amount of information acquired by hedge funds. However, mutual funds are more sensitive to the initial signal than are hedge funds. For example, a positive signal will induce a large number of mutual funds to forgo acquiring information, compared to a relatively smaller number of hedge funds who now learn. The added informed hedge funds are unable to impound the same amount of information into the price that the large amount of mutual funds would have. Thus, the model predicts that positive signals generate less informative prices than negative signals.

To empirically test the model's predictions, the EDGAR data has been refined to produce two measures of information acquisition. The first, *Requests*, captures the total number of requests for information from mutual funds about a given stock over a period of time. The second variable, IPs, records the total number of unique IP addresses from mutual funds making requests for information. The resultant panel provides a unique setting in which mutual fund information acquisition is measured at the stock level. Furthermore, as a proxy for the initial signal, this paper uses the event sentiment score provided by RavenPack. This score, which measures the sentiment of news articles, is viewed as a proxy for the initial signal since it is found to be related to both retail trading and the probability that prices will move away from fundamental value.

Consistent with the model, empirical analyses find that when the sentiment of news has been negative mutual funds acquire relatively more information than hedge funds. Specifically, if sentiment has been negative instead of positive, a stock can expect up to 16% more requests from mutual funds relative to hedge funds. Further, if sentiment has been negative, a stock can expect as much as 14% more IPs from mutual funds than hedge funds making requests.

To test implications for price informativeness, this paper exploits earnings announcements and utilizes the price jump ratio, a measure of price informativeness developed in Weller (2017). The price jump ratio proxies for price informativeness by measuring price movements in the days leading up to and including earnings announcements. The intuition behind this measure is that earnings announcements will result in small price movements when the price is informative. Alternatively, when the announcement causes large price movements, prices contained less information. Using this proxy, among others, the model's prediction that prices are more informative following news with negative sentiment is verified: prices are up to 14% more informative when sentiment has been negative instead of positive.

Finally, tests also verify the prediction that hedge fund information acquisition is more sensitive to that of mutual funds than vice versa. When mutual funds anticipate that a small amount of hedge funds will become informed, between 9% and 19% more mutual fund IP addresses will acquire information. By contrast, when hedge funds anticipate a low level of information acquisition from mutual funds, 40% more hedge fund IP addresses will request information.

This paper is related to the literature studying information acquisition when investors face multiple dimensions of uncertainty. In particular, it is related to papers such as Romer (1993), Gervais (1997), Avery and Zemsky (1998), Li (2013), Back, Crotty and Li (2013), and Wang and Yang (2016), which consider scenarios where the precision of signals or the proportion of informed traders is unknown. These papers focus on implications for market microstructure, while this paper studies whether uncertainty about other informed investors impacts incentives to acquire costly information. In, Gao, Song and Wang (2013), the proportion of informed traders is unknown, but information acquisition is exogenous and there are no short-sale constraints. Banerjee and Green (2015) studies asset prices when some investors are uncertain whether others are trading with information. Similar to Banerjee and

Green (2015), my paper utilizes both rational expectations and differences of opinions approaches, but in contrast, my paper uses a common short-sale constraint to drive uncertainty about whether informed investors are participating in the market.[5] Further, while Banerjee and Green (2015) focuses on risk-premia and volatility, my paper focuses on conditional information acquisition in the face of short-sale prohibitions.

Much of the existing work using rational expectations models employs a setting with normally-distributed variables that generate linear prices. However, using insights from Breon-Drish (2015), the model in this paper can handle non-linear prices generated by the short-sale prohibition. Using the "residual demand" approach instead of the typical "conjecture and verify" approach, investors are able to update from the price even though they are uncertain regarding the participation of informed investors.

This paper also contributes to a large literature regarding short selling.[6] Although loosely related, the closest paper in this respect is Nezafat, Schroder and Wang (2017), which examines information acquisition in the presence of short-sale constraints. Similar to Nezafat, Schroder and Wang (2017), this paper allows information acquisition to be endogenously determined when investors are prohibited from short selling. In contrast, my paper allows for two sets of investors (constrained and unconstrained) who endogenously respond to each other. Also, my paper focuses on the interaction between short-sale prohibitions and recent news releases.

To date, empirical studies of information acquisition faced the difficulty of measuring investors' information acquisition activity. This has been overcome in part by inferring information acquisition from returns around earnings announcements (Morse (1981), Meulbroek (1992), Heflin, Subramanyam and Zhang (2003), Weller (2017)). Other papers have used

[5]Related papers utilizing the rational expectations approach include Grossman and Stiglitz (1980), Hellwig (1980), and Verrecchia (1982), those using differences of opinion include Harrison and Kreps (1978), Banerjee and Kremer (2010).

[6]Miller (1977), Diamond and Verrecchia (1987), Boehme, Danielsen and Sorescu (2006), Bris, Goetzmann and Zhu (2007), Chang, Cheng and Yu (2007), Saffi and Sigurdsson (2010), Engleberg, Reed and Ringgenberg (2012), Beber and Pagano (2013), Boehmer, Jones and Zhang (2013), and Kolanski, Reed and Thornock (2013).

Google or Yahoo searches to measure information acquisition (Da, Engleberg Gao (2011), Drake, Roulstone and Thornock (2012), Lawrence, Ryans, Sun and Laptev (2018)). This paper adds to this literature by using a rich measure of information acquisition derived from EDGAR search records. By using a hand-unmasked sample of mutual fund and hedge fund IP addresses, this paper is able to quantify information acquisition activity and clearly attribute it to specific institutional investors. While other papers explore the EDGAR log files, this is among the first to refine the data by focusing on a sample of identifiable institutional investors.[7]

The remainder of this paper is structured as follows: Chapter 2 details the theoretical model and its predictions. Chapter 3 describes the data and discusses the empirical results. Chapter 4 provides a series of robustness checks. Chapter 5 contains concluding remarks.

[7] Other papers using the EDGAR log files data include, Drake, Roulstone and Thornock (2016), Dyer (2017), Chen, Cohen, Gurun, Lou and Malloy (2018), Crane, Crotty and Umar (2018), and Gibbons, Iliev and Kalodimos (2018).

CHAPTER 2: A MODEL OF INFORMATION ACQUISITION

The model is an extension of the classic Grossman and Stiglitz (1980) setting. The key modifications are (*i*) the imposition of a short-sale prohibition on a subset of potentially informed investors, and (*ii*) the addition of an initial stage in which news generates disagreement prior to the information acquisition decision.

Section 2.1. Assets, Investors, and Information

There is a single, risky asset with payoff θ, where $\theta \sim N(0, \tau_\theta^{-1})$. The supply of the risky asset is random and denoted by u, where $u \sim N(0, \tau_u^{-1})$.[1] There is also a risky-free asset that has a gross return of one and is in perfectly elastic supply.

There are three types of investors: hedge funds, mutual funds, and retail. Hedge funds (of measure ν) are institutional investors without a short-sale prohibition. Hedge funds can learn θ perfectly by paying κ_θ. The portion of hedge funds who become informed is denoted by μ. Mutual funds (of measure ω) are institutional investors who are prohibited from short selling. Mutual funds can pay κ_α to observe a noisy signal: $S_\alpha = \theta + \alpha$ where $\alpha \sim N(0, \tau_\alpha^{-1})$. The portion of mutual funds who become informed is denoted by δ.

Retail investors (of measure one) have access to a *free* signal, which they believe provides information about the payoff: they think $S_\eta = \theta + \eta$ where $\eta \sim N(0, \tau_\eta^{-1})$. Institutional investors also observe this initial signal, but know that it is just noise: $S_\eta = \eta$. Retail investors do not update their beliefs upon observing the price as they believe the information acquired by institutional investors is only noise.[2]

---

[1]Noisy supply is included in the model to prevent prices from being fully revealing. Modeling shares of the risky asset to be in random supply is equivalent to a setting where shares are in zero net supply and noise traders submit random demands.

[2]This disagreement is akin to the differences of opinions approach.

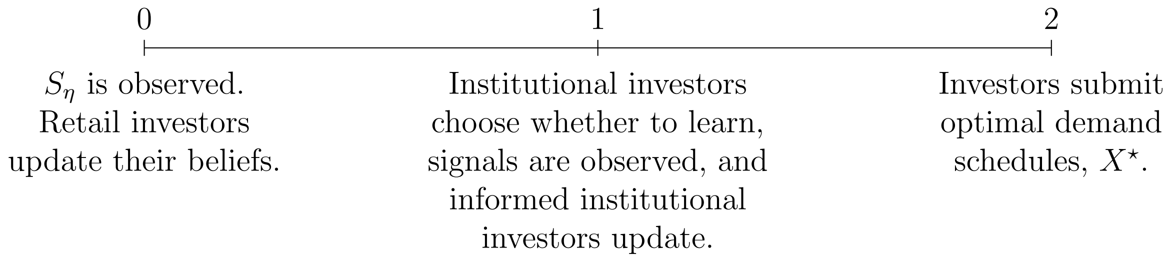

Figure 2.1: Model Timeline

The order of events is as follows: Upon observing the initial signal, retail investors believe that $\mathbb{E}[\theta|S_\eta] = \frac{\tau_\eta S_\eta}{\tau_\theta+\tau_\eta}$. Importantly, institutional investors decide whether to become informed after observing S_η. From their perspective, S_η generates predictable price pressure from retail investors. For example, a positive signal leads retail investors to believe the payoff of the risky asset will be greater than zero, whereas the correct expectation of θ is still zero. In expectation, retail investors will demand positive shares of the risky asset, thereby pushing the price of the asset above its payoff.

Institutional investors who choose to become informed receive their additional signals and update their beliefs regarding θ. Trading then occurs in a rational expectations equilibrium where investors submit optimal demand schedules, $X^\star$, which condition on the information in the price, P. All investors are risk-neutral and maximize their expected profits, Π, subject to quadratic inventory cost γ:[3]

$$\Pi = \big(\theta - P\big)X - \frac{\gamma}{2}X^2. \tag{2.1}$$

Following trading, the payoff of the risky asset is observed and trading gains/losses are realized. The timeline of events in this model is shown below.

[3]Results are expected to be similar in a setting with risk-averse investors, but given the non-linearity in prices, characterizing the results would be more difficult. Vives (2011), Rostek and Weretka (2012), Duffie and Zhu (2017), and Banerjee, Davis and Gondhi (2018).

Section 2.2. Equilibrium

An equilibrium consists of the optimal proportion of mutual funds and hedge funds who become informed, $\delta^\star$ and $\mu^\star$, such that: (*i*) investor demands are optimal, (*ii*) the price clears the market, (*iii*) hedge funds and mutual funds are indifferent between becoming informed and remaining uninformed, and (*iv*) institutional investors rationally update from the market clearing price.

Given the universal objective function, the optimal demand for every investor type can be expressed with the following equations:

$$X^\star = \frac{1}{\gamma}\Big(\mathbb{E}\big[\theta|I\big] - P\Big) \qquad \text{or} \qquad X_M^\star = \max\Big\{0, \frac{1}{\gamma}\Big(\mathbb{E}\big[\theta|I\big] - P\Big)\Big\}. \tag{2.2}$$

The optimal demand for investors who face a short-sale prohibition, $X_M^\star$, includes the maximization function to ensure they can only demand positive shares. The information sets, I, of all investor types contain the price and the initial signal; in addition, informed hedge funds know θ perfectly while informed mutual funds have observed S_α.

The market clearing condition requires that total demand from mutual funds, hedge funds, and retail investors is equal to total supply:[4]

$$\nu\mu X_{H\theta} + \nu(1-\mu)X_{HP} + \omega\delta X_{MS} + \omega(1-\delta)X_{MP} + X_R = u. \tag{2.3}$$

In equilibrium, institutional investors must be indifferent between becoming informed and remaining uninformed. Put differently, the marginal benefit of becoming informed must

[4]The subscripts to optimal demands refer to investor types and their information sets. $H\theta$ represents hedge funds who have observed θ while HP represents hedge funds who rely on the price. The subscripts for mutual funds and retail investors follow.

be equal to its marginal cost:

$$\mathbb{E}\Big[\big(\theta - P\big)X_{H\theta} - \frac{\gamma}{2}X_{H\theta}^2\Big|S_\eta\Big] - \mathbb{E}\Big[\big(\theta - P\big)X_{HP} - \frac{\gamma}{2}X_{HP}^2\Big|S_\eta\Big] = \kappa_\theta, \tag{2.4a}$$

$$\mathbb{E}\Big[\big(\theta - P\big)X_{MS} - \frac{\gamma}{2}X_{MS}^2\Big|S_\eta\Big] - \mathbb{E}\Big[\big(\theta - P\big)X_{MP} - \frac{\gamma}{2}X_{MP}^2\Big|S_\eta\Big] = \kappa_\alpha. \tag{2.4b}$$

Updating from the price is typically straightforward in this class of models. Following the conjecture and verify approach, the price is expressed as a linear function of the payoff, noise, and other factors. Then, the price can be transformed into a normally-distributed signal of the payoff and standard Bayesian updating leads to a tractable expression for the posterior belief regarding θ.

This conventional updating procedure does not apply in this setting for two reasons. First, due to the maximization function in the optimal demand of mutual funds, the price is non-linear. Second, uninformed investors are uncertain regarding the quality of the signal they observe from the price. The quality of the price's signal is determined by the trading activity of informed investors: as more informed investors trade, the information content of the price increases. In this model, the trading activity of informed, potentially-constrained investors (i.e., informed mutual funds) is unknown to uninformed investors. As a result, uninformed investors cannot update in the typical fashion.

Instead, uninformed mutual funds and hedge funds use the price to update as follows: At the trading stage, uninformed investors have observed S_η and the price, thus they know their own demand as well as the demand from retail investors. The uninformed remain uncertain, however, regarding the demands from informed mutual funds and hedge funds as well as noisy supply. With this in mind, uninformed investors construct an observable signal, S, by rearranging the market clearing condition according to what they know and what they do not know:

$$S \;=\; \underbrace{\nu(1-\mu)X_{HP} + \omega(1-\delta)X_{MP} + X_R}_{\text{Known/Observable}} \;=\; \underbrace{u - \nu\mu X_{H\theta} - \omega\delta X_{MS}}_{\text{Unknown}}. \tag{2.5}$$

Using optimal demand functions $X_{H\theta}$ and X_{MS}, the expression for S can be updated:

$$S \;=\; u - \nu\mu\frac{1}{\gamma}\big(\theta - P\big) - \omega\delta \max\left\{0, \frac{1}{\gamma}\Big(\mathbb{E}\big[\theta|S_\alpha, P\big] - P\Big)\right\}. \tag{2.6}$$

Equation (2.6) cannot be transformed into a standard, normally-distributed signal due to the maximization function in the optimal demand of informed mutual funds. Instead, the signal, S, comes from one of two regimes. The first arises when informed mutual funds believe the risky asset is undervalued ($\mathbb{E}\big[\theta|S_\alpha, P\big] > P$) and demand positive shares. In the second regime, informed mutual funds believe the asset is overvalued ($\mathbb{E}\big[\theta|S_\alpha, P\big] < P$) and demand zero shares as their short-sale prohibition binds. Although the signal extracted from the price is observable by uninformed investors using equation (2.5), which regime produces the signal is unclear: there is not a one-to-one mapping from price to regime. However, since the uninformed hold coherent beliefs regarding the likelihood of either regime being realized, they can update from the price using S. With this insight, the uninformed can form an expectation of the payoff upon observing the price:

$$\mathbb{E}\big[\theta|P\big] = \int_{-\infty}^{\infty} \theta f(\theta|P) d\theta = \int_{-\infty}^{\infty} \theta \frac{f(\theta, P)}{f(P)} d\theta, \tag{2.7}$$

where

$$f(\theta, P) = f_\theta(\theta) \int_{-\infty}^{\infty} f_\alpha(\alpha)\Big[F_u(\tilde{u}) f_u(u_1) + [1 - F_u(\tilde{u})] f_u(u_2)\Big] d\alpha. \tag{2.8}$$

The joint density of the payoff and the price accounts for the probabilities of being in either of the two regimes and the probability of observing S. The first regime is realized when $\mathbb{E}\big[\theta|S_\alpha, P\big] > P$. Given θ and α, this occurs with probability $F_u(\tilde{u})$. The second regime is realized when $\mathbb{E}\big[\theta|S_\alpha, P\big] < P$, which occurs with probability $[1 - F_u(\tilde{u})]$. The cutoff point between the two regimes is denoted with $\tilde{u}$. The probability of observing S

given the first regime is realized is $f_u(u_1)$, while $f_u(u_2)$ is the probability of observing S in the second regime.[5]

Informed mutual funds know which regime will be realized. As such, they use the price and the market clearing condition to construct an observable signal, S_I, and update as follows:[6]

$$\mathbb{E}\big[\theta|S_\alpha, P\big] = \frac{\tau_\alpha S_\alpha + \tau_I S_I}{\tau_\theta + \tau_\alpha + \tau_I}, \tag{2.9}$$

where

$$S_I = \theta - (\frac{\gamma}{\nu\mu})u, \tag{2.10}$$

and $\tau_I = (\frac{\nu\mu}{\gamma})^2 \tau_u$.

