Trading Volume and Dispersion of Signals [*]

Nikhil Vidhani [†]

December 15, 2020

[*] This work is based on my Ph.D. thesis, and I would like to thank Sankarshan Basu (committee chair), Srinivasan Rangan, Abhinav Anand, Xinjie Wang (discussant), Franziska Peter (discussant), Siyuan Yan (discussant), Yen-Cheng Chang (discussant), Prasanna Tantri (discussant), Murgie Krishnan, seminar participants at IIM Bangalore, and conference participants at International Conference on Derivatives and Capital Markets (ICDCM 2020), International Risk Management Conference (IRMC 2020), Southern Finance Association (SFA 2020), Conference on Asia-Pacific Financial Markets (CAFM 2020), World Finance and Banking Symposium (WFBS 2020), Securities and Financial Markets Conference (SFM 2020), and Emerging Market Finance Conference (EMF 2020) for helpful comments. I acknowledge the Mirae Asset Scholar grant for supporting this research.

[†] Indian Institute of Management, Bangalore. *email*: nikhil.vidhani16@iimb.ac.in. Ph: +91-797-555-7296

Trading Volume and Dispersion of Signals

Abstract

I propose a new measure of investor disagreement based on trading signals identified from the return-predicting anomaly literature. Disagreement significantly explains the next period trading volume over and above the factors identified in earlier literature. A move from 25th to 75th disagreement percentile predicts 27.4% higher turnover next period. The positive and significant relationship is robust to different specifications, alternative measures of turnover and disagreement, across size groups, and different time-periods. I also document that disagreement is higher for small, growth, and riskier stocks, which exhibit high fundamental uncertainty. Disagreement-volume relationship captures the information dissemination inefficiency of firms. Using properties of firms' 10-K reports, I find that investors rely more on return anomalies when the flow of information from firms is convoluted, opaque, disrupted, or delayed.

Keywords: Disagreement, Return Anomalies, Trading Volume
JEL Classification: G11, G12, G14

Trading Volume and Dispersion of Signals

1 Introduction

Disagreement among market participants can cause trading volume. If two traders hold opposing views regarding the future value of an asset relative to its current price, then the direction of their trades will differ, leading to volume. Disagreement can arise if investors use different signals to predict the future value of an asset. It could also arise when investors interpret the same signal differently when predicting future returns. In this paper, I propose a new method to measure investor disagreement that is derived from prior research on fundamental and price-related signals (anomalies) that have been shown to predict returns.

The intuition for my measure of disagreement is as follows. I assume that different investors believe that different anomalies predict future returns. An investor will initiate a buy (sell) trade if the anomaly predicts higher (lower) returns in the future. Because anomalies differ in the sign of their relationship with future returns, dispersion in these signs captures disagreement. To the best of my knowledge, this is the first study that has attempted to study the effect of dispersion in trading signals/anomalies on trading volume.

To measure disagreement, I consider 36 return anomalies from the literature (Linnainmaa and Roberts (2018)). Each anomaly is used to cross-sectionally sort the stocks based on their predicted future performance. Stocks are divided into three categories such that top 30% stocks get a buy signal, bottom 30% stocks get a sell signals and the rest are in hold category. This process is repeated for all 36 return anomalies giving a series of 36 buy/sell/hold signals for each stock in each period. For every firm-month, I measure disagreement as the standard deviation of these signals.

An important advantage of my measure of disagreement is that it is unlikely to be afflicted by data snooping bias. To reduce dependence on the particular choice of return anomalies in measuring disagreement, I only consider broad (buy/sell) trading signals generated by these anomalies and construct a rudimentary measure (standard deviation of signals) of disagreement which is unlikely to be affected by inclusion or exclusion of a few anomalies. To avoid cherry picking of anomalies, I choose a reasonable number of anomalies in their entirety from previously published studies. I consider all traded firms over the entire duration for which data is available to reduce sample selection bias.

Using a wide range of anomalies constructed from the firm's fundamentals and market price for a large panel of US stocks, I find that my disagreement measure is strongly related to subsequent monthly trading volume. This effect obtains after controlling for several factors identified in prior literature on determinants of the trading volume. A one standard deviation increase in disagreement causes roughly 15.5% additional trading volume in the next month. The relationship is robust to different regression specifications, different measures of turnover and disagreement, across time-periods, and size deciles.

I also document that disagreement is higher for small, growth, and riskier stocks, which exhibit high fundamental uncertainty. Because disagreement in signals is likely to increase uncertainty, my finding provides a channel through which small stocks and growth stocks are perceived as being more risky. I also find that firms belonging to pharmaceuticals, oil, computers and IT industries have the highest disagreement. These industries are characterized by higher level of uncertainty in the form of future cash flows, real options, and regulatory risks.

The sensitivity of trading volume to disagreement captures the firm's information dissemination inefficiency. Investor's reliance on return anomalies will increase if the flow of value relevant information from firms is convoluted, opaque, disrupted, or delayed. Using variables derived by parsing firm's 10-K reports, I find that disagreement coefficient is higher when a firm's disclosure is lengthier, has more complex words and when the disclosure becomes stale. Low analyst following, younger firms, and firms with low institutional ownership also exhibit a larger disagreement coefficient.

My work contributes to the literature on empirical studies of cross-sectional determinants of trading volume. Papers on volume prediction are relatively rare as compared to studies of stock returns[1]. Chordia, Huh, and Subrahmanyam (2006) is perhaps the first comprehensive empirical study of the cross-sectional determinants of trading activity. They consider several variables capturing liquidity trading, informed trading, and trading due to belief divergence in a single predictive regression framework. More recently, Jacobs and Hillert (2015), building on the empirical model of Chordia et al. (2006), add several other variables related to stock's visibility like S&P membership and advertising expense. They find that firm names appearing in the upper half of the alphabet experience significantly more trading than those in the lower half. Lo and Wang (2010) perform an exploratory investigation of trading volume and suggest other variables like alpha and residual standard deviation from CAPM regression and return auto-correlation to affect trading volume.