Given the non-linearity of the price and the complexity of $\mathbb{E}[\theta|P]$, the equilibrium price cannot be expressed analytically. Numerical simulation is required to solve for the price and to produce comparative statics. The objective of the simulation is to solve for the equilibrium proportions of mutual funds and hedge funds that become informed $(\delta^\star, \mu^\star)$. An overview of the simulation procedure is as follows: First, holding δ fixed and for a given S_η, a numerical solver is used to find the market clearing price for $10,000$ random draws of the (θ, α, u) triplet. For each of the $10,000$ draws, investors trade optimally (equation (2.2)), the price satisfies the market clearing condition (equation (2.3)), and investors use the price to rationally update (equations (2.7) and (2.9)). Next, the average profit differential between informed and uninformed hedge funds is calculated using the $10,000$ random draws. A numerical solver is used to find the optimal proportion of hedge funds who become informed, $\mu^\star$, such that this average profit differential is equal to the cost hedge funds pay to become informed (equation (2.4a)).

This procedure is followed across a grid of δ values from zero to one. Then, for each δ, the average profit differential between informed and uninformed mutual funds is calculated using

[5] A detailed derivation of $\mathbb{E}[\theta|P]$ is provided in Appendix A.

[6] The derivation of $\mathbb{E}[\theta|S_\alpha, P]$ is provided in Appendix B.

the $10,000$ random draws. The optimal proportion of mutual funds who become informed, $\delta^\star$, is identified as the δ such that the average profit differential is equal to the cost mutual funds pay to become informed (equation (2.4b)). Taken together, this process uses model parameters and a given S_η to find $\mu^\star$ and $\delta^\star$.[7]

The relationship between the initial signal, S_η, and average profit differentials between informed and uninformed mutual funds is illustrated in Figure 2.2. This figure shows that average profit differentials decrease with increasing S_η. Recall, S_η generates differing opinions between institutional investors and retail. Increasing the initial signal increases the likelihood that retail investors will have an optimistic view of the risky asset and will push its price above its fundamental value. The probability that mutual funds will refrain from trading and earn zero profits (whether informed of uninformed) thus increases with S_η. This is especially costly for informed mutual funds who are unable to benefit from their acquisition of costly, private information.

Figure 2.2 also shows that for a given S_η, average profit differentials strictly decrease with increasing δ: as more mutual funds become informed, the price will be more informative, providing a relative benefit to uninformed mutual funds. Furthermore, average profit differentials are relatively more sensitive to S_η than to δ. For example, Figure 2.2 shows that, holding S_η fixed, the difference in average profit differentials between $\delta = 0$ and $\delta = 1$ is less than 0.02. By contrast, fixing δ, the difference in average profit differentials between $S_\eta = -2$ and $S_\eta = +2$ is at least 0.06, which is three times as much as the previous case. It follows that when deciding whether to acquire private information, mutual funds are more sensitive to S_η than to δ. Thus, a common response of mutual funds to S_η will be to either all become informed ($\delta^\star = 1$) or to all remain uninformed ($\delta^\star = 0$).

The relationship between the initial signal, S_η, and optimal hedge fund information acquisition, $\mu^\star$, is illustrated in Figure 2.3. This figure highlights the effect of constrained mutual funds on the information acquisition of hedge funds. When mutual funds all remain

[7]The simulation algorithm is explained in more detail in Appendix C.

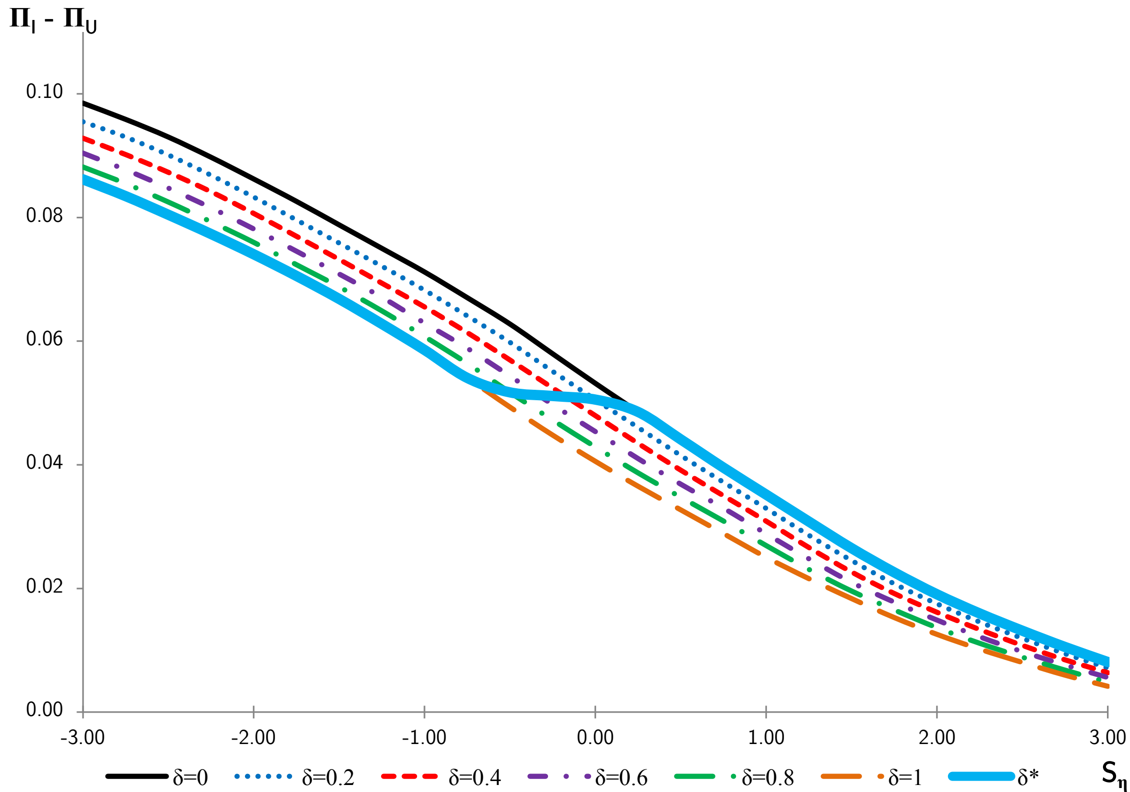

Figure 2.2: Mutual Fund Differential Profits

This figure illustrates the relationship between mutual fund differential profits ($\Pi_I - \Pi_U$) and the initial signal (S_η). The figure also highlights differential profits when mutual funds optimally acquire information, $\delta^\star$.

uninformed ($\delta = 0$), hedge funds do not condition their information acquisition on the initial signal. Under this scenario, the quality of the price's signal is perfectly known since there are zero informed mutual funds to be influenced by S_η. Hedge funds do condition their behavior on S_η when some proportion of mutual funds become informed ($\delta > 0$). When hedge funds expect more trading from informed mutual funds, they will acquire less information. Thus, hedge funds acquire less information given negative S_η and more when S_η is positive. This argument is made formally in the next section.

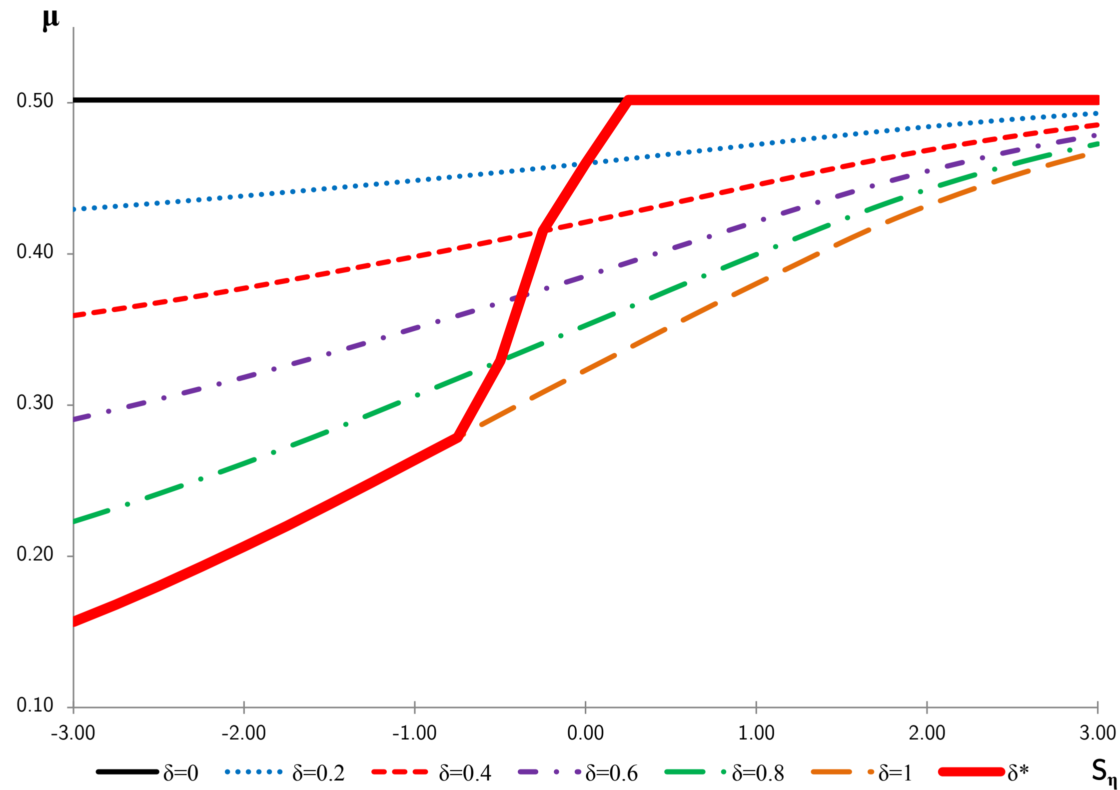

Figure 2.3: Hedge Fund Information Acquisition in Equilibrium

This figure illustrates the relationship between $\mu^{\star}$ and the initial signal, S_{η}. Mutual fund learning, δ, is taken as both exogenous and endogenous.

Section 2.3. Model Predictions

The model produces a series of novel predictions. First, the model generates predictions regarding the information acquisition of mutual funds and hedge funds. The model also predicts that the informativeness of prices is related to the initial signal and to the composition of informed investors. Finally, extending the model provides a unique prediction regarding information acquisition and institutional ownership.

Subsection 2.3.1. Information Acquisition and the Initial Signal

Figure 2.4 shows the optimal proportions of informed mutual funds and hedge funds with respect to the initial signal. It is clear that mutual funds prefer to become informed following a negative signal and prefer to remain uninformed following a positive signal. Consider the

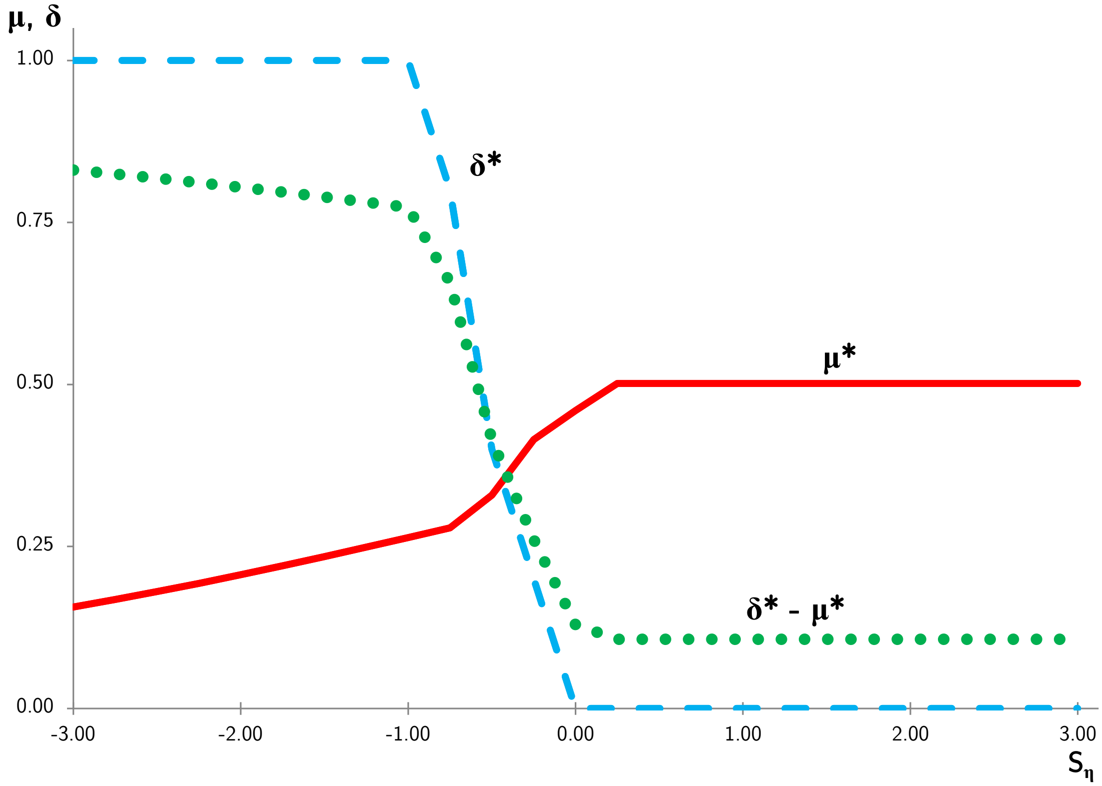

Figure 2.4: Relative Institutional Information Acquisition
This figure shows the optimal information acquisition of mutual funds ($\delta^{\star}$) and hedge funds ($\mu^{\star}$) in equilibrium. This figure also shows their difference ($\delta^{\star} - \mu^{\star}$).

case where the initial signal is positive, resulting in retail investors holding a relatively optimistic opinion of the payoff. From the perspective of mutual funds, retail investors are likely to generate upward price pressure by demanding positive shares of the asset. Thus, in expectation the asset will be overpriced and mutual funds will be constrained by their short-sale prohibition. Since a positive initial signal leads mutual funds to anticipate being unable to trade, their incentive to acquire costly information is diminished. Following positive S_{η}, mutual funds endogenously choose to remain uninformed. The opposite occurs when the initial signal is negative: mutual funds expect to trade profitably and are more willing to pay the cost to acquire information.

More hedge funds acquire information when they perceive fewer mutual funds will become informed and when these mutual funds are less likely to trade on their information. Following a negative signal, the pessimistic opinion of retail investors makes it more likely that mutual funds will be able to trade. More mutual funds will acquire information in this scenario, and in expectation more of their information will be impounded into the price. Hedge funds respond by substituting out the costly acquisition of private information and relying more on the freely observed and increasingly informative price.

Prediction 1. *Relative to hedge funds, mutual funds acquire more (less) information when the initial signal is negative (positive).*

Subsection 2.3.2. Price Informativeness

At first glance, it is unclear whether the price should be more informative following positive or negative initial signals. More hedge funds and fewer mutual funds are acquiring information when the signal is positive, but when the signal is negative fewer hedge funds and more mutual funds become informed. Closer examination of the model, however, provides both the answer and intuition.

Recall that mutual funds are acutely sensitive to the initial signal and their short-sale prohibition. As the signal moves from negative to positive, a relatively large portion of mutual funds will refrain from acquiring information, and for those who do become informed, they have a decreased likelihood of trading. A relatively large amount of information, therefore, is not acquired and not impounded into prices by mutual funds. On the other hand, more hedge funds become informed in response to fewer expected informed mutual funds. But whereas a large portion of mutual funds change their behavior, only a relatively small amount of hedge funds change from being uninformed to acquiring information. This relatively small group of informed hedge funds are unable to inject into the price the same amount of information that the large bloc of mutual funds would have.

Prediction 2. *Prices are more informative following negative signals than following positive signals.*

One way to see Prediction 2 is by measuring price informativeness as the covariance of the price and the payoff scaled by the variance of the price.[8] In other words, the informativeness of the price is the ratio of its signal and its noise. Using this terminology, Figure 2.5 illustrates that both the signal and the noise increase with S_η. That is, when the initial signal is positive, the price is both more variable and more closely associated with the payoff. Changes in the price's variance, however, are larger than changes in its covariance with the payoff. Figure 2.5 indicates that the variance of the price increases by approximately 45% as S_η changes from negative to positive, while the price's covariance with the payoff increases by only 25%. It follows that the informativeness of the price decreases with the initial signal since the price is becoming relatively more noisy.

Another way to view price informativeness is to consider the price's accuracy by measuring how close the equilibrium price is to the asset's payoff. This can be done by measuring the absolute difference between θ and the equilibrium price. While it is expected that this measure is smallest when the initial signal is near zero, comparing this measure between positive and negative signals should be instructive as to when prices are more informative. Figure 2.6 shows that it is when the initial signal is positive that the absolute difference between θ and the price is the largest, and thus the price is least accurate. This supports the evidence from Figure 2.5: prices are more informative following negative signals than positive.

Prediction 3. *Prices are more informative when relatively more mutual funds than hedge funds are acquiring information.*

Prediction 3 follows from Figures 2.5 and 2.6 and from the discussion above. When the signal is negative and relatively more mutual funds than hedge funds are informed, the price is more informative in terms of both the signal-to-noise ratio and price accuracy.

---

[8] Bai, Philipon and Savov (2016) suggests this ratio as a measure of price informativeness.