Prior research has generally used dispersion in analyst earnings forecasts to measure market-level disagreement. My paper makes a methodological contribution by proposing a new measure of disagreement that complements analyst forecast-based measure of disagreement[2]. Although commonly used, the analyst forecast based measure suffers from several weaknesses. First, these analysts are sell-side agents employed by brokerage firms who earn commissions on the number of shares sold following the issue of their forecasts. There is also evidence that analyst forecasts are positively biased (Beyer and Guttman (2011)) and exhibit herding behavior. Secondly, this measure is not a market-based measure in that it only includes views of a small class of agents (Atiase and Bamber (1994)). Moreover, analyst following is highly skewed towards larger firms, and forecast dispersion with small analyst following is imprecise. Finally, the most common forecast issued is for the earnings, and it may differ from beliefs about the asset's future value. Thus, forecast dispersion may not accurately capture the disagreement relating asset's future price.

[1] Lo and Wang (2010) begin their article by highlighting the remarkable amount of studies exploring the field of asset prices while there being no parallel *asset quantities* literature.

[2] Kandel and Pearson (1995) compute a measure of disagreement by isolating flips and diveregences of the same analyst's successive forecast surrounding an earnings announcement and attribute their frequency of occurrence to the extent of differential interpretation. More recently, Fischer, Kim, and Zhou (2020) regress analysts' forecast on prior forecasts of other analysts and use the extent of deviation of the slope coefficient from one as a measure of disagreement.

2 Review of Literature

In standard rational expectations (RE) asset pricing models with common priors and homogenous interpretation of information (like CAPM), there is no role for trading volume once the hedging and insurance needs of traders are satisfied (Varian (1989)). In a fully revealing RE equilibrium, i.e., where prices reflect all available information, investors without private information are not willing to trade since they anticipate the superior nature of information possessed by other investors (Milgrom and Stokey (1982)). Asymmetric information dries up liquidity, and the market fails as it does in Akerlof (1978).

To overcome the impossibility of trade in a fully revealing RE setting, researchers assume some exogenous source of randomness either through noise trading (Kyle (1985)), endowment shocks (Diamond and Verrecchia (1981)) or liquidity shocks. Some form of exogenous noise must be assumed, even with differential information, to conceal trading intentions of insiders holding superior information. This assumption makes equilibrium only partially revealing as opposed to fully revealing[3]. For instance, in Kyle (1985), privately informed investors trade because of the camouflage provided by noise traders.

However, the assumption of exogenous noise, which creates room for trading, is also an inherent weakness of partially revealing RE equilibrium models. Any interesting patterns in trade emerge as a direct consequence of the assumed structure of noise, limiting the role of differential information in predicting trading patterns (Kruger (2020)). Disagreement models can overcome the shortcomings of RE models. They assume some form of investor belief heterogeneity, which creates disagreement regarding asset payoffs. Agents agree to disagree in equilibrium and hence interpret information differently, giving rise to trade as investors revise their beliefs (Banerjee and Kremer (2010)). Investors can disagree when one group of investors (specialists) acquire value relevant information earlier than the other investors causing them to disagree and hence trade more. This effect is evident with trading volume spiking at information events like earnings announcements. Disagreement can also arise when investors hold heterogeneous priors or interpret public information such as earnings announcements differently. Hong and Stein (2007) give a brief overview of this literature.

Several theoretical papers model volume as a function of disagreement. In these papers, disagreement comes in many flavors. Investors can disagree because of differential access to information, different tastes and endowments, and divergent beliefs. Varian (1989) shows that mere differences in information, tastes, or endowments, without any belief heterogeneity, cannot give rise to trading. Investors must have differing beliefs for trading to occur. Investor's beliefs can differ if investors hold different opinions (priors), or they interpret the common information differently. Investor's posterior belief is a function of her prior belief and likelihood function. The latter is akin to how information is interpreted. Next, I summarize the leading papers that link investor disagreement to trading volume.

Karpoff (1986) presents a model of trade that assumes random pairing where each owner

[3] In a partially revealing RE equilibrium, agents' private information is only partly reflected in the price. When there are two sources of randomness, i.e., asset's final payoff, and exogenous noise, then prices can't fully reveal the information pertinent to asset's payoff (Diamond and Verrecchia (1981)). Without exogenous noise, asset's payoff related information is fully reflected in the price, and in that sense, prices are fully revealing.

(seller) and nonowner (buyer) are blindly paired. Trade happens when the buyer's demand price exceeds that of the paired seller due to idiosyncratic demand price revisions. Volume increases if agents, in addition to their idiosyncratic revisions, also have independent differential interpretations of information. The differential interpretation further adds to the variance of demand price revisions resulting in increased trading.

Varian (1989) finds that new information causes trade but only through its differential interpretation. When agents interpret information differently, they also interpret the adjusted price (which contains all new information) differently, and hence trading ensues. In a heterogeneous agent model with different priors and interpretation of information signals, Varian finds that the equilibrium trade for an agent is proportional to the extent by which her opinions and interpretations differ from the average opinion and average interpretations, respectively.

Harris and Raviv (1993) consider two groups of agents disagreeing on how they interpret commonly available information signals like announcements of dividends, earnings, stock issues, and macroeconomic events using different likelihood functions. Both groups agree on the "direction" of the common information, i.e., whether it is good or bad but differ in the impact the information has on asset's future payoffs. A positive signal makes the responsive group more optimistic, and this group owns all the shares. In contrast, a negative signal makes the unresponsive group relatively optimistic, resulting in it owning all the shares. Such a mechanism emerges as a result of risk-neutrality and short-sale constraints, where at each period, only one group ends up owning all the shares.

Kandel and Pearson (1995) (hereafter KP) model differential interpretation of public information signals as arising from investors differing in the mean and precision of their likelihood function, i.e., the model to interpret the public signal. KP model the information signal as consisting of two parts: asset payoff and an additive likelihood function, which is interpreted differently across agents. Agents have different means and precisions of the likelihood term. KP find that public announcements (like EA) can have an abnormal trading reaction even when there is no price reaction and attribute this to different likelihoods used by agents to gauge the future asset payoff. In a two-period heterogeneous agent model, KP find that, in equilibrium, the quantity traded is driven by two factors. The first factor is independent of the price change and proportional to the extent by which agents disagree about the mean of the public signal. The second factor is proportional to price change and differences in the precision of the prior beliefs and the public signal. Using analysts' revision of earnings forecasts, they find significant proportions of revisions to be flips or divergences, which can only appear if analysts interpret information differently. Overall, KP provide substantial empirical evidence that investors interpret public signals (EA) differentially evident from the observed patterns in stock trading volumes.

Banerjee and Kremer (2010) consider a difference in opinion (DO) model of trading volume where investors disagree about the interpretation of public information. In a DO model, agents agree to disagree in equilibrium and hence interpret information differently, giving rise to trade as investors revise their beliefs. Disagreement can lead to two types of trading volume response, 1) convergence trade when investors had a prior disagreement but agreed on the current information, and 2) idiosyncratic trade when investors agree on prior information but disagree on the current information. A period of high disagreement leads to a higher volume, which falls slowly and hence exhibits clustering.

My paper is most closely related to KP. I study trading signals generated from widely used return predictors, including fundamental ratios, accounting anomalies, and return momentum acting as different models of asset valuation used by investors. Consistent with the model of KP, the public signal can be thought of as the current market price, and the different likelihood functions are the different return predicting signals investors use. For instance, one trader may base her trades using the current book value to calculate the market to book ratio while another may use past price to base her trade on return momentum.

Turning to empirical work, Chordia et al. (2006) perform a comprehensive empirical analysis of the cross-sectional determinants of trading activity. They consider several variables believed to impact trading volume in a single predictive regression framework. Trading volume can be associated with several determinants like liquidity trading, informed trading, and trading due to dispersion in beliefs. Liquidity trading arises due to portfolio rebalancing needs and is positively related to the magnitude of past returns. Trading is supposedly higher in more visible stocks where visibility can be proxied by firm size, firm age, price level, and market to book ratio, all of which predict higher trading volume. If the estimation of a firm's fundamentals is uncertain, then this should lead to more learning-based trade as investors correct their estimation errors. Fundamental uncertainty is proxied by absolute earnings surprise, earnings volatility, and CAPM beta. Lastly, analyst forecast dispersion proxies for belief dispersion and the number of analysts following a firm represents the mass of informed agents, both of which positively predict trading volume.

Carlin, Longstaff, and Matoba (2014) examine how disagreement affects asset returns, volatility, and trading volume using data from the MBS market. Using a VAR framework of disagreement, volatility, and volume, authors find that increased disagreement leads to increased volatility and volume. Volume also increases with volatility, but only when disagreement is high. Investors learn from their trades - higher disagreement leads to higher volume, and subsequently, disagreement falls, resulting in a mean-reverting time series for disagreement.

Jacobs and Hillert (2015) find that firm names appearing in the upper half of the alphabet experience significantly more trading than those in the lower half. They build on the empirical model of Chordia et al. (2006) and propose several other variables that might influence trading like advertising expense, 52-week high/low events, idiosyncratic volatility, market model alpha, media coverage, S&P 500, and DJIA membership.

In addition to work that has predicted monthly trading volume, a few papers in the accounting literature have linked disagreement with the abnormal trading volume observed during an earnings announcement (EA). Beaver (1968) emphasizes that a price reaction conveys average market-wide changes in expectations, while a volume reaction indicates changes in individual expectations. The change in individual expectations might arise due to differential precision of prior information or increased disagreement following an EA (see Bamber, Barron, and Stevens (2010) for a review). Bamber, Barron, and Stober (1997) find incremental explanatory power using three measures of disagreement in explaining abnormal trading volume around EA: dispersion in prior beliefs, differential interpretation of information signal, and consensus effect arising from post information belief dispersion. All three measures are constructed using analyst's forecasts, and in particular, the second measure for differential interpretation is the dispersion in forecast revisions by the same analyst.

3 Research Design

I study the relation between disagreement and trading volume in a broader setting than just an earnings announcement (EA). Financial markets are continuously bombarded with new information, be it public announcements, dividend announcements, new issues and buybacks, mergers and acquisitions, macroeconomic news, and geopolitical developments. Studying the relationship surrounding an EA is restrictive, given the rich set of the information environment in which firms operate. A comprehensive and parsimonious model of trading activity should explain the determinants of volume at all times.

Limited attention and costly acquisition of information make it difficult for investors to study the entirety of fundamental anomalies. As such investors are likely to anchor their trading decisions on a small set of anomalies. Dispersion in information signals would trigger trading activity since investors using different anomalies as their model of asset valuation would come up with differing estimates. Some investors would interpret the asset to be undervalued while others would find it to be overvalued.