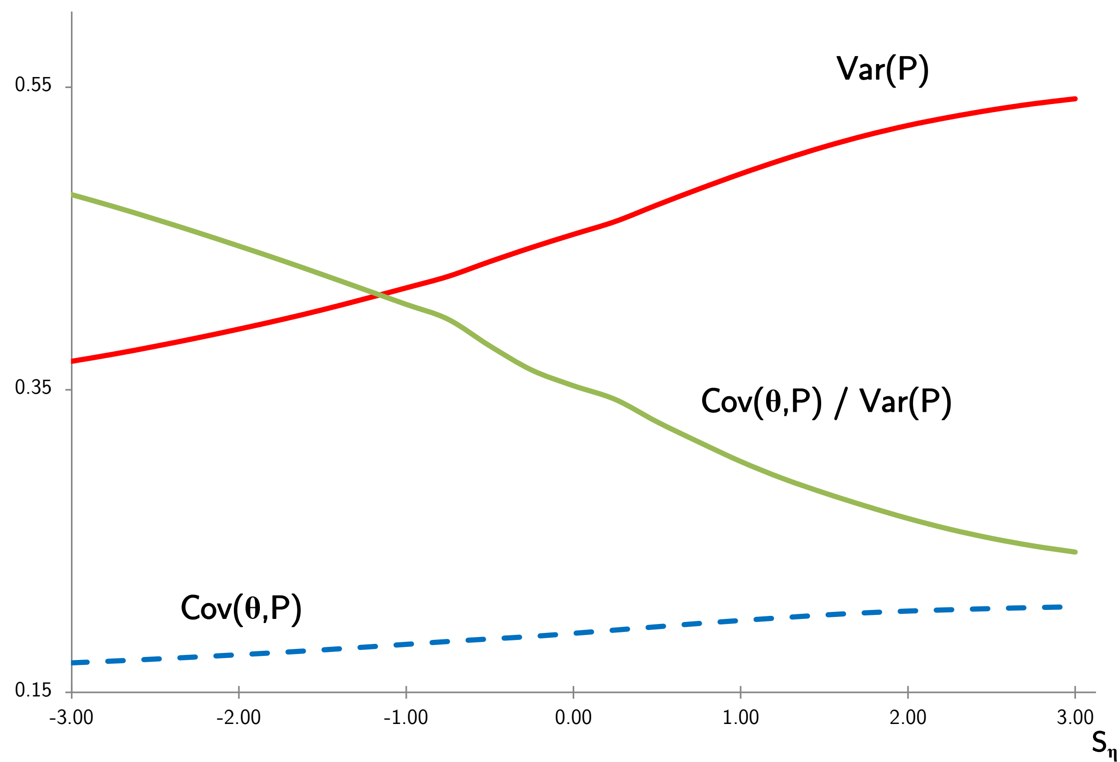

Figure 2.5: Price Informativeness: Signal-to-Noise

This figure shows the variance of the price, the covariance between the price and the payoff, and the price informativeness with respect to S_η. Price informativeness is defined as the covariance between the price and the payoff divided by the variance of the price.

Subsection 2.3.3. Information Acquisition and Other Investors

In addition to highlighting the relationship between information acquisition and the initial signal, the model also describes how mutual funds and hedge funds respond to each other. Consider the model analyzed without hedge funds. In this scenario, the initial signal still generates buying or selling pressure from retail investors and mutual funds remain prohibited from short selling. Since the initial signal changes the likelihood that mutual funds will be able to trade, they still consider the signal when deciding whether to acquire information. Further, Figure 2.4 shows that when hedge funds are present in the model, their level of information acquisition has a small influence on how many mutual funds become informed.

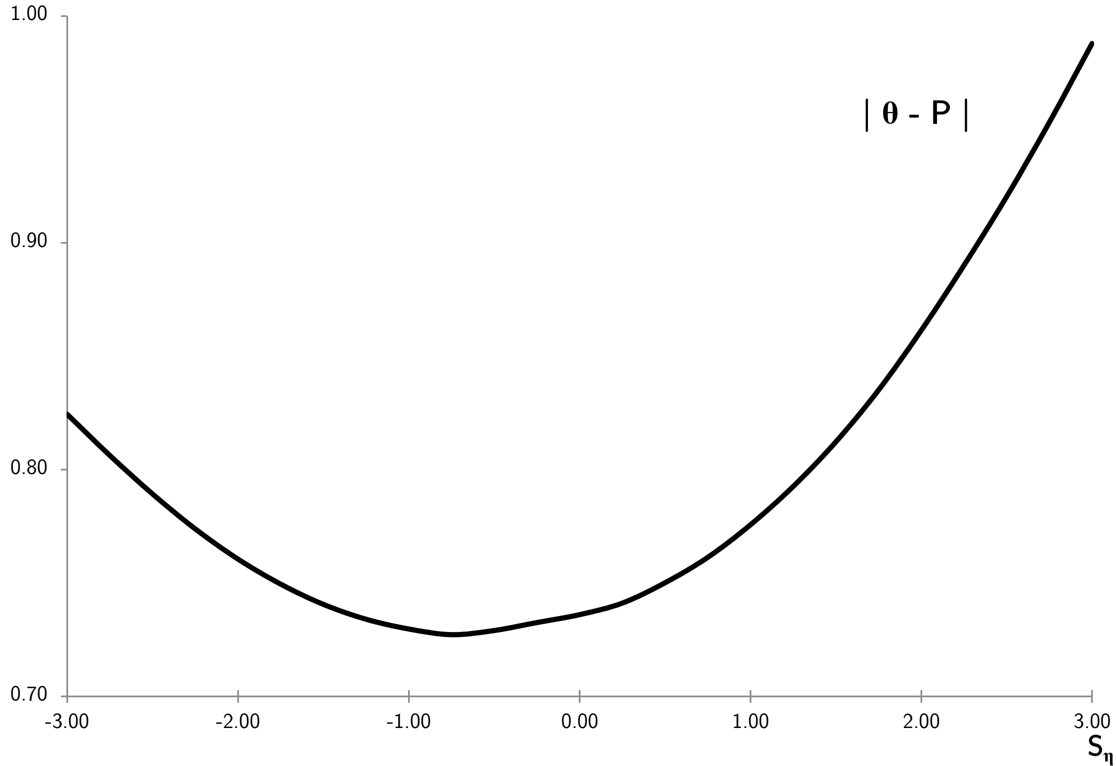

Figure 2.6: Price Informativeness: Accuracy

This figure shows the absolute difference between the payoff of the risky asset and its price as a function of S_η.

When deciding whether to acquire information, the chief concern of mutual funds is whether the initial signal will trigger their short-sale prohibition.

Now consider the model without mutual funds. This scenario is depicted by the horizontal line in Figure 2.3, which shows that without informed mutual funds, hedge funds acquire information independent of the initial signal. Hedge funds only condition their information acquisition on the signal to the extent that it influences the behavior of mutual funds. When deciding whether to become informed, hedge funds are primarily concerned with the information acquired by mutual funds, thus they are only indirectly affected by the signal, S_η.

Prediction 4. *In terms of information acquisition, hedge funds are sensitive to the behavior of mutual funds: more hedge funds acquire information when they expect fewer informed mutual funds. In comparison, mutual funds are less sensitive to the information acquisition behavior of hedge funds.*

Subsection 2.3.4. Information Acquisition and Institutional Ownership

The model can be readily extended to account for mutual funds who are not prohibited from short selling because they either own the asset or do not subject themselves to such a prohibition. If it is assumed that these unconstrained mutual funds can acquire private information, the model extension speaks to the responses of hedge funds and constrained mutual funds to institutional ownership.

Similar to an earlier argument, hedge funds primarily care about the activity of other informed investors, while constrained mutual funds are mainly concerned with the initial signal and their short-sale prohibition. As the mass of unconstrained mutual funds increases, constrained mutual funds do not alter their information acquisition. On the other hand, fewer hedge funds acquire information due to the substitutability of information. In other words, hedge funds substitute out costly private information and instead rely on a more informative and costless signal from the price.

Prediction 5. *Fewer hedge funds acquire information as institutional ownership increases. In comparison, information acquisition by constrained mutual funds is less sensitive to institutional ownership.*

CHAPTER 3: EMPIRICAL ANALYSIS

Testing the model's predictions requires data measuring the information acquisition activity of institutional investors. Given the model explicitly differentiates between types of institutions (mutual funds and hedge funds), the data should also distinguish between these institutional investor types. This paper uses a dataset that captures the information acquisition of investors and classifies them as mutual funds or hedge funds. The predictions from the model are empirically tested using this unique dataset.

Section 3.1. Data Sources and Sample Construction

The primary dataset used to test the predictions from the model has been derived from the EDGAR log files, which record requests to view documents on the EDGAR filing system, the SEC's online repository for public information. An observation in the EDGAR log files contains details from an electronic request to view, for example, a 10-K filing for IBM. This request is recorded along with the date, the time, the CIK of the filer, the specific filing requested, the size of the electronic file requested, and a masked version of the IP address from whence the request originated. Requesting IP addresses are masked as the last of their four octets is reported as a random set of three letters in place of the actual digits.[1] For example, the SEC reports the IP address 152.19.255.34 as 152.19.255.*xxx*, where *xxx* is a combination of three letters. Using the first three octets of an IP address combined with historical IP registration records, many IP addresses requesting information from EDGAR have been linked to specific financial institutions.[2] The result is a hand-unmasked sample of mutual fund and hedge fund IP addresses and their history of requesting (acquiring)

[1] In terms of IP addresses, an *octet* is a group of eight bits, or the one to three digit numbers separated by periods in the example above.

[2] IP registration records were acquired from MaxMind, https://www.maxmind.com/en/home.

information from EDGAR. Over 400 mutual fund and hedge fund companies have been unmasked, and 17.9 million of their requests for information between January 2012 and June 2017 have been included in the sample.

Mutual fund and hedge fund information acquisition are defined at the stock level and aggregated both monthly and in the weeks leading up to earnings announcements. Two variables measuring information acquisition are derived from the EDGAR data: (i) $Requests^m_{it}$ counts the total number of requests for information by mutual funds regarding stock i during time period t, and (ii) IPs^m_{it} records the total number of unique IP addresses from mutual funds requesting information (superscript h in place of m would indicate hedge funds instead of mutual funds). These two variables are referred to as the *Learning* variables in what follows, and serve as proxies for δ and μ from the model.

This study also uses data from RavenPack, which provides a history of news releases linked to specific firms. In addition, RavenPack conducts sentiment analysis of news items and produces an event sentiment score, which ranges from zero to 100, with 50 indicating neutral sentiment.[3] These stock-level sentiment scores are aggregated both monthly and in the weeks leading up to earnings announcements. The main variable derived from RavenPack is ESS_{it}, which is calculated as the median event sentiment score across all news stories regarding stock i over period t. This measure is used as a proxy for the model's initial signal, S_η.[4]

This paper also utilizes the CRSP, Compustat, TAQ, MIDAS, and Thomson Reuters databases.[5]

Section 3.2. Information Acquisition and the Sentiment of Recent News

Prediction 1 states that, relative to hedge funds, mutual funds will acquire more information when the sentiment of recent news is negative compared to when it is positive. This

---

[3]Previous work to utilize the RavenPack event sentiment scores include Green, Hand and Penn (2012), Ho, Shi and Zhang (2013), and Dang, Moshirian and Zhang (2015).

[4]The mapping from the model's S_η to ESS from RavenPack is discussed in more detail in Section 4.1.

[5]Further details regarding the sources of data are provided in Appendix D.

prediction is tested using the following regression model:

$$Diff_{it} = \beta_1 Negative_{it} + \beta_2 Positive_{it} + \beta_3 X_{it} + \zeta_i + \phi_t + \epsilon_{it}, \tag{3.1}$$

where $Diff$ is equal to the difference between mutual fund and hedge fund information acquisition. For instance, when using $Requests$ to measure information acquisition, $Diff$ is equal to total requests from mutual funds less total requests from hedge funds ($Requests^m_{it} - Requests^h_{it}$). The independent variable $Negative$ is an indicator equal to one when $ESS \leq 47$. The variable $Positive$ takes the value of one when $ESS \geq 53$ (recall that neutral sentiment is 50). The matrix X contains the natural logarithm of the following control variables: the number of news releases ($Articles$), total trading volume ($Volume$), average market capitalization over the time period, the number of owning institutional investors ($No.\ Inst.$), and total institutional ownership of shares ($Inst.\ Own.$). The X matrix also contains the average daily price volatility over the period ($Volatility$) and indicator variables for whether period t contains or corresponds to stock i's quarterly or annual filings. Stock and year-month fixed effects (ζ_i and ϕ_t) are included to control for unobserved heterogeneity at the stock-level and in the time dimension.

Two different settings are considered. In the first, t refers to monthly time periods: ESS, $Diff$ and all other variables are measured monthly. The second setting focuses on the weeks leading up to earnings announcements. In this setting, all variables are measured in the four week period prior to stock i's earnings announcement. This second setting is considered for two reasons. First, the EDGAR data show that mutual funds and hedge funds acquire more information in the weeks leading up to earnings announcements. Second, the model makes predictions regarding price informativeness. These predictions will be tested by examining prices and information acquisition leading up to earnings announcements.

Testing Prediction 1 is done by analyzing the difference between β_1 and β_2. Finding $\beta_1 > \beta_2$ would provide support for the prediction that relative to hedge funds, more mutual funds acquire information when the sentiment of news has been negative. Results are reported in Table 3.1.

Table 3.1 shows that relative to hedge funds, mutual funds acquire more information when the sentiment of news is negative. As an example, column (4) shows that when news sentiment is negative instead of positive in the weeks leading up to an earnings announcement, mutual funds will make 7 more requests relative to requests by hedge funds. Given that on average mutual funds make 45 more requests than hedge funds, mutual funds increase their information acquisition by 16% relative to hedge funds when sentiment is negative. Similarly, column (2) shows that mutual funds make 3.75 more requests, relative to requests by hedge funds, when sentiment has been negative for a given month. The average monthly difference between mutual fund and hedge fund requests is 22; thus an increase of 3.75 requests indicates that mutual funds increase their information acquisition by 17% relative to hedge funds when sentiment is negative for the month.

When focusing on IPs, column (8) of Table 3.1 shows that almost 2 additional mutual fund IP addresses, relative to hedge fund IPs, acquire information when the sentiment of news is negative. This suggests that when measured by IP addresses, mutual funds increase their information acquisition by 14% relative to hedge funds when sentiment is negative. Overall, Table 3.1 provides support for Prediction 1: relative to hedge funds, more mutual funds acquire information when the sentiment of news is negative instead of positive.

Section 3.3. Price Informativeness

The model makes two predictions regarding the informativeness of prices. Prediction 2 states that prices should be more informative when the sentiment of news is negative instead of positive. Prediction 3 states that prices should be more informative when relatively more mutual funds are acquiring information compared to hedge funds. Both of these predictions can be tested using the data previously detailed with the addition of price informativeness measures.

One proxy for the information content of prices is the price jump ratio (PJR), a measure developed in Weller (2017). The price jump ratio compares price movements in the weeks

Table 3.1: Information Acquisition and the Sentiment of Recent News

This table presents results from testing the relationship between information acquisition and the sentiment of recent news. The tests use the regression model in equation (3.1). The dependent variable, $Diff$ is measured with either *Requests* or IPs and both the monthly setting and the earnings announcement (EA Dates) setting are utilized. The variable ESS_{it} captures the sentiment of recent news regarding stock i during period t. ESS is measured using the median sentiment rating from RavenPack. The indicator variable $Negative$ takes on the value of one if $ESS \leq 47$. The indicator variable $Positive$ takes on the value of one if $ESS \geq 53$. The table presents $Neg - Pos$ as a test of whether negative and positive news induce differential information acquisition. The matrix X_{it} contains the natural logarithm of the following control variables: the number of news releases (*Articles*), total trading volume ($Volume$), average market capitalization over the time period (not shown), the number of owning institutional investors ($No.\,Inst.$), and total institutional ownership of shares ($Inst.\,Own.$). The X matrix also contains the average daily price volatility over the period ($Volatility$) and indicator variables for whether period t contains or corresponds to stock i's quarterly or annual filings. Stock and year-month fixed effects are included. Standard errors are clustered by stock and year-month. The sample is made up of approximately 3,300 stocks and 66 months.