The factor search literature is vast, with over 300 factors identified in Harvey, Liu, and Zhu (2016). To the best of my knowledge, no study has combined both the model of trading volume with return anomaly research and attempted to study the impact of these signals on trading volume. To the extent that investors differ in their model to predict future stock prices by way of focusing on a particular signal, then dispersion across the signals as measured by their standard deviation should relate to trading volume. I hypothesize that the relationship is positive and a higher dispersion causes increased trading volume in the future. We expect the coefficient, β_1 to be positive in equation 3.1:

$$Volume_{i,t+1} = \beta_0 + \beta_1 \cdot Signal\_Deviation_{i,t} + \gamma \cdot Controls_{i,t} + u_{i,t+1} \tag{3.1}$$

The disagreement-volume relationship hinges on the trading reaction following differential interpretation of public information. Smaller firms are less visible, meagerly followed, issue infrequent disclosures, and have an opaque information environment. The amount of public information in the form of management disclosures, analyst recommendations, and media coverage is limited for small firms. Many times the only disclosures are the ones mandated by regulators, which comprise quarterly earnings and annual balance sheet announcements. On the contrary, large firms routinely issue management forecasts, voluntary disclosures, and major sales announcements in addition to earnings and balance sheet disclosures. This leaves stock price and accounting information as the only source of information for small firms. Since not much is known about smaller firms, investors are more likely to estimate future stock prices using widespread anomalies and fundamental ratios. Hence disagreement-volume relationship should be stronger for such firms, and we should expect the coefficient on β_1 to be comparatively higher for small firms.

Investment using accounting fundamentals and related ratios is popular among value investors.

Several books on value investing give prominence to price to book ratio as an indicator of value firms[4]. Value investment attempts to estimate the intrinsic value of a stock using information from the balance sheet and profit and loss statements. The intrinsic value is then compared to the current stock price, and a buy (sell) trade is initiated when the intrinsic value is smaller than the current price. Price to book ratio is also a proxy for visibility where firm with high valuations, i.e., growth firms, are often talked about in media and followed more by analysts[5]. Thus the information environment of value firms is limited and hence the use of anomalies would be higher. We should expect the disagreement-volume relation to be stronger for the high book to market firms since these firms fall into the category of value firms and are more likely to be evaluated using return anomalies originating from accounting fundamentals.

3.1 Measuring Volume

Trading volume has seen explosive growth in the last few decades. The monthly cross-sectional average dollar volume (shares traded) has skyrocketed from $3 Million (70 K) to $1.2 Billion (23 Mn) in the period between 1962 to 2019. In the same period, turnover (defined as a ratio of dollar volume to market capitalization) has gone from 19% to 225%. Thus on average, a company's entire shares changed hands more than twice in December 2019. Figure 1 shows a time series of three measures of volume, averaged across all firms every month. The logarithmized version of the series shows a clear time trend in all three measures.

[Insert Figure 1 here.]

There are several ways to measure volume like share volume, dollar volume, or turnover. Section 2 of Lo and Wang (2010) gives a brief exposition of different approaches. Of the three measures, dollar volume depends on the size of the firm and share volume depends on the share price making turnover the only measure independent of share price and firm size. Moreover, dollar volume and share volume vary a lot across the sample. The ratio of 75th percentile to 25th percentile observation of dollar volume, share volume, and turnover is 66.5, 38.9, and 6.6 respectively across all NYSE stocks over the 1962-2019 period. Lo and Wang (2010) also find that if the two fund separation holds, then turnover is the most natural measure for studying the relation between trading volume and equilibrium market models like CAPM.

However, turnover is highly non-stationary. As can be seen in Figure 1, log turnover has a clear linear time trend, which means turnover has grown exponentially over the years. To account for the time trend, I include year dummies in all regressions. Chordia et al. (2006) adjust turnover and other non-stationary time-series using an adjustment procedure proposed in Gallant, Rossi, and Tauchen (1992) (hereafter GRT). In their study, GRT remove linear and quadratic time trends, monthly calendar dummies, day of the week effects, and trading gaps (due to holidays) from both the mean and variance of the trading volume. The adjustment procedure involves regressing time series of firm wise turnover on the above effects and subsequently regressing squared residuals from the last regression again on the same effects. This removes any trend or calendar effects

[4] Lower P/B ratio (or high book to market) signifies that a firm's market price doesn't accurately reflect it's book value and hence is undervalued. Graham (2006) is a popular investing book following value investing tenets.

[5] Number of analyst following and book to market ratio has a rank correlation of -0.14.

from both the mean and variance of the turnover time series. In the regression specifications, I do not consider the above adjustments for the following reasons[6]:

- GRT's motivation for using quadratic detrending is mostly based on visual examination of volume series, which exhibited a downward trend in post-depression years followed by an upward trend since the World War II. This gives a nonlinear visual appearance to the volume series. The sample period for my study is from 1962 to 2019, and there is no apparent second-order non-linearity in turnover to consider quadratic detrending.

- In a bid to achieve stationarity, GRT adjustment also removes the effects mentioned above from the variance of turnover, which is an overkill and reduces the analysis to a statistical exercise. After two-tier detrending, it is impossible to assign any economic meaning to the "GRT adjusted" turnover.

- Removing calendar effects and seasonalities reduces the efficacy of the model in explaining these interesting patterns. For instance, increased trading in January is believed to originate from tax-induced trading. Removing monthly seasonality in a model studying tax effects on trading activity will destroy any explanatory power.

3.2 Anomalies

Disagreement arises since investors interpret the available information present in stock price and fundamentals differentially by employing different return anomalies. I use 31 anomalies studied in a recent paper by Linnainmaa and Roberts (2018) and complement that with five momentum anomalies from McLean and Pontiff (2016). Construction of these anomalies only requires data on stock price and annual reports.

Table 1 contains the list of anomalies used in constructing the deviation measure. The 36 anomalies span seven categories: profitability (6), earnings quality (3), valuation (5), momentum (5), investment (9), financing (6), and distress (2). Column 3 of the table depicts the predicted return association in the future. For instance, with respect to Gross Profitability anomaly, higher gross profit predicts higher returns in the future. Within each category, anomalies generally have a similar sign of predicted relationship. All six profitability anomalies are positively associated with future returns, while 8 out of 9 investment anomalies predict a negative future return. All anomalies predicting a negative relationship as per their original study are scaled by minus one to make the entire set of 36 anomalies to be positively related to future returns.

[Insert Table 1 here.]

Figure 2 gives a correlation heat map for the 36 anomaly signals. The positive correlation is represented by blue circles while negative correlation is shown by red circles. The size of the circle is proportional to the magnitude of the correlation coefficient. A heat map helps visualize the correlation between signals better than a correlation matrix. All categories of signals, except momentum (15-19) are positively correlated within their respective groups. Profitability (1-6)

[6] For completeness I include GRT adjusted turnover in one of the robustness checks (Table 5).

signals have a strong negative correlation with investment (20-28) signals, a slightly negative association with momentum signals, and a positive correlation with distress (35-36) signals. Earnings quality (7-9) signals are positively correlated with investment signals. Valuation (10-14) and investment signals negatively correlate with distress signals.

[Insert Figure 2 here.]

3.3 Measuring Disagreement

I propose a disagreement measure which tries to capture the differential interpretation of public information like stock price and accounting fundamentals, where each investor employs a different return anomaly to estimate stock's future performance. As a first step, each of the 36 return anomalies is used to cross-sectionally rank stocks based on their expected future performance. For instance, a stock with a low market to book value is expected to earn higher than average returns in the future; hence for each period, stocks will be reverse sorted on the market to book value. Stocks are divided into three categories such that the top 30% stocks get a buy (+1) signals, bottom 30% stocks get a sell signal (-1), and the rest are in hold category (0). This process is repeated for all 36 return anomalies giving a series of 36 buy/sell/hold signals for each stock in each period. For every firm-month, I measure disagreement as the standard deviation of these signals.

There are several ways to capture dispersion other than the standard deviation. Absolute deviation uses absolute differences from the mean value instead of squared differences in standard deviation. One limitation of using a broad buy/sell categorization is that other measures of deviation like interquartile range (IQR) and absolute median deviation are not properly defined. To safeguard against the choice of deviation measure and the splitting criterion for assigning buy/sell signals, I also check for robustness using absolute deviation as the measure of disagreement as well as using an 80/20 split where top 20% stocks are assigned a buy signal, and bottom 20% get a sell signal.

Let $A_{f,t,s}$ denote the s^{th} anomaly for firm f at time t. $L_{t,s}$ and $H_{t,s}$ are the low and high cross-sectional cut-offs for determining whether the stock gets a sell signal or a buy signal. For the 70/30 split, $L_{t,s}$ ($H_{t,s}$) is the 30$^{\text{th}}$ (70$^{\text{th}}$) percentile of anomaly s across all firms at time t. The corresponding trading signal $T_{f,t,s}$ is given by,

$$T_{f,t,s} = \begin{cases} -1, & A_{f,t,s} < L_{t,s} \\ 0, & L_{t,s} \leq A_{f,t,s} \leq H_{t,s} \\ 1, & A_{f,t,s} > H_{t,s} \end{cases} \tag{3.2}$$

The mean trading signal, $\overline{T_{f,t}} = \frac{\sum_s T_{f,t,s}}{|S|}$ and the deviation among the signals $\sigma^2_{f,t} = \frac{\sum_s (T_{f,t,s} - \overline{T_{f,t}})^2}{|T_{f,t}|-1}$, where $|T_{f,t}|$ is the number of anomaly signals for firm f at time t and $|S|$ is the number of anomalies. $\sigma_{f,t}$ is the measure of disagreement and the primary explanatory variable of interest in this study.

In the rest of the paper, I work with 70/30 cross-sectional stock splits[7]. For each period and firm pair, disagreement is the standard deviation of the 36 signals. Figure 3 gives a time-series of

[7] I use 80/20 stock split as a robustness check.

average disagreement[8] and its confidence interval (of two standard errors) over time. For each month, averaging is done across all firms. The average signal dispersion over the entire sample period is 0.753, and the standard deviation is 0.103. Unlike turnover, there is neither a trend nor a spike around 2008 in the time-series of disagreement. The stationary appearance of disagreement gives us confidence that the volume-disagreement relationship isn't affected by a common trend[9].

[Insert Figure 3 here.]

4 Data and Sample

The first step to construct the disagreement measure is to construct the anomalies. I employ 31 anomalies from Linnainmaa and Roberts (2018) and five momentum anomalies from McLean and Pontiff (2016). From the universe of NYSE, AMEX and NASDAQ stocks, I obtain monthly stock returns, shares outstanding and volume data from the Centre for Research in Security Prices (CRSP) database. Stock's annual fundamentals are obtained from COMPUSTAT's annual files, and earnings data is acquired from quarterly files. Both CRSP and COMPUSTAT data are accessed from Wharton Research Data Services (WRDS) account. Analyst forecast summary data is obtained from the Institutional Brokers' Estimate System (IBES) database via the Thomson Reuters account. Matching across the datasets is done using 8-character CUSIP identifier. Annual accounting data is available from January 1962, quarterly data from July 1971, monthly stock return from December 1925 and analyst forecast data from January 1976. The above availability dictates the sample to start from 1976. The sample ended in 2019. For quarterly earnings data, it is required that an announcement is made within 180 days after the fiscal quarter-end. For calculating analyst forecast dispersion, I require at least two annual forecasts. If forecast dispersion could not be calculated due to the unavailability of data, then previous estimates are carried forward. All these requirements leave us with roughly 670 thousand firm-month observations. Finally, all variables[10] are winsorized at 0.5% on either extreme before estimating a regression.