	Requests				IPs			
	Monthly		EA Dates		Monthly		EA Dates	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Negative$	2.25***	1.42*	4.85***	4.39***	1.03***	0.77***	1.28***	1.18***
(s.e.)	(0.65)	(0.71)	(1.27)	(1.44)	(0.15)	(0.16)	(0.20)	(0.23)
$Positive$	-2.49***	-2.32***	-4.62***	-2.67**	-0.71***	-0.52***	-0.87***	-0.70***
	(0.54)	(0.59)	(1.03)	(1.11)	(0.07)	(0.09)	(.141)	(0.17)
Articles		4.28***		6.17***		1.27***		1.66***
		(0.38)		(0.60)		(0.09)		(0.10)
$Volume$		4.46***		5.00***		1.25***		1.29***
		(0.45)		(0.50)		(0.11)		(0.09)
$No.\,Inst.$		0.91		1.15		0.04		-0.030
		(0.98)		(1.15)		(0.19)		(0.19)
$Inst.\,Own.$		-0.23		-0.50		-0.15		-0.04
		(0.55)		(0.56)		(0.10)		(0.09)
$Volatility$		-0.92***		-0.68**		-0.19***		-0.07
		(0.23)		(0.27)		(0.05)		(0.05)
$Neg - Pos$	4.74***	3.75***	9.47***	7.05***	1.74***	1.29***	2.15***	1.88***
(s.e.)	(0.96)	(1.11)	(2.03)	(2.26)	(0.18)	(0.19)	(.305)	(0.36)
R^2	.560	.568	.595	.601	.778	.799	.811	.820
Observations	141,146	141,146	48,782	48,782	141,146	141,146	48,782	48,782

leading up to an earnings announcement to price movements at the announcement:

$$PJR_{it} = \frac{CAR_{it}^{(T-1,T+1)}}{CAR_{it}^{(T-30,T+1)}} \tag{3.2}$$

where $CAR_{it}^{(T-1,T+1)}$ is the cumulative abnormal return from the day before stocks i's earnings announcement to the day after, while the denominator contains the cumulative abnormal return beginning one month before an earnings announcement.[6]

The intuition behind this measure can be developed with an example. Consider that at its earnings announcement a stock's price experiences a significant jump, especially compared to how much the price moved over the last month. It can be said that the price just prior to the earnings announcement was weakly informative as to the content of the announcement. On the other hand, consider that at the earnings announcement, the stock's price moved very little, particularly when compared to how much the price moved over the last month. This is an example of a price accurately corresponding to the content of the announcement. Following this intuition, high price jump ratios indicate low levels of price informativeness while low price jump ratios are indicative of high price informativeness.

Following the same logic, the numerator from the price jump ratio can be used as a second proxy for price informativeness. This second measure is referred to as the price jump (PJ). Using both PJR and PJ, Predictions 2 and 3 can be tested using the following regression model:

$$\left\{PJR_{it} \text{ or } ln|PJ_{it}|\right\} = \beta_1 L_{it-1} + \beta_2 Negative_{it} + \beta_3 Positive_{it} + \beta_4 X_{it} + \zeta_i + \phi_t + \epsilon_{it} \tag{3.3}$$

where,

$$L_{it-1} \in \left\{ln(Learning_{it-1}^{m}), ln(Learning_{it-1}^{h}), Diff_{it-1}\right\}. \tag{3.4}$$

[6]Abnormal returns are calculated using the three-factor model (Fama and French (1993)).

This model is similar to the regressions used previously. The independent variable, L, measures learning from either mutual funds or hedge funds, and also measures their difference, $Diff$. Testing whether prices are more informative when news sentiment is negative is done by comparing β_2 to β_3. Finding $\beta_2 < \beta_3$ would provide support for Prediction 2. The prediction that prices are more informative when relatively more mutual funds are learning is tested by analyzing β_1 in the specification where $L = Diff$. If $\beta_1 < 0$ in this specification, then Prediction 3 would be supported.

Using $Learning$ instead of $Diff$ as an independent variable provides a validity check for the data. For instance, finding $\beta_1 \nless 0$ would imply that as institutional investors acquire more information, the price of a stock fails to become more informative. However, it should be the case that more information acquisition will lead to more informative prices. If institutional information acquisition as measured from the EDGAR data is a good proxy, it should provide estimates for β_1 that are less than zero.

Finally, L is lagged one time period. This is done to address reverse causality. If $\beta_1 < 0$ using $Learning_{it}$, it could be the case that investors are acquiring more information when they believe the price is more informative. This reverse causality story is ruled out by using lagged $Learning$.

Results using equation (3.3) are displayed in Table 3.2. In this table, odd-numbered columns measure price informativeness with PJR and even-numbered columns use PJ, while $Learning$ is measured using IPs. Notice that β_1 is significantly less than zero in the first four columns. This indicates that prices are more informative when more information is acquired by mutual funds and hedge funds. Specifically, the estimate in column (1) of -0.017 indicates that the price jump ratio will fall by 50 basis points when 30% more mutual fund IP addresses request information. In terms of the price jump, the estimate of -0.056 in column (4) suggests that it will decrease by 1% when 18% more hedge funds acquire information.

Prediction 2 is supported as negative sentiment leads to price jump ratios being reduced by 4 percentage points. Given that the average price jump ratio in the sample is 50 percentage

Table 3.2: Price Informativeness

This table presents results from testing whether information acquisition and the sentiment of recent news are related to price informativeness. The tests use the regression model in equation (3.3). Odd-numbered columns use PJR as the dependent variable while even-numbered columns use PJ. *Learning* is measured using IPs and is lagged one period in all specifications. ESS is measured using the median sentiment rating from RavenPack. The indicator variable $Negative$ takes on the value of one if $ESS \leq 47$. The indicator variable $Positive$ takes on the value of one if $ESS \geq 53$. The table presents $Neg - Pos$ as a test of whether negative and positive news induce differential price informativeness. The matrix X_{it} contains the natural logarithm of the following control variables: the number of news releases ($Articles$), total trading volume ($Volume$), average market capitalization over the time period (not shown), the number of owning institutional investors ($No.\,Inst.$), and total institutional ownership of shares ($Inst.\,Own.$). The X matrix also contains the average daily price volatility over the period ($Volatility$) and indicator variables for whether period t contains or corresponds to stock i's quarterly or annual filings. Stock and year-month fixed effects are included. Standard errors are clustered by stock and year-month. The sample is made up of approximately 3,300 stocks and 66 months.

	Price Jumps					
	Mutual Funds		Hedge Funds		Difference	
	(1)	(2)	(3)	(4)	(5)	(6)
Learning	-.017***	-.063***	-.012**	-.056***	-.0013**	-.007***
(s.e.)	(.005)	(.013)	(.006)	(.012)	(.0006)	(.001)
Negative	-.022*	-.079***	-.022*	-.078***	-.023*	-.080***
	(.011)	(.026)	(.011)	(.026)	(.011)	(.026)
Positive	.018	.065***	.018	.064***	.018	.064***
	(.011)	(.019)	(.011)	(.019)	(.011)	(.019)
Articles	-.018***	-.046***	-.019***	-.047***	-.019***	-.047***
	(.004)	(.009)	(.004)	(.009)	(.004)	(.009)
Volume	-.052***	.036**	-.052***	.034**	-.053***	.033**
	(.006)	(.014)	(.006)	(.013)	(.006)	(.013)
No. Inst.	-.017	.073**	-.017	.071**	-.017	.074**
	(.017)	(.033)	(.017)	(.033)	(.017)	(.033)
Inst. Own.	.018*	-.018	.018*	-.017	.018*	-.018
	(.010)	(.017)	(.010)	(.017)	(.010)	(.017)
Volatility	.000	.098***	.000	.098***	.000	.098***
	(.003)	(.007)	(.003)	(.007)	(.003)	(.007)
$Neg - Pos$	-.040**	-.144***	-.040**	-.143***	-.041**	-.144***
(s.e.)	(.018)	(.036)	(.018)	(.037)	(.018)	(.036)
R^2	.304	.233	.304	.232	.304	.233
Observations	10,613	39,643	10,613	39,643	10,613	39,643

points, price jump ratios are 8% lower when news sentiment is negative compared to when it is positive. In other words, negative sentiment improves price informativeness by 8%, as measured by the price jump ratio. When measuring price informativeness with price jumps, Table 3.2 shows that prices are 14% more informative following news with negative sentiment instead of positive.

Columns (5) and (6) provide support for Prediction 3. These columns show that as mutual funds acquire more information relative to hedge funds, prices become more informative. Highlighting column (5), the price jump ratio drops by 1.3 percentage points when 10 more mutual fund IP addresses request information. This suggests that prices are 2.6% more informative when, compared to hedge funds, 10 more mutual fund IP addresses acquire information.

Section 3.4. Information Acquisition and Other Investors

Prediction 4 states that hedge funds acquire less information when they expect more mutual funds will be learning. In comparison, mutual funds are less sensitive to hedge funds' level of information acquisition. This prediction can be tested using similar techniques to the previous section, but requires two steps. In the first step, news sentiment and other information from period $t-1$ are used to predict the level of information acquisition in period t from hedge funds or mutual funds. For the sake of exposition, suppose that the first step predicts information acquisition by hedge funds using the following regression equation:

$$ln(Learning_{it}^{h}) = \beta_1 Negative_{it-1} + \beta_2 Positive_{it-1} + \beta_3 X_{it-1} + \zeta_i + \phi_t + \epsilon_{it}, \tag{3.5}$$

which is similar to equation (3.1) except the right-hand-side variables are calculated from period $t-1$ and the dependent variable measures hedge fund learning instead of $Diff$. In the second step, the following regression is utilized to test the response of mutual funds to the expected level of information acquisition from hedge funds:

$$ln(Learning_{it}^{m}) = \beta_1 Low_{it}^{h} + \beta_2 High_{it}^{h} + \beta_3 Y_{it} + \zeta_i + \phi_t + \epsilon_{it}. \tag{3.6}$$

In this setup, the level of hedge fund information acquisition predicted from the first step is used to create variables indicating whether hedge funds are expected to acquire a high or low amount of information. The indicator variable Low^h is equal to one when hedge fund information acquisition is predicted to be in the lower quartile for a given stock. Similarly, $High^h$ is equal to one when hedge fund learning is predicted to be in the upper quartile.[7] The control variables contained in Y include $Negative$, $Positive$, and X, similar to previous specifications. Again, the regressions include stock and year-month fixed effects.

Equations (3.5) and (3.6) detail how to test whether mutual funds are sensitive to hedge funds. These same equations can be used to test whether hedge funds are sensitive to mutual funds by making the obvious changes. Testing these sensitivities is done by analyzing the difference between β_1 and β_2 from the second-step regression. Finding $\beta_1 > \beta_2$ when Low^m and $High^m$ represent predicted mutual fund learning would indicate that hedge funds acquire more information when they expect mutual funds are acquiring less.

Table 3.3 reports results using monthly time periods (odd-numbered columns) and the weeks prior to earnings announcements (even-numbered). Three points can be made about these results. First, in terms of acquiring information, hedge funds are highly sensitive to the expected level of mutual fund learning. In Panel A, hedge funds increase their level of information acquisition by as much as 42% when they anticipate a low level of mutual fund information acquisition. Second, mutual funds are sensitive to the expected level of hedge fund learning. Mutual funds increase their own level of information acquisition by between 9% and 19% when they anticipate low information acquisition by hedge funds.

Third, hedge funds are more sensitive to mutual funds than vice versa. This can be seen by comparing the estimates of $\beta_1 - \beta_2$ between hedge funds and mutual funds from Panel A. Notice that hedge fund estimates are between 38% and 42%, while mutual fund estimates

[7] In addition to defining Low and $High$ using quartiles, extreme deciles and terciles are also used.

Table 3.3: Information Acquisition and Other Investors

This table presents results from testing whether hedge funds and mutual funds condition their information acquisition on expected learning from each other. The tests utilize the regression model in equation (3.5) to calculate the expected level of learning from the other investor type. Then the regression model in equation (3.6) is used to test whether investors learn more when it is expected that the other investors will learn less. Odd-numbered columns define t in monthly terms. Even-numbered columns define t as the time period 15 days prior to an earnings announcement date while $t-1$ corresponds to 15 day period just prior to t. The dependent variables $Learning^h$ and $Learning^m$ are measured using either *Requests* or IPs. The independent variable ESS captures the sentiment of recent news. ESS is measured using the median sentiment rating from RavenPack. The indicator variable $Negative$ takes on the value of one if $ESS \leq 47$. The indicator variable $Positive$ takes on the value of one if $ESS \geq 53$. The matrix X_{it} contains the natural logarithm of the following control variables: the number of news releases (*Articles*), total trading volume ($Volume$), average market capitalization over the time period (not shown), the number of owning institutional investors ($No.\, Inst.$), and total institutional ownership of shares ($Inst.\, Own.$). The X matrix also contains the average daily price volatility over the period ($Volatility$) and indicator variables for whether period t contains or corresponds to stock i's quarterly or annual filings. *Low* is equal to one when learning from the other investor type for a given stock is predicted to be in the lower quartile (Panel A), decile (Panel B), or tercile (Panel C). $High$ is equal to one when learning from the other investor type for a given stock is predicted to be in the upper quartile, decile, or tercile. The table presents $Low - High$ as a test of whether low expected learning from others results in different information acquisition than when learning from others is expected to be high. The control variables contained in Y include $Negative$, $Positive$, and X. Stock and year-month fixed effects are included. Standard errors are clustered by stock and year-month and the Murphy-Topel correction has been applied for predicted regressors. The sample is made up of approximately 3,300 stocks and 66 months.

	Mutual Funds				Hedge Funds			
	Requests		IPs		*Requests*		IPs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Quartile Cutoff								
Low	.054**	.131***	.063***	.121***	.329***	.279***	.288***	.234***
(s.e.)	(.024)	(.034)	(.017)	(.024)	(.106)	(.094)	(.065)	(.055)
High	-.033	-.038	-.028	-.070***	-.092**	-.119**	-.114***	-.148***
	(.025)	(.031)	(.017)	(.022)	(.037)	(.049)	(.027)	(.029)
$Low - High$	.087**	.168***	.091***	.192***	.422***	.398***	.402***	.382***
(s.e.)	(.037)	(.048)	(.024)	(.034)	(.120)	(.116)	(.070)	(.071)
R^2	.745	.688	.819	.776	.611	.592	.688	.663
Observations	126,574	38,983	126,574	38,983	126,574	38,983	126,574	38,983
Panel B. Decile Cutoff								
Low	.068*	.165***	.093***	.157***	.600***	.421***	.457***	.337***
(s.e.)	(.039)	(.048)	(.025)	(.034)	(.166)	(.152)	(.090)	(.080)
High	-.048*	-.075**	-.033*	-.098***	-.133***	-.254***	-.148***	-.178***
	(.029)	(.036)	(.019)	(.032)	(.049)	(.072)	(.042)	(.040)
$Low - High$	.116**	.239***	.126***	.255***	.733***	.675***	.605***	.514***
(s.e.)	(.050)	(.063)	(.030)	(.046)	(.180)	(.177)	(.102)	(.097)
R^2	.745	.688	.819	.776	.612	.592	.689	.663
Observations	126,574	38,983	126,574	38,983	126,574	38,983	126,574	38,983
Panel C. Tercile Cutoff								
Low	.044*	.132***	.050***	.130***	.283***	.264***	.249***	.214***
(s.e.)	(.022)	(.032)	(.015)	(.021)	(.088)	(.081)	(.055)	(.047)
High	-.042*	-.029	-.038**	-.051**	-.086**	-.091**	-.095***	-.120***
	(.022)	(.034)	(.016)	(.023)	(.033)	(.043)	(.023)	(.027)
$Low - High$	.086**	.162***	.089***	.181***	.369***	.355***	.343***	.333***
(s.e.)	(.034)	(.046)	(.022)	(.031)	(.101)	(.101)	(.063)	(.060)
R^2	.745	.688	.819	.776	.610	.592	.688	.663
Observations	126,574	38,983	126,574	38,983	126,574	38,983	126,574	38,983

are between 9% and 19%. Furthermore, when comparing the 95% confidence intervals of these estimates, the intervals are disjoint for three out of the four settings considered.[8]

Overall, Table 3.3 provides supporting evidence for Prediction 4: when it comes to information acquisition, hedge funds are more sensitive to mutual funds than vice versa.

Section 3.5. Information Acquisition and Institutional Ownership

The final prediction from the model speaks to the influence of institutional ownership on information acquisition. Prediction 5 states that when it comes to information acquisition, hedge funds are more sensitive to institutional ownership than are mutual funds. Hedge funds will acquire less information when institutional ownership is higher. Mutual fund information acquisition, on the other hand, is less sensitive to institutional ownership.

Institutional ownership has been included in each of the previous tests using both the number of owning institutions and the number of shares owned by institutions. Table 3.3 included these two variables in the regression models explaining mutual fund and hedge fund information acquisition. For mutual funds, Table 3.3 provides evidence that they acquire more information as the number of owning institutions increases. However, the number of shares owned by institutions is insignificant for mutual funds. With respect to hedge funds, Table 3.3 shows that the number of owning institutions has no relationship with information acquisition, but the number of shares owned by institutions does. The results suggest that increasing institutional ownership decreases hedge fund information acquisition.

Causal interpretations should not be drawn from Table 3.3 as institutional ownership and information acquisition are measured contemporaneously. It could be the case that information acquisition leads to changes in institutional ownership, not the other way around. To account for reverse causality, Table 3.4 displays results using equation (3.1) in the monthly setting with lagged institutional ownership and *Learning* for mutual funds or hedge funds as the dependent variable.

---

[8] The comparison must be done within similar settings. For example, column (1) compares to column (5) while column (2) compares to column (6).

Table 3.4: Information Acquisition and Institutional Ownership

This table presents results from testing the relationship between information acquisition and institutional ownership. The tests use the regression model similar to equation (3.1). All columns utilize the monthly setting. The dependent variable *Learning* is measured with either *Requests* or *IPs*. The variable ESS_{it} captures the sentiment of recent news regarding stock i during period t. ESS is measured using the median sentiment rating from RavenPack. The indicator variable *Negative* takes on the value of one if $ESS \leq 47$. The indicator variable *Positive* takes on the value of one if $ESS \geq 53$. The matrix X_{it} contains the natural logarithm of the following control variables: the number of news releases (*Articles*), total trading volume (*Volume*), average market capitalization over the time period (not shown), the number of owning institutional investors (*No. Inst.*), and total institutional ownership of shares (*Inst. Own.*). Importantly, *No. Inst.* and *Inst. Own.* are lagged one month. The X matrix also contains the average daily price volatility over the period (*Volatility*) and indicator variables for whether period t contains or corresponds to stock i's quarterly or annual filings. Stock and year-month fixed effects are included. Standard errors are clustered by stock and year-month. The sample is made up of approximately 3,300 stocks and 66 months.

Panel A. Mutual Funds

	Requests			*IPs*		
	(1)	(2)	(3)	(4)	(5)	(6)
No. Inst.	.024		.056***	.010		.034***
(s.e.)	(.015)		(.017)	(.010)		(.012)
Inst. Own.		-.006	-.030**		-.008	-.022***
		(.010)	(.011)		(.007)	(.008)
R^2	.745	.745	.745	.818	.818	.818
Observations	126,619	126,619	126,619	126,619	126,619	126,619

Panel B. Hedge Funds

	Requests			*IPs*		
	(1)	(2)	(3)	(4)	(5)	(6)
No. Inst.	-.045**		.009	-.031***		.010
(s.e.)	(.021)		(.025)	(.011)		(.013)
Inst. Own.		-.046***	-.050***		-.033***	-.037***
		(.011)	(.014)		(.006)	(.008)
R^2	.609	.609	.609	.684	.684	.684
Observations	126,619	126,619	126,619	126,619	126,619	126,619

With respect to mutual funds, Table 3.4 provides inconsistent evidence for the influence of institutional ownership on their information acquisition. When using both institutional ownership variables in the regression, it appears that institutional ownership matters, but the two variables disagree as to the direction. As the number of institutional owners increases mutual funds acquire more information. However, as the number of shares owned by institutions increases, mutual funds acquire less information. Overall, there is no clear relationship between mutual fund information acquisition and institutional ownership.

When focusing on hedge funds, Table 3.4 provides support for Prediction 5. That is, as institutional ownership increases, hedge funds acquire less information. In particular, as institutional ownership increases by 20%, hedge funds acquire between 0.7% and 1.0% less information. In summary, Prediction 5 is supported as hedge fund information acquisition is clearly related to institutional ownership, but mutual fund information acquisition is not.

CHAPTER 4: ROBUSTNESS TESTS

Table 3.1 through Table 3.4 provide the main empirical findings of this paper. In this section, these findings are subject to several checks for robustness. These checks include using subsamples and alternative measurements of key variables. This section also provides detail regarding the mapping from the model's S_η to ESS derived from the RavenPack data.

Section 4.1. Mapping Between S_η and ESS

Recall that ESS is measured using event sentiment scores from news stories collected by RavenPack. The measurement of ESS uses stories from a variety of sources, including the Dow Jones Newswires, Wall Street Journal, NBC, Reuters, New York Times, and Yahoo! News, among others. Also, news stories containing tabular (i.e., quantitative) information as well as commentary are included in this measurement. Thus, by construction ESS is not solely a measure of fundamental information type, but also includes sentiment or tone. For this reason, ESS is used as a proxy for S_η.

Further, for ESS to be a reasonable proxy for the initial signal, it should be related to trading pressure from retail investors. To test whether this is the case, the buy-sell ratio using signed odd-lot trades is used as a proxy for retail trading pressure.[1] A two-step approach is used in order to account for algorithmic trading since odd-lots are not solely from retail investors, as demonstrated in O'Hara, Yao, and Ye (2014). In the first step, odd-lot buys and sells are explained using the cancel-to-trade ratio (CTR), which is a proxy for algorithmic trading. In this specification, $oddTrades_{it}$ is either the total amount of odd-lot buys or sells

[1] The idea that small trades, or odd-lot trades, are more likely to come from retail traders has been promoted in papers such as, Lamont and Frazzini (2007), Hvidkjaer (2008), and Barber, Odean and Zhu (2009). Trading data is taken from TAQ. The results presented use the Lee and Ready (1991) algorithm to sign trades. Results are robust to using the algorithms developed by Ellis, Michaely and O'Hara (2000) and Chakrabarty, Li, Nguyen and Van Ness (2007).

for stock i during time period t. Stock and year-month fixed effects are included:

$$ln(oddTrades_{it}) = \beta_1 ln(CTR_{it}) + \zeta_i + \phi_t + \epsilon_{it}. \tag{4.1}$$

The buy-sell ratio from odd-lot trades ($OddBSR$) is then calculated using the residuals from the first-step regressions. These residuals are interpreted as odd-lot trades that are orthogonal to algorithmic trading, thus they are more likely to represent retail trading. Next, the influence of ESS on this retail buy-sell ratio is tested using the following regression:

$$LowBSR_{it} = \beta_1 Negative_{it} + \beta_2 Positive_{it} + \beta_3 RoundBSR_{it} + \beta_4 X_{it} + \zeta_i + \phi_t + \epsilon_{it}, \tag{4.2}$$

where the dependent variable, $LowBSR$, takes the value of one when $OddBSR$ is below either 0.50 or 0.25. Thus, $LowBSR_{it}$ indicates whether stock i has experienced relative selling pressure from retail investors in time period t. The control variable $RoundBSR$ is the buy-sell ratio from round lots. Using residual odd-lot trades from the first stage controls for algorithmic trading, and using $RoundBSR$ in the second stage controls for general trading direction. The X matrix contains control variables such as the absolute difference between $OddBSR$ and 0.50 ($RatioSize$) and the natural logarithm of the number of news releases ($Articles$). Stock and year-month fixed effects are included.

If ESS is a reasonable proxy for the initial signal, it should be the case that news with negative sentiment leads to more selling pressure from retail investors and news with positive sentiment leads to more buying pressure. Thus, following equation (4.2), β_1 should be greater than β_2. This would indicate that negative sentiment increases the probability that there will be significant selling pressure from retail investors. Results using equation (4.2) are reported in Panel A of Table 4.1.

Panel A shows that news with negative sentiment increases the probability of selling pressure from retail investors while news with positive sentiment decreases that probability. Specifically, examining $\beta_1 - \beta_2$ shows that when the sentiment of news has been negative,

Table 4.1: ESS as a Proxy for the Initial Signal

This table presents results from testing the relationships between the sentiment of recent news, retail trading (Panel A), and price movements around earnings announcements (Panel B). The *OddBSR* variable is created using residual buys and sells from regressions using equation (4.1). The results below use the regression model in equation (4.2). The dependent variable *LowBSR* is an indicator of whether *OddBSR* is below 0.50 or 0.25. The variable *RatioSize* measures the absolute difference between *OddBSR* and 0.50. *RoundBSR* measures the buy-sell ratio from round-lot trading. *Under* (*Over*) is equal to one when stock i's stock price jumps up (down) by at least 25% at its earning announcement. *JumpSize* is the absolute value of a stock's price jump. The matrix X_{it} contains the natural logarithm of the the number of news releases (*Articles*). Stock and year-month fixed effects are included. Standard errors are clustered by stock and year-month . The sample is made up of approximately 3,300 stocks and 42 months.

Panel A. ESS and Retail Trades

	Monthly				Earnings Announcements			
	$OddBSR < 0.50$		$OddBSR < 0.25$		$OddBSR < 0.50$		$OddBSR < 0.25$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Negative	.013*	.014*	.011	.012	.064***	.039**	.053***	.030*
(s.e.)	(.007)	(.007)	(.008)	(.007)	(.021)	(.017)	(.017)	(.016)
Positive	-.013**	-.011*	-.010	-.009	-.030**	-.013	-.027**	-.012
	(.006)	(.006)	(.006)	(.006)	(.012)	(.010)	(.012)	(.011)
RoundBSR		-5.87***		-5.82***		-6.00***		-5.40***
		(0.16)		(0.15)		(0.23)		(0.16)
RatioSize	-.580***	-.675***	2.23***	2.139***	-1.09***	-1.24***	4.94***	4.81***
	(.063)	(.090)	(0.12)	(0.14)	(0.02)	(0.10)	(.210)	(0.21)
Articles	-.033***	-.021***	-.031***	-.020***	-.049***	-.037***	-.038***	-.027***
	(.003)	(.003)	(.003)	(.003)	(.007)	(.005)	(.007)	(.005)
$Neg - Pos$	.026***	.025***	.022**	.021**	.094***	.052**	.080***	.042*
(s.e.)	(.010)	(.010)	(.011)	(.010)	(.029)	(.024)	(.027)	(.025)
R^2	.043	.218	.054	.229	.057	.234	.117	.284
Observations	99,803	99,803	99,803	99,803	35,058	35,058	35,058	35,058

Panel B. ESS and Earnings Announcements

	Under				Over			
	Raw		Abnormal		Raw		Abnormal	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Negative	.001	.001	.002**	.001	-.001	-.001	-.001	-.001
(s.e.)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Positive	-.001	-.001**	-.001	-.001**	.002**	.001	.002*	.001
	(.000)	(.000)	(.001)	(.000)	(.001)	(.001)	(.001)	(.001)
JumpSize		.154***		.149***		.180***		.190***
		(.021)		(.020)		(.015)		(.015)
Articles	.000	.000*	.001	.000	.000	.000	.000	.000
	(.000)	(.000)	(.000)	(.000)	(.003)	(.000)	(.000)	(.000)
$Neg - Pos$	.002*	.002**	.003**	.002*	-.003**	-.002**	-.003*	-.002
(s.e.)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
R^2	.098	.229	.121	.235	.141	.237	.140	.243
Observations	44,722	44,722	44,722	44,722	44,722	44,722	44,722	44,722

the probability of selling pressure from retail investors is increased by between 2 and 10 percentage points. This finding supports the use of ESS as a proxy for the initial signal described in the model.

In addition to being related to retail trading, ESS should also be related to prices moving away from fundamental value. This relationship is investigated using techniques similar to equation (4.2) and examining price movements around earnings announcements:

$$Under_{it} = \beta_1 Negative_{it-1} + \beta_2 Positive_{it-1} + \beta_3 X_{it} + \zeta_i + \phi_t + \epsilon_{it}. \tag{4.3}$$

In this setup, $Under$ equals one when a stock experiences a price jump above 25% at its earnings announcement.[2] Thus, $Under$ indicates undervaluation, or scenarios where prices jump upward to reflect the new fundamental information provided in the earnings announcement. The $Negative$ and $Positive$ variables have been previously described. However, in this test they are lagged such that they measure news sentiment in a four week period, but do not include the two weeks immediately prior to the earnings announcement. The X matrix contains similar control variables to the previous tests, with the addition of $JumpSize$, which is the absolute value of the earnings announcement price jump.

If ESS is a reasonably proxy for the initial signal, negative sentiment should increase the probability that stocks will be undervalued while positive sentiment will decrease that probability. Thus, following equation (4.3), β_1 should be greater than β_2. This would indicate that when compared to positive sentiment, negative sentiment increases the probability of undervaluation. Results are shown in Panel B of Table 4.1. The results show that when compared to positive sentiment, negative sentiment increases (decreases) the probability of undervaluation (overvaluation). These results hold whether measuring earnings announcement returns using raw returns or abnormal returns.[3]

[2] Defining $Under$ using jumps between 20% and 50% provide similar results.

[3] In Panel B of Table 4.1, $Over$ is the opposite of $Under$.

Together, finding that ESS derived from RavenPack is related to both retail trading and mispricing provides support for its use as a proxy for the model's initial signal, S_η.

Section 4.2. Information Acquisition and the Sentiment of Recent News - Robustness

Table 3.1 tests the differential effect of negative and positive sentiment on institutional information acquisition using contemporaneous measures of ESS and $Learning$. In other words, in the monthly setting $Learning$ and ESS were both measured in the same month. However, it could be the case that learning takes place at the beginning of the month while news articles cluster near the end of the month. Also, it could be the case that information acquisition is driving news sentiment, not the other way around. To control for these hypotheticals, the tests from Table 3.1 are repeated using lagged ESS. That is, equation (3.1) is amended by using $Negative$ and $Positive$ from period $t-1$. As can be seen in the first four columns of Table 4.2, similar results are obtained. Column (2) shows that mutual funds increase their requests relative to hedge funds by almost 3 requests when news sentiment is negative, which is an increase of 6%. Thus, it can be said that negative sentiment in April leads mutual funds to acquire relatively more information in May.

Also in Table 3.1, ESS of 50 was considered neutral. However, average ESS over the sample is 52, suggesting that 52 instead of 50 may better indicate neutral sentiment. With this in mind, the last four columns of Table 4.2 show results from measuring $Negative$ and $Positive$ considering ESS scores of 52 as neutral. As can be seen, the results hold and Prediction 1 is again confirmed.

Recall that $Negative$ and $Positive$ have been defined as news sentiment at least three units away from neutral. It could be the case that other cutoffs provide better definitions of negative and positive news sentiment. Table 4.3 reproduces the results from Table 3.1 while using cutoffs of two and four units away. The results are quantitatively and qualitatively similar to those presented in Table 3.1.

RavenPack measures how relevant news releases are to a given stock, with their relevance metric ranging from zero to 100. Relevance has not been considered in the results presented

Table 4.2: Information Acquisition and the Sentiment of Recent News - Lags and Neutral

This table presents results from testing the relationship between information acquisition and the sentiment of recent news while accounting for lagged news and using a different value for neutral. The tests use the regression model in equation (3.1). Odd-numbered columns utilize the monthly setup. Even-numbered columns define t as the 30 days prior to an earnings announcement date. The dependent variable, $Diff$ is measured with either $Requests$ or IPs. The variable ESS_{it} captures the sentiment of recent news regarding stock i during period t. In the first four columns, ESS is measured similar to Table 3.1, but is lagged one period, thus ESS_{it-1} is used to derive $Negative$ and $Positive$. In the last four columns, ESS is used contemporaneously, but 52 instead of 50 is considered neutral sentiment. ESS is measured using the median sentiment rating from RavenPack. For columns (1) through (4), the indicator variable $Negative$ takes on the value of one if $ESS \leq 47$ and the indicator variable $Positive$ takes on the value of one if $ESS \geq 53$. For columns (5) through (8), the indicator variable $Negative$ takes on the value of one if $ESS \leq 49$ and the indicator variable $Positive$ takes on the value of one if $ESS \geq 55$. The table presents $Neg - Pos$ as a test of whether negative and positive news induce differential information acquisition. The matrix X_{it} contains the natural logarithm of the following control variables: the number of news releases ($Articles$), total trading volume ($Volume$), average market capitalization over the time period (not shown), the number of owning institutional investors ($No.\,Inst.$), and total institutional ownership of shares ($Inst.\,Own.$). The X matrix also contains the average daily price volatility over the period ($Volatility$) and indicator variables for whether period t contains or corresponds to stock i's quarterly or annual filings. Stock and year-month fixed effects are included. Standard errors are clustered by stock and year-month. The sample is made up of approximately 3,300 stocks and 66 months.

	Lagged ESS				Neutral ESS = 52			
	Requests		*IPs*		*Requests*		*IPs*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Negative	-0.85	0.69	0.14	0.19	3.52***	5.83***	1.24***	1.72***
(s.e.)	(0.66)	(0.75)	(0.12)	(0.14)	(0.77)	(1.53)	(0.16)	(0.24)
Positive	-0.58	-2.11***	-0.41***	-0.31**	-3.28***	-2.93**	-0.81***	-1.11***
	(0.54)	(0.67)	(0.08)	(0.12)	(0.63)	(1.40)	(0.12)	(0.21)
Articles	4.62***	1.95***	1.35***	0.56***	4.26***	6.11***	1.26***	1.64***
	(0.39)	(0.28)	(0.08)	(0.05)	(0.38)	(0.60)	(0.09)	(0.10)
Volume	5.21***	3.60***	1.47***	1.05***	4.44***	4.98***	1.25***	1.29***
	(0.51)	(0.43)	(0.12)	(0.08)	(0.45)	(0.50)	(0.10)	(0.09)
No. Inst.	1.01	1.33	0.13	-0.01	0.91	1.14	0.04	-0.03
	(1.11)	(1.07)	(0.21)	(0.22)	(0.99)	(1.15)	(0.19)	(0.18)
Inst. Own.	-0.53	-0.68	-0.23*	0.01	-0.22	-0.48	-0.15	-0.03
	(0.64)	(0.53)	(0.12)	(0.11)	(0.55)	(0.56)	(0.10)	(0.09)
Volatility	-1.04***	-0.38*	-0.21***	-0.02	-0.91***	-0.68**	-0.19***	-0.07
	(0.25)	(0.20)	(0.06)	(0.04)	(0.23)	(0.27)	(0.05)	(0.05)
$Neg - Pos$	-0.27	2.80**	0.55***	0.51**	6.79***	8.76***	2.05***	2.82***
(s.e.)	(0.86)	(1.21)	(0.15)	(0.21)	(1.20)	(2.62)	(0.23)	(.395)
R^2	.566	.507	.797	.770	.568	.601	.799	.820
Observations	127,116	39,352	127,116	39,352	141,146	48,782	141,146	48,782

Table 4.3: Information Acquisition and the Sentiment of Recent News - Cutoffs

This table presents results from testing the relationship between information acquisition and the sentiment of recent news while using various cutoffs to define *Negative* and *Positive*. The tests use the regression model in equation (3.1). Odd-numbered columns utilize the monthly setup. Even-numbered columns define t as the 30 days prior to an earnings announcement date. The dependent variable, $Diff$ is measured with either *Requests* or IPs. The variable ESS_{it} captures the sentiment of recent news regarding stock i during period t. ESS is measured using the median sentiment rating from RavenPack. The indicator variable *Negative* takes on the value of one if $ESS \leq 50 - C$. The indicator variable *Positive* takes on the value of one if $ESS \geq 50 + C$. The first four columns use $C = 2$ while the last four columns use $C = 4$ as the cutoffs between negative, neutral, and positive sentiment. The table presents $Neg - Pos$ as a test of whether negative and positive news induce differential information acquisition. The matrix X_{it} contains the natural logarithm of the following control variables: the number of news releases (*Articles*), total trading volume (*Volume*), average market capitalization over the time period (not shown), the number of owning institutional investors (*No. Inst.*), and total institutional ownership of shares (*Inst. Own.*). The X matrix also contains the average daily price volatility over the period (*Volatility*) and indicator variables for whether period t contains or corresponds to stock i's quarterly or annual filings. Stock and year-month fixed effects are included. Standard errors are clustered by stock and year-month. The sample is made up of approximately 3,300 stocks and 66 months.