For evaluating how a firm's information environment impacts the disagreement volume relationship, I use the parsed EDGAR filings provided by Bill McDonald at the *Software Repository for Accounting and Finance* section of the University of Notre Dame. I also use the summary file created by Loughran and McDonald using the parsed EDGAR files. The summary file is used to get the number of total words, number of unique words, and the size of 10-K filing. To measure firm complexity, I find the number of complex words by searching for the occurrence of 374 complex words identified by Loughran and McDonald (2020) in each of the parsed 10-K EDGAR files[11]. The

[8]Cross-sectional average of the firm by firm disagreement is $\left(\frac{\sum_f \sigma_{f,t}}{|F_t|}\right)$ where $|F_t|$ is the number of firms at time t.

[9]The unit-root (or the AR(1) coefficient) is 0.90, which is statistically different from 1 thus making the time-series of average disagreement stationary. Banerjee and Kremer (2010) and Carlin et al. (2014) argue that investors learn from disagreement, and periods of high disagreement are followed by periods of low disagreement. Thus disagreement should form a mean-reverting time-series. Figure 3 shows that disagreement varies on a time scale of business cycles.

[10]Variable representing ranks, dummies, terciles, or deciles are not winsorized. Further, the number of analysts following a firm is also not winsorized since most firms are followed just 2 or 3 analysts. Following similar logic, the number of business segments from Compustat is also not winsorized.

[11]The parsed EDGAR files are present at https://sraf.nd.edu/, and the summary file is located at https://

number of firm segments is from Compustat's historical segments. These variables are merged with CRSP and Compustat data using CIK firm identifier and fiscal yearends. Lastly, I get the institutional ownership data from the 13-F database, short interest from Compustat, and firm segments from the historical segments database.

4.1 Data Snooping

Data snooping is a significant concern in empirical asset pricing research. I try to be as agnostic and comprehensive as I can be regarding my choice of factors. However, certain limitations on data availability restrict the choice of factors. The most comprehensive recent list of fundamental factors is McLean and Pontiff (2016), which employs 97 past determinants of returns. Many of these factors use IBES and other proprietary data, which prohibits their inclusion in my analysis. Computational and time constraints at my end also restrict the universe of return anomalies I can consider while calculating the disagreement measure. Finally, I settle with all 31 factors studied by Linnainmaa and Roberts (2018) and the five momentum anomalies studied in McLean and Pontiff (2016).

The construction of disagreement measures can still be infected with data-snooping concerns. Lo and MacKinlay (1990) argue that portfolios formed using previously studied return characteristics are more prone to data snooping. I offer several justifications as to why the disagreement measure may not be susceptible to data mining:

- I am not aware of any previous work using the dispersion among return predicting anomalies to predict trading volume, and hence the arguments of Lo and MacKinlay (1990) don't apply. Moreover, my work is focused on explaining volume, whereas data snooping as such is attributed widely to return anomalies literature where factor searching is intense (Harvey et al. (2016)).

- I am using only a rudimentary dispersion measure – standard deviation – and as such, this dispersion measure is unlikely to be significantly affected by the inclusion/exclusion of a few factors.[12]

- In constructing the disagreement measure the exact degree by which a particular signal predicts return is not used. For instance, if prior research has documented that factor A predicts 1% higher returns while factor B predicts only 0.1% higher returns, a stock falling in the "buy" categories for both the factors will get a +1 signal w.r.t. each factor. In other words, looking just at the signal, it is impossible to tell how much of a future return is predicted by the underlying factor. A more involved approach exploiting the degree of return prediction would assign a higher number to signal A than to signal B. This prevents basing the disagreement measure on previously observed returns and addresses the concern raised by Lo and MacKinlay (1990).

sraf.nd.edu/textual-analysis/resources/#LM_10X_Summaries. The list of complex words is also available at https://sraf.nd.edu/data/complexity/.

[12]In Internet Appendix, I show that average disagreement depends only on the aggregate sum of individual signal correlations. Hence, one particular signals can't significantly affect the average level of disagreement.

- I am choosing a reasonable number of factors in their entirety from a previously published study (Linnainmaa and Roberts (2018)) and complementing it with the full set of momentum-based factors from McLean and Pontiff (2016). Hence, there is no cherry-picking of factors that would fit the data.

- The study covers all firms over the entire duration for which stock's price and fundamental data is available over the CRSP and COMPUSTAT databases, thus nulling the p-hacking argument due to sample selection (Harvey et al. (2016)). The large sample at hand also avoids any false/lucky discovery due to small sample usage.

- I use the multi testing threshold of $|t| \geq 2.78$ for ascertaining significance instead of the usual 1.96 (Harvey et al. (2016)). Moreover, all standard errors are double clustered by year-month and firm,[13] and are more conservative.

5 Nature of Disagreement

The ingredients of the disagreement measure are the return anomaly signals based on accounting data and stock price. Disagreement arises when signals diverge about the predicted return for the stock. In this section, I explore the nature of disagreement and how it is characterized and evaluate questions like: Does disagreement vary by the number of buy and sell trading signals? Do all anomalies contribute equally to the disagreement measure? Which industries cluster more when disagreement is low or high? And lastly, how do firm characteristics like size and return vary with disagreement?

Figure 4 presents average disagreement as a function of the difference in the number of buy and sell signals. Disagreement is highest when we have an equal number of buy and sell signals. This is expected since a buy signal is given a value of +1 while a sell signal is given -1. If we think of signals as a sequence of ± 1 Bernoulli trials, then its variance[14] is maximized when the likelihood of buy signals is the same as that of sell signals.

[Insert Figure 4 here.]