	Cutoff = 2				Cutoff = 4			
	Requests		IPs		*Requests*		IPs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Negative	0.44	3.72***	0.57***	0.99***	1.80**	5.03***	0.84***	1.16***
(s.e.)	(0.72)	(1.37)	(0.16)	(0.22)	(0.76)	(1.52)	(0.16)	(0.25)
Positive	-0.03	-1.40	-0.01	-0.36**	-2.49***	-2.60**	-0.60***	-0.71***
	(0.64)	(0.99)	(0.10)	(0.15)	(0.60)	(1.14)	(0.09)	(0.18)
Articles	4.31***	6.17***	1.27***	1.66***	4.28***	6.17***	1.27***	1.66***
	(0.38)	(0.60)	(0.09)	(0.10)	(0.38)	(0.60)	(0.09)	(0.10)
Volume	4.45***	5.01***	1.25***	1.29***	4.45***	5.00***	1.25***	1.29***
	(0.45)	(0.50)	(0.10)	(0.09)	(0.45)	(0.50)	(0.11)	(0.09)
No. Inst.	0.87	1.15	0.03	-0.03	0.91	1.16	0.04	-0.03
	(0.99)	(1.15)	(0.19)	(0.19)	(0.98)	(1.15)	(0.19)	(0.18)
Inst. Own.	-0.23	-0.50	-0.15	-0.04	-0.23	-0.51	-0.15	-0.04
	(0.55)	(0.56)	(0.10)	(0.09)	(0.55)	(0.56)	(0.10)	(0.09)
Volatility	-0.94***	-0.69**	-0.19***	-0.08	-0.92***	-0.68**	-0.19***	-0.07
	(0.23)	(0.27)	(0.05)	(0.05)	(0.23)	(0.27)	(0.05)	(0.05)
$Neg - Pos$	0.47	5.12**	0.58***	1.36***	4.29***	7.63***	1.44***	1.87***
(s.e.)	(1.16)	(2.07)	(0.21)	(0.33)	(1.17)	(2.36)	(0.20)	(0.38)
R^2	.568	.601	.799	.820	.568	.601	.799	.820
Observations	141,146	48,782	141,146	48,782	141,146	48,782	141,146	48,782

thus far. Table 4.4 takes relevance into account by limiting the set of news articles to those with relevance scores of at least 80 or 90. The results shown in Table 4.4 are qualitatively unchanged from Table 3.1.

It may be the case that the relationship between news sentiment and information acquisition is different based on the size of the stock. It may also be the case that the results found in the previous tables are driven by one subset of stocks, either small or large. To explore these ideas, Table 4.5 provides results using size-based subsamples. Large, small, and micro stocks are identified using the NYSE Breakpoints following Fama and French (2012). The relationship between information acquisition and sentiment holds within each subsample.

Finally, in unreported results the same tests were repeated using different years within the 2012 through 2017 period and by measuring ESS using mean instead of median sentiment scores. In all cases, the results hold and provide the same qualitative interpretations as Table 3.1. Overall, the finding that mutual funds acquire relatively more information than hedge funds when sentiment is negative is robust.

Section 4.3. Price Informativeness - Robustness

Similar to the previous section, the robustness checks from Section 4.2 have been applied to the results in Table 3.2, and again, the main findings hold. To further test the price informativeness predictions and the findings in Table 3.2, this section uses $Requests$ as the measure of $Learning$. Table 4.6 recreates Table 3.2 using $Requests$ instead of IPs. The results hold: increased learning by either hedge funds or mutual funds improves price informativeness. As relatively more mutual funds become informed compared to hedge funds, PJR and PJ are reduced. This finding provides additional support to Prediction 3. Furthermore, Table 4.6 shows that price jumps are lower following negative sentiment than following positive; providing additional support for Prediction 2.

Price jumps have also been analyzed within size-based subsamples. Table 4.7 shows that the previous findings generally hold for large, small, and micro stocks. Indeed, prices are between 11% and 15% more informative following negative sentiment than positive.

Table 4.4: Information Acquisition and the Sentiment of Recent News - Relevance

This table presents results from testing the relationship between information acquisition and the sentiment of recent news while accounting for the relevance of news. The tests use the regression model in equation (3.1). Odd-numbered columns utilize the monthly setup. Even-numbered columns define t as the 30 days prior to an earnings announcement date. The dependent variable, $Diff$ is measured with either *Requests* or IPs. The variable ESS_{it} captures the sentiment of recent news regarding stock i during period t. In the first four columns, only news items with a relevance score of 80 or more are used to measures ESS. In the last four columns, only news with relevance of greater than 90 are used. ESS is measured using the median sentiment rating from RavenPack. The indicator variable *Negative* takes on the value of one if $ESS \leq 47$. The indicator variable *Positive* takes on the value of one if $ESS \geq 53$. The table presents $Neg - Pos$ as a test of whether negative and positive news induce differential information acquisition. The matrix X_{it} contains the natural logarithm of the following control variables: the number of news releases (*Articles*), total trading volume (*Volume*), average market capitalization over the time period (not shown), the number of owning institutional investors (*No. Inst.*), and total institutional ownership of shares (*Inst. Own.*). The X matrix also contains the average daily price volatility over the period (*Volatility*) and indicator variables for whether period t contains or corresponds to stock i's quarterly or annual filings. Stock and year-month fixed effects are included. Standard errors are clustered by stock and year-month. The sample is made up of approximately 3,300 stocks and 66 months.

	Relevance $\geq$ 80				Relevance $\geq$ 90			
	Requests		IPs		*Requests*		IPs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Negative	1.50**	4.07***	0.78***	1.10***	1.51**	4.10***	0.78***	1.11***
(s.e.)	(0.72)	(1.43)	(0.16)	(0.23)	(0.72)	(1.43)	(0.16)	(0.23)
Positive	-2.41***	-2.85**	-0.55***	-0.74***	-2.42***	-2.83**	-0.55***	-0.73***
	(0.59)	(1.10)	(0.09)	(0.17)	(0.59)	(1.10)	(0.09)	(0.17)
Articles	8.10***	13.8***	2.24***	3.54***	8.12***	14.0***	2.24***	3.58***
	(0.57)	(1.10)	(0.12)	(0.19)	(0.58)	(1.13)	(0.12)	(0.19)
Volume	4.16***	4.92***	1.20***	1.29***	4.16***	4.92***	1.20***	1.29***
	(0.43)	(0.48)	(0.10)	(0.09)	(0.43)	(0.48)	(0.10)	(0.09)
No. Inst.	0.84	1.27	0.03	0.01	0.83	1.24	0.03	-0.00
	(0.98)	(1.16)	(0.19)	(0.19)	(0.98)	(1.16)	(0.19)	(0.19)
Inst. Own.	-0.22	-0.61	-0.15	-0.07	-0.22	-0.61	-0.15	-0.07
	(0.55)	(0.56)	(0.10)	(0.09)	(0.55)	(0.56)	(0.10)	(0.09)
Volatility	-1.05***	-0.86***	-0.22***	-0.12**	-1.06***	-0.88***	-0.23***	-0.13**
	(0.23)	(0.27)	(0.05)	(0.05)	(0.23)	(0.27)	(0.05)	(0.05)
$Neg - Pos$	3.91***	6.91***	1.32***	1.83***	3.92***	6.94***	1.33***	1.84***
(s.e.)	(1.08)	(2.24)	(0.19)	(0.35)	(1.08)	(2.24)	(0.19)	(0.35)
R^2	.570	.603	.802	.822	.570	.603	.802	.822
Observations	141,146	48,782	141,146	48,782	141,146	48,782	141,146	48,782

Table 4.5: Information Acquisition and the Sentiment of Recent News - NYSE Breakpoints

This table presents results from testing the relationship between information acquisition and the sentiment of recent news and partitions the sample by size using NYSE Breakpoints. Columns (1) through (4) use large stocks only. Columns (5) through (8) use small stocks. Columns (9) through (12) use micro stocks. The tests use the regression model in equation (3.1). Odd-numbered columns utilize the monthly setup. Even-numbered columns define t as the 30 days prior to an earnings announcement date. The dependent variable, $Diff$ is measured with either *Requests* or IPs. The variable ESS_{it} captures the sentiment of recent news regarding stock i during period t. ESS is measured using the median sentiment rating from RavenPack. The indicator variable $Negative$ takes on the value of one if $ESS \leq 47$. The indicator variable $Positive$ takes on the value of one if $ESS \geq 53$. The table presents $Neg - Pos$ as a test of whether negative and positive news induce differential information acquisition. The matrix X_{it} contains the natural logarithm of the following control variables: the number of news releases ($Articles$), total trading volume ($Volume$), average market capitalization over the time period (not shown), the number of owning institutional investors ($No. Inst.$), and total institutional ownership of shares ($Inst. Own.$). The X matrix also contains the average daily price volatility over the period ($Volatility$) and indicator variables for whether period t contains or corresponds to stock i's quarterly or annual filings. Stock and year-month fixed effects are included. Standard errors are clustered by stock and year-month. The sample is made up of approximately 3,300 stocks and 66 months.

	Large				Small				Micro			
	Requests		IPs		*Requests*		IPs		*Requests*		IPs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Negative$	1.34	9.60**	0.45*	1.09*	3.02***	4.15*	1.06***	1.40***	1.23**	2.23**	0.67***	0.82***
(s.e.)	(1.78)	(4.55)	(0.25)	(0.63)	(0.85)	(2.50)	(0.18)	(0.45)	(0.50)	(1.06)	(0.11)	(0.19)
$Positive$	-4.58***	-5.68**	-0.86***	-0.94**	-2.05**	-2.52	-0.52***	-0.72**	-0.36	-0.60	-0.17**	-0.25
	(1.61)	(2.79)	(0.20)	(0.38)	(0.84)	(1.84)	(0.13)	(0.32)	(0.42)	(0.92)	(0.08)	(0.16)
$Articles$	8.88***	14.7***	2.36***	3.39***	2.98***	4.95***	1.03***	1.66***	1.91***	1.93***	0.69***	0.81***
	(0.89)	(1.92)	(0.18)	(0.31)	(0.39)	(1.03)	(0.08)	(0.17)	(0.21)	(0.45)	(0.05)	(0.08)
$Volume$	19.7***	21.39***	4.37***	4.73***	13.0***	13.0***	3.38***	3.08***	2.31***	2.85***	0.87***	0.85***
	(2.00)	(2.90)	(0.40)	(0.46)	(1.20)	(1.38)	(0.20)	(0.24)	(0.25)	(0.36)	(0.05)	(0.07)
$No. Inst.$	2.06	4.01	0.57	0.62	-2.07	0.78	-1.15**	-1.09*	0.98	0.75	0.16	0.14
	(2.27)	(3.95)	(0.44)	(0.67)	(3.03)	(3.78)	(0.52)	(0.63)	(0.68)	(0.94)	(0.11)	(0.14)
$Inst. Own.$	-3.65	-8.86**	-1.45**	-1.47**	-10.6***	-10.9***	-1.56***	-1.47***	-0.12	0.02	-0.04	-0.01
	(4.07)	(4.24)	(0.67)	(0.66)	(2.25)	(2.46)	(0.36)	(0.53)	(0.35)	(0.43)	(0.06)	(0.07)
$Volatility$	-1.11	-0.83	-0.07	-0.00	-2.40***	-1.54***	-0.54***	-0.31***	-1.28***	-1.17***	-0.38***	-0.25***
	(0.74)	(1.03)	(0.15)	(0.16)	(0.43)	(0.56)	(0.08)	(0.10)	(0.16)	(0.21)	(0.03)	(0.04)
$Neg - Pos$	5.92**	15.3***	1.31***	2.03**	5.07***	6.68*	1.58***	2.12***	1.59**	2.83*	0.84***	1.08***
(s.e.)	(3.04)	(6.19)	(0.39)	(0.89)	(1.40)	(3.71)	(0.25)	(0.67)	(0.78)	(1.74)	(0.16)	(0.30)
R^2	.541	.601	.759	.799	.352	.426	.611	.657	.335	.413	.618	.654
Observations	44,489	14,764	44,489	14,764	41,840	14,014	41,840	14,014	61,388	21,559	61,388	21,559

Table 4.6: Price Informativeness - *Requests*

This table presents results from testing whether information acquisition and the sentiment of recent news are related to price informativeness. The tests use the regression model in equation (3.3). Odd-numbered columns use PJR as the dependent variable while even-numbered columns use PJ. *Learning* is measured using *Requests* and is lagged one period in all specifications. ESS is measured using the median sentiment rating from RavenPack. The indicator variable $Negative$ takes on the value of one if $ESS \leq 47$. The indicator variable $Positive$ takes on the value of one if $ESS \geq 53$. The table presents $Neg - Pos$ as a test of whether negative and positive news induce differential price informativeness. The matrix X_{it} contains the natural logarithm of the following control variables: the number of news releases ($Articles$), total trading volume ($Volume$), average market capitalization over the time period (not shown), the number of owning institutional investors ($No.\ Inst.$), and total institutional ownership of shares ($Inst.\ Own.$). The X matrix also contains the average daily price volatility over the period ($Volatility$) and indicator variables for whether period t contains or corresponds to stock i's quarterly or annual filings. Stock and year-month fixed effects are included. Standard errors are clustered by stock and year-month. The sample is made up of approximately 3,300 stocks and 66 months.

	Price Jumps					
	Mutual Funds		Hedge Funds		Difference	
	(1)	(2)	(3)	(4)	(5)	(6)
Learning	-.013***	-.041***	-.006*	-.028***	-.0002***	-.001***
(s.e.)	(.003)	(.008)	(.003)	(.006)	(.0001)	(.000)
$Negative$	-.022*	-.079***	-.022*	-.079***	-.023**	-.081***
	(.011)	(.026)	(.011)	(.026)	(.011)	(.026)
$Positive$	.018	.064***	.018	.065***	.018*	.064***
	(.011)	(.019)	(.011)	(.019)	(.011)	(.019)
$Articles$	-.019***	-.047***	-.019***	-.048***	-.019***	-.050***
	(.004)	(.009)	(.004)	(.009)	(.004)	(.009)
$Volume$	-.051***	.036***	-.052***	.034**	-.053***	.030**
	(.006)	(.013)	(.006)	(.013)	(.006)	(.013)
$No.\ Inst.$	-.017	.073**	-.017	.070**	-.017	.073**
	(.017)	(.034)	(.017)	(.033)	(.017)	(.033)
$Inst.\ Own.$	.018*	-.017	.018*	-.017	.018*	-.017
	(.010)	(.017)	(.010)	(.017)	(.010)	(.017)
$Volatility$	.000	.098***	.000	.098***	.000	.098***
	(.003)	(.007)	(.003)	(.007)	(.003)	(.007)
$Neg - Pos$	-.040**	-.143***	-.040**	-.143***	-.041**	-.146***
(s.e.)	(.018)	(.037)	(.018)	(.037)	(.018)	(.037)
R^2	.304	.233	.304	.232	.304	.232
Observations	10,613	39,643	10,613	39,643	10,613	39,643

Table 4.7: Price Informativeness - NYSE Breakpoints

This table presents results from testing whether information acquisition and the sentiment of recent news are related to price informativeness when measured using price jumps at earnings announcements and partitioning the sample by size using NYSE Breakpoints. Panel A uses large stocks only. Panel B uses small stocks. Panel C uses micro stocks. The tests use the regression model in equation (3.3). Odd-numbered columns use *Requests* and the *Learning* variable. Even-numbered columns use IPs. *Learning* is lagged one period in all specifications. The dependent variable, PJ, measures price informativeness. ESS is measured using the median sentiment rating from RavenPack. The indicator variable $Negative$ takes on the value of one if $ESS \leq 47$. The indicator variable $Positive$ takes on the value of one if $ESS \geq 47$. The table presents $Neg - Pos$ as a test of whether negative and positive news induce differential price informativeness. The matrix X_{it} contains the natural logarithm of the following control variables: the number of news releases (*Articles*), total trading volume ($Volume$), average market capitalization over the time period (not shown), the number of owning institutional investors ($No.Inst.$), and total institutional ownership of shares ($Inst.Own.$). The X matrix also contains the average daily price volatility over the period ($Volatility$) and indicator variables for whether period t contains or corresponds to stock i's quarterly or annual filings. Stock and year-month fixed effects are included. Standard errors are clustered by stock and year-month. The sample is made up of approximately 3,300 stocks and 66 months.