Disagreement and Anomalies

The correlation heat map (Figure 2), although helpful in the visual distinction of groups of anomaly signals, doesn't provide any information on how different anomalies affect disagreement. Figure

[13] Clustering by time is equivalent to Fama-MacBeth (FM) procedure where time series mean and standard error of per period cross-sectional regression estimates are reported as coefficient estimates and its standard error. In time clustering entire cross-section of a period is taken as one observation, and it preserves the cross-correlation structure of asset returns. Similarly, firm clustering takes entire time-series of a firm as one observation and therefore takes into account the auto-correlation structure of persistent variables like trading volume. This is equivalent to Newey-West (NW) adjustment. In general, double clustering performs better than FM regression combined with NW adjustment as long as there is a sufficient number of clusters (~ 50) in each dimension (see Petersen (2009) for a comparison of different approaches to adjust errors).

[14] With 70/30 splits, the average number of buy and sell signals across firms is fixed and equal to $0.3 * 36 = 10.8$. However, within a time-period number of buy and sell signals will vary across firms. Since the signals are not independent, the number of buy signals is not a binomial random variable. Hence, finding the analytical variance of the number of buy signals is much more involved and depends on the correlation structure of signals.

5 plots the semi-annual average disagreement constructed using all but one group of anomaly signals.[15] If excluding a particular group reduces the disagreement substantially, then that group contributes significantly to the disagreement measure. Removing profitability or earnings quality signals doesn't have any significant change in disagreement. The exclusion of any other groups of anomalies dampens the disagreement with the momentum group affecting it the most. Surprisingly, the exclusion of just two distress anomalies reduces average disagreement from 0.743 to 0.714. This is because the two anomalies, O_Score and Z_Score are mostly negatively correlated with other anomalies and hence contribute significantly to the disagreement measure. This is evident from the last two rows of the heat map in Figure 2 where the two distress anomaly mostly have a negative correlation with other anomalies.

[Insert Figure 5 here.]

Disagreement and Firm Characteristics

Next, I look at how firm characteristics vary with disagreement deciles. Many of the firm characteristics are also the controls in regression analysis in section 6. I look at stock returns, size, earnings surprise, forecast dispersion, and other variables across ten disagreement deciles computed at each month. To facilitate comparison across deciles and characteristics, I use cross-sectional ranks of firm characteristics[16]. Figure 6 presents the variation in ranks of firm characteristics over ten disagreement deciles. Subplot 1 has turnover measures, subplot 2 has disagreement measures, subplots 3-5 has firm characteristics, and subplots 6-8 has firm's 10-K report characteristics.

[Insert Figure 6 here.]

Different measures of turnover increase monotonically with disagreement. Log turnover rank rises from 0.45 in first decile to 0.63 in tenth decile. Changes in GRT adjusted turnover ranks closely follow the changes in log tunover rank while changes in value-weighted and detrended turnover are much more gentle. All disagreement measures move sharply with disagreement. Thus, the particular choice of disagreement should then be irrelevant in explaining trading volume. This is expected since different measures only differ in their mathematical formulae, not in the trading signals generated by return anomalies. Since disagreement captures standard deviation of return predicting signals, it is expected to strongly correlate with standard deviation of daily returns as well (RET_VOL)

Excess return is stable for the first six deciles and then starts falling. It's rank changes from 0.5 to 0.35 as we move from sixth to tenth decile. CAPM alpha and beta are much more stable across the ten deciles. Interestingly, firm's dividend payout falls dramatically between first and fifth decile reaching a 0 rank and then stays there for the next five deciles. A high disagreement firm pays no dividend.

[15] For constructing disagreement from the full set of 36 signals, I require at least 10 anomaly signals to be available. However, I relax this requirement for disagreement constructed in Figure 5 as the number of anomalies are reduced.

[16] Stocks are ranked each month by each characteristic and then scaled to fall between 0 and 1. Thus they are equivalent to percentiles. Details of variable construction appear in Appendix A.2.

Fundamental uncertainty captured by earnings surprise and earnings volatility also associates positively with disagreement. Moving from first to last disagreement decile, earnings surprise, and volatility rank change from 35% to 80%. The signals making up the disagreement measure are mostly constructed using fundamental information from the balance sheet and P/L statements. A more uncertain income stream and business forecast cast doubt on the future asset value, and the signals which measure future value will indicate this uncertainty as increased disagreement.

Forecast dispersion, derived from analyst estimates, is also a measure of disagreement. Unlike other disagreement measures in subplot 2, $FDISP$ ranks doesn't vary as strongly with STD_DEV deciles. Over the ten deciles, dispersion ranks goes from 0.3 to 0.6. Even though the two measures of disagreement are related, they measure two different types of disagreement. Dispersion in analyst forecast represents the dispersion in the private information content of analysts (Barron, Kim, Lim, and Stevens (1998)), and hence it is a proxy for differential content of private information possessed by analysts. It is thus different in form and structure from the fundamental disagreement, which originates as a result of differential interpretation of commonly available return predicting signals. Forecast dispersion and fundamental disagreement are two measures of disagreement that differ by their origins and hence capture different types of heterogeneity among investor's trading motives. We should not expect these to be substitutes in a regression setup where disagreement explains some property of asset prices like returns, volatility, or trading activity.

Similarly, disagreement measure is largely unrelated to the number of analysts following a firm. The mass of analyst following represents the interest of informed agents in a particular stock and hence is a measure of the level of informed trading in the stock. Trading due to disagreement arises purely from non-informational[17] reasons. The relationship between the number of analysts and disagreement is slightly negative and $NUMEST$ ranks decrease from 0.15 to 0.05 over ten deciles.