	Mutual Funds		Hedge Funds		Difference	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Large Stocks						
Learning	-.059***	-.103***	.022*	.036	-.000	-.003*
(s.e.)	(.016)	(.026)	(.012)	(.022)	(.000)	(.002)
Negative	-.073*	-.073*	-.073*	-.073*	-.081**	-.081**
	(.041)	(.040)	(.041)	(.041)	(.040)	(.040)
Positive	.060*	.061*	.062*	.062*	.071**	.070**
	(.035)	(.035)	(.035)	(.035)	(.035)	(.035)
$Neg - Pos$	-.133**	-.134**	-.134**	-.135**	-.152**	-.151**
(s.e.)	(.062)	(.062)	(.062)	(.062)	(.062)	(.062)
Panel B. Small Stocks						
Learning	-.001	-.006	.002	-.013	-.001*	-.010***
(s.e.)	(.014)	(.022)	(.012)	(.024)	(.001)	(.003)
Negative	-.110**	-.110**	-.110**	-.109**	-.095**	-.093**
	(.042)	(.042)	(.042)	(.042)	(.043)	(.043)
Positive	.021	.021	.021	.021	.024	.023
	(.037)	(.037)	(.037)	(.037)	(.035)	(.035)
$Neg - Pos$	-.131*	-.130*	-.131*	-.130*	-.118*	-.116*
(s.e.)	.070	.070	.069	.069	.067	.067
Panel C. Micro Stocks						
Learning	-.023	-.050**	-.007	-.032	-.002***	-.011***
(s.e.)	(.015)	(.022)	(.011)	(.022)	(.001)	(.004)
Negative	-.058	-.058	-.059	-.059	-.070*	-.070*
	(.040)	(.040)	(.040)	(.040)	(.038)	(.038)
Positive	.057	.057	.058	.058	.056	.056
	(.036)	(.036)	(.036)	(.036)	(.035)	(.035)
$Neg - Pos$	-.114*	-.114*	-.117*	-.117*	-.126**	-.127**
(s.e.)	.059	.059	.059	.059	.057	.058

Section 4.4. Information Acquisition and Other Investors - Robustness

The robustness checks from the previous section have been applied to the results in Table 3.3. The results quantitatively and qualitatively remain unchanged. Table 3.3 defines *Low* and *High* as the lower and upper quartiles, deciles, and terciles of expected learning by other investors. Upon observing Table 3.3, it could be said that the results are stronger when using a more strict definition of *Low* and *High* (i.e., when using deciles instead of quartiles or terciles). However, the main takeaway is that the results remain qualitatively the same. That is, when it comes to acquiring information, hedge funds care more about what mutual funds are doing than vice versa.

As another robustness check and to further study Prediction 4, the analysis is repeated on sized-based subsamples. The results in Table 4.8 show that the relationship described in Prediction 4 is strongest among large and small stocks.

Table 4.8: Information Acquisition and Other Investors - NYSE Breakpoints

This table presents results from testing whether hedge funds and mutual funds condition their information acquisition on expected learning from their each other and partitions the sample by size using NYSE Breakpoints. Panel A uses large stocks only. Panel B uses small stocks. Panel C uses micro stocks. The tests utilize the regression model in equation (3.5) to calculate the expected level of other investor type learning. Then the regression model in equation (3.6) is used to test whether investors learn more when it is expected that the other investor type will learn less. Odd-numbered columns define t in monthly terms. Even-numbered columns define t as the time period 15 days prior to an earnings announcement date while $t-1$ corresponds to 15 day period just prior to t. The dependent variables $Learning^h$ and $Learning^m$ are measured using either $Requests$ or IPs. The matrix X_{it} contains the natural logarithm of the following control variables: the number of news releases ($Articles$), total trading volume ($Volume$), average market capitalization over the time period (not shown), the number of owning institutional investors ($No.\,Inst.$), and total institutional ownership of shares ($Inst.\,Own.$). The X matrix also contains the average daily price volatility over the period ($Volatility$) and indicator variables for whether period t contains or corresponds to stock i's quarterly or annual filings. Low is equal to one when learning from the other investor type for a given stock is predicted to be in the lower quartile. $High$ is equal to one when learning from the other investor type for a given stock is predicted to be in the upper quartile. The table presents $Low - High$ as a test of whether low expected learning from others results in different information acquisition than when learning from others is expected to be high. The control variables contained in Y include $Negative$, $Positive$, and X. Stock and year-month fixed effects are included. Standard errors are clustered by stock and year-month and the Murphy-Topel correction has been applied for predicted regressors. The sample is made up of approximately 3,300 stocks and 66 months.

	Mutual Funds				Hedge Funds			
	Requests		*IPs*		*Requests*		*IPs*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Large Stocks								
Low	.016	.020	.030	.070**	.380**	.346***	.323***	.330***
(s.e.)	(.032)	(.047)	(.021)	(.028)	(.145)	(.107)	(.091)	(.071)
High	-.055	-.044	.013	-.018	-.083	-.128	-.127**	-.163**
	(.036)	(.049)	(.030)	(.040)	(.081)	(.112)	(.054)	(.062)
$Low - High$	.071	.064	.017	.089*	.463***	.474***	.450***	.493***
(s.e.)	(.052)	(.062)	(.038)	(.047)	(.164)	(.154)	(.103)	(.090)
R^2	.677	.626	.783	.742	.588	.565	.698	.656
Observations	42,473	13,956	42,473	13,956	42,473	13,956	42,473	13,956
Panel B. Small Stocks								
Low	-.016	.018	.026	.040	.254**	.292**	.207***	.190**
(s.e.)	(.027)	(.047)	(.018)	(.031)	(.121)	(.123)	(.078)	(.078)
High	.010	.042	.018	-.023	-.046	-.052	-.061	-.078
	(.032)	(.052)	(.023)	(.039)	(.084)	(.107)	(.038)	(.059)
$Low - High$	-.026	-.024	.008	.063	.300**	.344***	.268***	.268***
(s.e.)	(.042)	(.070)	(.030)	(.047)	(.144)	(.167)	(.086)	(.092)
R^2	.564	.494	.664	.587	.523	.502	.601	.565
Observations	37,838	11,926	37,838	11,926	37,838	11,926	37,838	11,926
Panel C. Micro Stocks								
Low	.039	.123***	.048**	.093***	.257***	.136	.184***	.125***
(s.e.)	(.024)	(.032)	(.020)	(.025)	(.075)	(.084)	(.042)	(.046)
High	.003	-.028	-.013	-.046	.025	.088	-.038	-.031
	(.027)	(.040)	(.021)	(.030)	(.049)	(.067)	(.025)	(.037)
$Low - High$	.035	.152***	.061**	.138***	.232**	.048	.223***	.157***
(s.e.)	(.038)	(.054)	(.031)	(.040)	(.092)	(.106)	(.050)	(.058)
R^2	.575	.496	.649	.571	.468	.450	.504	.470
Observations	48,900	13,480	48,900	13,480	48,900	13,480	48,900	13,480

CHAPTER 5: CONCLUSION

This paper extends the classic Grossman and Stiglitz (1980) setting to explore the interaction between short-sale prohibitions and information acquisition. When news generates differences of opinions, constrained investors will be relatively optimistic or pessimistic. When relatively pessimistic, they expect to be bound by their short-sale prohibition. Fewer constrained investors acquire costly information in this scenario since they anticipate being unable to trade. In response, more unconstrained investors acquire information.

The model makes several unique predictions. First, relative to hedge funds, fewer mutual funds acquire information following news with positive sentiment than following news with negative sentiment. Second, hedge fund information acquisition is more sensitive to the information acquisition of mutual funds than vice versa. Third, since the informativeness of prices depends on information acquisition, the model also highlights a link between the sentiment of recent news and the information content of prices: prices are more informative following negative sentiment than following positive sentiment.

A new dataset is employed to verify the model's predictions. This dataset was derived from records of activity on the EDGAR filing system and a hand-collected sample of IP addresses from hedge funds and mutual funds. Using direct measures of mutual fund and hedge fund information acquisition, tests verify that relatively more mutual fund's acquire information following news with negative sentiment than positive. Also, the data show that hedge funds are more sensitive to the information acquisition of mutual funds than vice versa. Finally, tests verify that prices are more informative when more information is acquired and when the sentiment of recent news has been negative.

Overall, this paper highlights an important link between short-sale prohibitions and information acquisition. Namely, in the presence of short sale prohibited investors, information acquisition is conditional on the sentiment of recent news.

APPENDIX A: UPDATING BY UNINFORMED MUTUAL FUNDS AND HEDGE FUNDS

Deriving $\mathbb{E}[\theta|P]$ entails following the logic developed in Section 2.2. The uninformed do not know which regime will obtain, but they do know that there are only two possibilities. Using this insight, the uninformed can update by considering being in either regime, and accounting for the probability that either one obtains.

Being in this Regime 1 means (i) informed mutual funds demand positive shares and (ii) informed mutual funds believe the asset to be undervalued. The observable signal, denoted here by S_{in}, can be derived from equation (2.6) by removing the maximization function since informed mutual funds have positive demand:

$$S_{in} = u - \frac{\nu\mu}{\gamma}\big(\theta - P\big) - \frac{\omega\delta}{\gamma}\Big(\mathbb{E}\big[\theta|S_\alpha, P\big] - P\Big). \tag{A.1}$$

This signal can be refined by expanding $\mathbb{E}\big[\theta|S_\alpha, P\big]$ using equations (2.9) and (2.10):

$$S_1 = \underbrace{\Big[P(\nu\mu + \omega\delta) - S_{in}\gamma\Big]\left[\frac{\tau_\theta + \tau_\alpha + \tau_I}{\big(\nu\mu + \omega\delta\big)\big(\tau_\alpha + \tau_I\big) + \nu\mu\tau_\theta}\right]}_{\text{Observable Signal}} = \theta + Z_1. \tag{A.2}$$

Thus, the price has been transformed into a signal equal to the payoff plus Z_1, where,

$$Z_1 = \frac{\alpha\Big(\omega\delta\tau_\alpha\Big) - u\Big(\frac{\gamma}{\nu\mu}\Big)\Big[(\nu\mu + \omega\delta)\tau_I + \nu\mu(\tau_\theta + \tau_\alpha)\Big]}{\big(\nu\mu + \omega\delta\big)\big(\tau_\alpha + \tau_I\big) + \nu\mu\tau_\theta}. \tag{A.3}$$

Notably, Z_1 contains only two unknowns, α and u. These two variables are independent normal random variables and are combined additively after being multiplied by known scalars. As such, Z_1 is a mean-zero normal random variable with precision τ_1. To see τ_1, rewrite equation (A.3) using scalars A and B:

$$Z_1 = A\alpha + Bu. \tag{A.4}$$

Since A and B are known and α and u are independent normal random variables, the precision of Z_1 is as follows:

$$\tau_1 = \left(\frac{A^2}{\tau_\alpha} + \frac{B^2}{\tau_u}\right)^{-1}. \tag{A.5}$$

Being in Regime 2 means (i) informed mutual funds do not trade and (ii) informed mutual funds believe the asset to be overvalued. The observable signal, denoted here by S_{out}, can be written from equation (2.6) by replacing the maximization function with zero since informed mutual funds do not trade:

$$S_{out} = u - \frac{\nu\mu}{\gamma}\big(\theta - P\big). \tag{A.6}$$

This signal can be refined into the sum of θ and noise:

$$S_2 = \underbrace{\Big[P\nu\mu - S_{out}\gamma\Big]\left[\frac{1}{\nu\mu}\right]}_{\text{Observable Signal}} = \theta + Z_2. \tag{A.7}$$

Z_2 is a mean-zero normal random variable with precision τ_2. Deriving Z_2 and τ_2 is shown as follows:

$$Z_2 = -u\left(\frac{\gamma}{\nu\mu}\right), \tag{A.8}$$

$$\tau_2 = \left(\frac{\nu\mu}{\gamma}\right)^2 \tau_u. \tag{A.9}$$

Uninformed investors observe only one signal from the price, $S = S_{in} = S_{out}$. However, they can transform this signal into two different refined signals, S_1 and S_2. Uninformed investors cannot simply use the refined signal from the regime that is more likely, since there remains some probability that the observable signal comes from the less-likely regime. Updating from the price requires using the explicit function for $\mathbb{E}\big[\theta|P\big]$ and giving consideration to both regimes. Indeed, the goal of understanding the price's signal to uninformed investors is to calculate $\mathbb{E}[\theta|P]$, which can be written by definition as follows:

$$\mathbb{E}\big[\theta|P\big] = \int_{-\infty}^{\infty} \theta f(\theta|P) d\theta = \int_{-\infty}^{\infty} \theta \frac{f(\theta, P)}{f(P)} d\theta. \tag{A.10}$$

The joint density of θ and P can be expanded to recognize that the price can exist under two regimes:

$$f(\theta, P) = f(\theta, P, \underbrace{\mathbb{E}[\theta|S_\alpha, P] > P}_{\text{Regime 1}}) + f(\theta, P, \underbrace{\mathbb{E}[\theta|S_\alpha, P] \leq P}_{\text{Regime 2}}). \tag{A.11}$$

To explicitly write out the density functions comprising equation (A.11), note that each contains three random variables that remain unknown to the uninformed, θ, α, and u. The joint density of the payoff, the price, and the realization of Regime 1, for example, is the probability that θ, α, and u combine such that a given θ is observed, Regime 1 is realized, and a given price is obtained (equivalent to S_1 being observed). This joint probability density can be written as follows:

$$f(\theta, P, \mathbb{E}[\theta|S_\alpha, P] > P) = f(\theta, S_1, \text{Regime 1}) = f_\theta(\theta) \int_{-\infty}^{\infty} f_\alpha(\alpha) F_u(\tilde{u}) f_u(u_1) d\alpha. \tag{A.12}$$

This density can be understood as follows: First, take the probability of obtaining any given payoff, $f_\theta(\theta)$. For every payoff there exists infinite combinations of α and u that produce the observed price (or S_1) and Regime 1. Thus, the density function integrates over every possible combination by integrating over every possible α. For every (θ, α) pair, the probability that Regime 1 obtains is equivalent to the probability that u is below a certain threshold, denoted $\tilde{u}$, which is the cutoff point between regimes. That is, $\tilde{u}$ is the value at which $\mathbb{E}[\theta|S_\alpha, P] = P$. Finally, for every (θ, α) pair, there exists only one possible u that will produce the observed S_1 in Regime 1. When matched with the (θ, α) pair, u_1 is the value that produces the S_1 observed.

The joint density of the payoff, the price, and the realization of Regime 2 can be written in similar fashion:

$$f(\theta, P, \mathbb{E}[\theta|S_\alpha, P] \leq P) = f(\theta, S_2, \text{Regime 2}) = f_\theta(\theta) \int_{-\infty}^{\infty} f_\alpha(\alpha)[1 - F_u(\tilde{u})] f_u(u_2) d\alpha. \quad \text{(A.13)}$$

The joint density of the payoff and the price can be seen by combining equations (A.12) and (A.13):

$$f(\theta, P) = f_\theta(\theta) \int_{-\infty}^{\infty} f_\alpha(\alpha) \Big[F_u(\tilde{u}) f_u(u_1) + [1 - F_u(\tilde{u})] f_u(u_2) \Big] d\alpha. \quad \text{(A.14)}$$

Finally, the marginal density of the price, needed in the denominator of equation (A.10), can be found by integrating equation (A.14) over θ. Thus, using equations (A.10) through (A.14), uninformed investors can update their beliefs using the price.

The variables used in equation (A.14) are derived below. As noted, u_1 is the value that when combined with a given θ and α produces S_1. Recall, equation (A.2) shows that S_1 is the sum of θ with Z_1. Using this identity combined with the derivation of Z_1 in equation (A.3), u_1 can be written as follows:

$$u_1 = \left[\theta - S_1 + \alpha \Big(\frac{\omega\delta\tau_\alpha}{(\nu\mu + \omega\delta)(\tau_\alpha + \tau_I) + \nu\mu\tau_\theta} \Big) \right] \left[\frac{(\nu\mu + \omega\delta)(\tau_\alpha + \tau_I) + \nu\mu\tau_\theta}{(\nu\mu + \omega\delta)\tau_I + \nu\mu(\tau_\theta + \tau_\alpha)} \right] \left[\frac{\nu\mu}{\gamma} \right]. \quad \text{(A.15)}$$

Similarly, u_2 is the value that when combined with a given θ produces S_2. Using equations (A.7) and (A.8), u_2 can be seen as follows:

$$u_2 = \Big[\theta - S_2 \Big] \frac{\nu\mu}{\gamma}. \quad \text{(A.16)}$$

Finally, $\tilde{u}$ is the cutoff point between Regime 1 and Regime 2. That is, $\tilde{u}$ is the value that when combined with θ and α equates the price with $\mathbb{E}[\theta|S_\alpha, P]$. The complete form of

$\tilde{u}$ can be shown using equations (B.2a), (B.2b), and (B.3):

$$\tilde{u} = \Big[\theta\big(\tau_\alpha + \tau_I\big) + \alpha\tau_\alpha - P\big(\tau_\theta + \tau_\alpha + \tau_I\big)\Big]\Big(\frac{\nu\mu}{\gamma\tau_I}\Big). \tag{A.17}$$

APPENDIX B: UPDATING BY INFORMED MUTUAL FUNDS

At the trading stage, informed mutual funds have observed the initial signal (S_η), the noisy signal (S_α), and the price, thus they know their own demand as well as the demand from retail investors and the uninformed. Informed mutual funds remain uncertain, however, regarding the demand from informed hedge funds and noisy supply. With this in mind, informed mutual funds consider what they know upon receipt of the price to form an observable signal. They produce this signal by rearranging the market clearing condition according to what they know and what remains unknown:

$$\underbrace{-\nu(1-\mu)X_{HP} - \omega\delta X_{MS} - \omega(1-\delta)X_{MP} - X_R}_{\text{Known/Observable}} = \underbrace{\nu\mu X_{H\theta} - u}_{\text{Unknown}}. \tag{B.1}$$

This signal can by refined by expanding the optimal demand function of informed hedge funds into known and unknown elements:

$$S_I = \underbrace{\Big(-\nu(1-\mu)X_{HP} - \omega\delta X_{MS} - \omega(1-\delta)X_{MP} - X_R\Big)\Big(\frac{\gamma}{\nu\mu}\Big) + P}_{\text{Observable Signal}} \tag{B.2a}$$

$$= \theta - \Big(\frac{\gamma}{\nu\mu}\Big)u. \tag{B.2b}$$

Informed mutual funds can transform the price into a signal that is the sum of θ and noise. Equipped with S_I, informed mutual funds update their beliefs on θ:

$$\mathbb{E}\big[\theta|S_\alpha, P\big] = \frac{\tau_\alpha S_\alpha + \tau_I S_I}{\tau_\theta + \tau_\alpha + \tau_I}, \tag{B.3}$$

where $\tau_I = (\frac{\nu\mu}{\gamma})^2\tau_u$.

APPENDIX C: STEPS IN NUMERICAL SIMULATION

Numerical simulation is required to solve for equilibrium learning and produce comparative statics. The steps to solve for equilibrium are detailed below. The parameters used in the simulation are summarized in Table **??**.

1. Set model parameters (i.e., masses, precisions, and costs). Fix δ and pick an initial signal, S_η.

2. Make a conjecture for the portion of informed hedge funds in equilibrium, μ.

3. Draw a random triplet, (θ, α, u) and conjecture a market clearing price. Compute updated beliefs from the price using equations (2.7) and (2.9). Compute optimal demands using equation (2.2). Alter the price conjecture until the market clearing price has been found (i.e., equation (2.3) holds). Repeat this step $10,000$ times, collecting the realized profits for each investor type using equation (2.1). Compute the average difference in realized profits between informed and uninformed hedge funds.

4. Repeat Steps 2 and 3, altering the μ conjecture, until the average profit differential is equal to κ_θ (i.e., enforce hedge funds' indifference condition, equation (2.4a)). The resultant μ represents the optimal portion of informed hedge funds given the parameters chosen in Step 1.

5. Repeat Steps 1 through 4 over a grid of δ values ranging from zero to one. For each δ, collect the average profit differential between informed and uninformed mutual funds which is the numerical equivalent to the left-hand side of equation (2.4b). Select $\delta^\star$ from the grid of δ values. If the average profit differential from each δ is larger than κ_α then $\delta^\star = 1$. If the average profit differential from each δ is smaller than κ_α then $\delta^\star = 0$. If neither of the previous cases fit, $\delta^\star$ is selected as the δ which produces the average profit differential nearest to κ_α. The μ from Step 4 and $\delta^\star$ from this step constitute equilibrium, $(\mu^\star, \delta^\star)$.

Table C.1: Summary of Simulation Parameters

This table summarizes the simulation parameters used to provide the figures shown in this paper. The simulation process is summarized in Section 2.2 of the text and is more detailed in the Appendix.

Parameter	Symbol	Values
Mass of Hedge Funds	ν	1.00
Mass of Mutual Funds	ω	1.00
Precision of Payoff	τ_θ	$1.00 - 3.00$
Precision of Initial Signal	τ_η	$1.00 - 3.00$
Precision of Noisy Signal	τ_α	$1.00 - 3.00$
Precision of Noisy Supply	τ_u	$1.00 - 3.00$
Inventory Holding Cost Parameter	γ	$1.15 - 2.00$
Cost to Observe θ	κ_θ	$0.10 - 0.40$
Cost to Observe S_α	κ_α	$0.025 - 0.10$

APPENDIX D: DATA SOURCES AND SUMMARY STATISTICS

The EDGAR log files contain records of requests from masked IP addresses to view filings within the SEC's EDGAR database. An example of data from the log files is found in Panel A of Table D.1. While the fourth octet of every IP address has been masked with a random set of three letters, the IP address can nevertheless be linked to investments firms using the first three octets. This is due to the fact that organizations typically register blocks of IP addresses, with the most common block containing 256 IPs. All 256 IP addresses usually have the same first three octets. Thus, investment firms that have acquired blocks of IP addresses are not fully masked by the randomized fourth octet. Using historical IP address records from MaxMind, many investment firms and their activity on EDGAR are identifiable. As an example, Panel B of Table D.1 shows that the masked IP addresses in Panel A can be linked to actual investment firms.

A summary of the data derived from the EDGAR log files is provided in Table D.2. Also included in Table D.2 is a summary of the RavenPack data. The RavenPack data is used to measure ESS, which is calculated as the median event sentiment score from all articles on RavenPack over a given time period, with 50 indicating neutral sentiment.

Data on stock returns, market capitalization, and trading volume has been acquired from CRSP. Weller (2017) has been followed in order to calculate price jump ratios using returns from CRSP. The TAQ dataset has been used to measure trading volumes and to calculate buy-sell ratios. Earnings announcement dates and other stock-level information has been acquired from Compustat. Institutional ownership data has been taken from Thomson Reuters. The SEC's MIDAS data has been used to measure trading and quote-cancellation activity.

Table D.1: EDGAR Log File Example

This table shows an example of the data contained in the EDGAR log files. Panel A shows the data as it appears in raw form. Panel B shows an example of how the data appears after investment firms have been unmasked and after filings have been specified.

Panel A. Example of EDGAR Log Files

IP Address	Date	Time	CIK	Accession
154.61.131.ecg	20170531	09:47:33	051143	000104746917001061
205.173.24.fhf	20170531	11:07:28	274191	000002741917000008
216.230.48.igg	20170924	12:27:02	320193	000032019317000009
165.71.0.aah	20170924	16:12:55	831259	000083125917000016

Panel B. Example of Unmasked EDGAR Log Files

Investment Firm	Date	Time	Ticker	Filing
Dodge & Cox	20170531	09:47:33	IBM	10-K for 2016
Eaton Vance	20170531	11:07:28	TGT	10-K for 2016
American Century	20170924	12:27:02	AAPL	10-Q for Q2 2017
John Hancock	20170924	16:12:55	FCX	Earnings for Q2 2017

Table D.2: Summary of EDGAR Log Files, Stocks, and RavenPack

This table provides summary statistics for the activity of mutual funds and hedge funds in the EDGAR log files and for the RavenPack data. Between the 444 investment firms unmasked, a total of 17.9 million requests from EDGAR have been observed. Panel A summarized the unmasked institutions and their activity on EDGAR. Panel B summarizes institutional activity on EDGAR at the stock level. The variables in Panel B measure the average activity for stocks either monthly or in the 30 days leading up to earnings announcements (EA Date). The figures in parentheses in Panel B are averages for large stocks only. The RavenPack data consists of almost 19 million news items for over the sample of 3,300 stocks. Panel C shows a summary of the RavenPack data. For example, the average stock has 56 news items per month and a median event sentiment score (ESS) of 52.68.

Panel A. Institutions, Stocks, and EDGAR Activity

	Number	
Mutual Fund Companies	171	(80% of AUM; 66% of total funds)
Hedge Fund Companies	273	
Stocks	3,375	(58% large/small; 42% micro)
Months	66	(Jan. 2012 - June 2017)
EA Dates	48,782	
Total Requests	17,912,730	
Total Requests prior to EA Date	3,365,703	

Panel B. Stock-Level EDGAR Activity

	Mean	
	Mutual Funds	Difference
Monthly Requests per Stock	28 (124)	22 (93)
EA Date Requests per Stock	59 (130)	45 (91)
Monthly IPs per Stock	8 (38)	6 (31)
EA Date IPs per Stock	17 (41)	13 (32)

Panel C. RavenPack Summary

	Mean	Median
No. Items per Month	56.24	24.47
No. Items per EA Date	68.79	34.00
Median ESS Monthly	52.68	52.23
Median ESS EA Date	52.60	51.92
Std. Dev. of ESS Monthly	7.57	7.19
Std. Dev. of ESS EA Date	5.40	5.22

REFERENCES

Almazan, A., K. C. Brown, M. Carlson, and D. A. Chapman (2004). Why constrain your mutual fund manager? *Journal of Financial Economics 73*(2), 289–321.

Avery, C. and P. Zemsky (1998). Multidimensional uncertainty and herd behavior in financial markets. *American economic review*, 724–748.

Back, K., K. Crotty, and T. Li (2013). Estimating the order-flow component of security returns. *Available at SSRN*.

Bai, J., T. Philippon, and A. Savov (2016). Have financial markets become more informative? *Journal of Financial Economics 122*(3), 625–654.

Banerjee, S., J. Davis, and N. Gondhi (2018). When transparency improves, must prices reflect fundamentals better? *The Review of Financial Studies 31*(6), 2377–2414.

Banerjee, S. and B. Green (2015). Signal or noise? uncertainty and learning about whether other traders are informed. *Journal of Financial Economics 117*(2), 398–423.

Banerjee, S. and I. Kremer (2010). Disagreement and learning: Dynamic patterns of trade. *The Journal of Finance 65*(4), 1269–1302.

Barber, B. M., T. Odean, and N. Zhu (2008). Do retail trades move markets? *The Review of Financial Studies 22*(1), 151–186.

Beber, A. and M. Pagano (2013). Short-selling bans around the world: Evidence from the 2007–09 crisis. *The Journal of Finance 68*(1), 343–381.

Boehme, R. D., B. R. Danielsen, and S. M. Sorescu (2006). Short-sale constraints, differences of opinion, and overvaluation. *Journal of Financial and Quantitative Analysis 41*(2), 455–487.

Boehmer, E., C. M. Jones, and X. Zhang (2013). Shackling short sellers: The 2008 shorting ban. *The Review of Financial Studies 26*(6), 1363–1400.

Breon-Drish, B. (2015). On existence and uniqueness of equilibrium in a class of noisy rational expectations models. *The Review of Economic Studies 82*(3), 868–921.

Bris, A., W. N. Goetzmann, and N. Zhu (2007). Efficiency and the bear: Short sales and markets around the world. *The Journal of Finance 62*(3), 1029–1079.

Chakrabarty, B., B. Li, V. Nguyen, and R. A. Van Ness (2007). Trade classification algorithms for electronic communications network trades. *Journal of Banking & Finance 31*(12), 3806–3821.

Chang, E. C., J. W. Cheng, and Y. Yu (2007). Short-sales constraints and price discovery: Evidence from the hong kong market. *The Journal of Finance 62*(5), 2097–2121.

Chen, H., L. Cohen, U. Gurun, D. Lou, and C. Malloy (2018). Iq from ip: Simplifying search in portfolio choice. Technical report, National Bureau of Economic Research.

Crane, A. D., K. Crotty, and T. Umar (2018). Do hedge funds profit from public information?

Crouzet, N., I. Dew-Becker, and C. G. Nathanson (2018). On the effects of restricting short-term investment. Technical report, Kellogg School Discussion Paper 120, Kellogg School of Management.

Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. *The Journal of Finance 66*(5), 1461–1499.

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

Diamond, D. W. and R. E. Verrecchia (1987). Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics 18*(2), 277–311.

Drake, M. S., D. T. Roulstone, and J. R. Thornock (2012). Investor information demand: Evidence from google searches around earnings announcements. *Journal of Accounting research 50*(4), 1001–1040.

Duffie, D. and H. Zhu (2017). Size discovery. *The Review of Financial Studies 30*(4), 1095–1150.

Dyer, T. A. (2018). *Does Public Information Acquisiton Level the Playing Field or Widen the Gap? An Analysis of Local and Non-local Investors.* Ph. D. thesis, The University of North Carolina at Chapel Hill.

Ellis, K., R. Michaely, and M. O'Hara (2000). The accuracy of trade classification rules: Evidence from nasdaq. *Journal of Financial and Quantitative Analysis 35*(4), 529–551.

Engelberg, J. E., A. V. Reed, and M. C. Ringgenberg (2012). How are shorts informed?: Short sellers, news, and information processing. *Journal of Financial Economics 105*(2), 260–278.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics 33*(1), 3–56.

Fama, E. F. and K. R. French (2012). Size, value, and momentum in international stock returns. *Journal of financial economics 105*(3), 457–472.

Gao, F., F. Song, and J. Wang (2013). Rational expectations equilibrium with uncertain proportion of informed traders. *Journal of Financial Markets 16*(3), 387–413.

Gervais, S. (1997). Market microstructure with uncertain information precision: A new framework. Technical report, Tech. rep., Working Paper, University of Pennsylvania.

Gibbons, B., P. Iliev, and J. Kalodimos (2018). Analyst information acquisition via edgar.

Green, J., J. Hand, and M. Penn (2012). The bad news dissemination bias in the business press. Technical report, Working paper.

Grossman, S. J. and J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *The American economic review 70*(3), 393–408.

Harrison, J. M. and D. M. Kreps (1978). Speculative investor behavior in a stock market with heterogeneous expectations. *The Quarterly Journal of Economics 92*(2), 323–336.

Heflin, F., K. Subramanyam, and Y. Zhang (2003). Regulation fd and the financial information environment: Early evidence. *The Accounting Review 78*(1), 1–37.

Hellwig, M. F. (1980). On the aggregation of information in competitive markets. *Journal of economic theory 22*(3), 477–498.

Hirshleifer, D., A. Subrahmanyam, and S. Titman (2006). Feedback and the success of irrational investors. *Journal of Financial Economics 81*(2), 311–338.

Ho, K.-Y., Y. Shi, and Z. Zhang (2013). How does news sentiment impact asset volatility? evidence from long memory and regime-switching approaches. *The North American Journal of Economics and Finance 26*, 436–456.

Hvidkjaer, S. (2008). Small trades and the cross-section of stock returns. *The Review of Financial Studies 21*(3), 1123–1151.

Kolasinski, A. C., A. Reed, and J. R. Thornock (2013). Can short restrictions actually increase informed short selling? *Financial Management 42*(1), 155–181.

Lamont, O. and A. Frazzini (2007). The earnings announcement premium and trading volume. Technical report, National Bureau of Economic Research.

Lawrence, A., J. Ryans, E. Sun, and N. Laptev (2018). Earnings announcement promotions: A yahoo finance field experiment.

Lee, C. M. and M. J. Ready (1991). Inferring trade direction from intraday data. *The Journal of Finance 46*(2), 733–746.

Li, T. (2013). Insider trading with uncertain informed trading.

Mendel, B. and A. Shleifer (2012). Chasing noise. *Journal of Financial Economics 104*(2), 303–320.

Meulbroek, L. K. (1992). An empirical analysis of illegal insider trading. *The Journal of Finance 47*(5), 1661–1699.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of Finance 32*(4), 1151–1168.

Morse, D. (1981). Price and trading volume reaction surrounding earnings announcements: A closer examination. *Journal of Accounting Research*, 374–383.

Nezafat, M., M. Schroder, and Q. Wang (2017). Short-sale constraints, information acquisition, and asset prices. *Journal of Economic Theory 172*, 273–312.

O'hara, M., C. Yao, and M. Ye (2014). What's not there: Odd lots and market data. *The Journal of Finance 69*(5), 2199–2236.

Romer, D. (1992). Rational asset price movements without news. Technical report, National bureau of economic research.

Rostek, M. and M. Weretka (2012). Price inference in small markets. *Econometrica 80*(2), 687–711.

Saffi, P. A. and K. Sigurdsson (2010). Price efficiency and short selling. *The Review of Financial Studies 24*(3), 821–852.

Verrecchia, R. E. (1982). Information acquisition in a noisy rational expectations economy. *Econometrica: Journal of the Econometric Society*, 1415–1430.

Vives, X. (2011). Strategic supply function competition with private information. *Econometrica 79*(6), 1919–1966.

Wang, Y. and M. Yang (2016). Insider trading when there may not be an insider.

Weller, B. M. (2017). Does algorithmic trading reduce information acquisition? *The Review of Financial Studies 31*(6), 2184–2226.