Disagreement is also significantly higher among small and younger firms. Across the ten disagreement deciles, the ranking for firm size and age falls by roughly 40 percentiles. O_SCORE, a measure of distress risk, is also significantly higher for high disagreement firms. Disagreement also associates negatively with book to market ratio where growth firms (having low BTM ratio) tend to be firms with high disagreement. Growth firms derive a major chunk of their value from real options and growth opportunities which makes their revenue projections difficult to estimate. This may lead to higher disagreement among investors about the future asset value of these firms.

Measures of firm complexity like $COMPLEX\_WORDS$ and NUM_SEG and measures of document readability like DOC_SIZE and NUM_WORDS vary mildly with disagreement. Institutional ownership declines sharply as disagreement rises while short interest rises with disagreement. Illiquidity, measured by $ILLIQ$ and $SPREAD$, significantly increases with disagreement deciles. It is important to note that both turnover and illiquidity rises with disagreement.

Overall, the average firm in high disagreement decile pays no dividend, experience volatile returns, is small and young, has high earnings uncertainty, has low BTM ratio and low institu-

[17]The superior information of the analyst is the private information available only to the analyst utilizing her stock/industry research, management interviews, and information of the macro-economy. This information is different from the information contained in fundamental signals considered in this study. These signals are common information, and investors are assumed to interpret the common information differently, leading to fundamental disagreement.

tional ownership, experience high distress risk and illiquidity when compared to a firm in low disagreement decile.

Disagreement and Industry Characteristics

Evidence from figure 6 suggests that disagreement is related to fundamental uncertainty, size, and risk characteristics of firms. Since different industries vary in their risk characteristics, should disagreement be different for firms in different industries? For instance, should disagreement be higher for firms in established industries like utilities that have a constant source of revenues or should it be higher for firms in the pharmaceutical industry whose value depends a lot on drug trials, regulatory approvals, and R&D investments? Looking at which industries cluster at low and high disagreement deciles gives more insight into the nature of disagreement.

Figure 7 gives for each disagreement decile the percentage of firm-month observations belonging to a particular industry. The relative concentration of each industry changes across disagreement deciles. The 48 industries (Fama and French (1997)) are presented in 8 subplots of 6 industries each. Industries are sorted on their relative concentration, and then the smallest group is presented in subplot 1 and the largest in subplot 8. This arrangement puts industries with several firms in the later subplots.

Looking at the last four subplots, industries like drugs, electronic chip making, IT services, computers, medical equipment, retail, and oil are the ones whose concentration increases with disagreement decile. While industries like utilities, machinery, building materials, chemicals, food, office supplies, lab equipment, steel, transport, and auto show the opposite trend, i.e., their concentration falls with disagreement deciles.

[Insert Figure 7 here.]

Industries with real options like IT, computers, pharmaceuticals, and oil[18] are harder to value since their future income streams are dependent on the exercise of these options. Similarly, other growth industries like financial trading, electronic equipment, and semiconductor chip-making are also highly concentrated in high disagreement deciles. Established industries with a smooth revenue stream like auto, chemicals, machinery, business supplies (paper), and utilities are easier to value. These industries are concentrated in lower disagreement deciles.

Overall, industries that possess fundamental uncertainty, be it uncertain future cash flows, growth opportunities, patents, global risks, real options, and regulatory hurdles, concentrate at high disagreement. Increased fundamental uncertainty makes valuation difficult, which is reflected in higher disagreement originating from return anomalies.

[18] Firms in the oil and drug industry derive their real options from exploration activities. Firms in the oil industry have an option to explore a potential oil field, whether to develop the site and finally, when to start drilling and extraction of petroleum products. Besides, the oil industry is also heavily dependent on global crude prices, which adds an extra degree of uncertainty. The drug development process is also full of real options at various stages of drug development, clinical trials, regulatory approvals, and patents.

6 Results and Discussion

In this section, I present the main results of the impact of disagreement on trading volume. Following Chordia et al. (2006), we estimate predictive regressions and use one month lagged values for all right-hand side variables. Turnover as measured by monthly dollar volume scaled by market capitalization, L_TURN is regressed on the disagreement measure, STD_DEV, and a host of control variables and fixed effect dummies. I use the same set of firm characteristics previously studied in Chordia et al. (2006) as control variables. To control for the time trend in volume and the possibility of it varying across industry types, I include dummy variables for each year of the sample and the 48 industry types identified in Fama and French (1997)[19]. I estimate the below regression (without intercept),

$$L\_TURN_{i,t} = \sum_k \alpha_k \cdot Controls_{k,i,t-1} + \beta \cdot STD\_DEV_{i,t-1} + \sum_d \gamma_d \cdot Dummies_{i,t-1} + \epsilon_{i,t} \tag{6.1}$$

The controls used are past returns separated into positive and negative returns (RET^+ and RET^-), firm leverage measured by debt to equity ratio (LEV), beta coefficient from firm-level CAPM regression ($CAPM\_BETA$), book to market ratio (BTM), current market price (L_PRC), firm's age as measured by logarithm of number of months since firm's first trading (L_FAGE), firm size (L_ME), absolute surprise in a firm's earnings and the volatility of earnings ($ESURP$ and $EVOL$), number of analyst following a firm and their forecast dispersion ($NUMEST$ and $FDISP$). To control for differences in trading structure between NYSE/AMEX and NASDAQ exchanges and any double counting issues identified by Atkins and Dyl (1997), I include a dummy for NASDAQ stocks in all regressions. A preceding $L\_$ ahead of a variable means its logarithm is used in the regression. A complete description of variable definitions and their construction appears in Appendix A.2.

Table 2 Panel A gives the correlations among several control variables and the disagreement measure, i.e., all variables that appear on the right-hand side of equation 6.1. High correlation between RET^+ and RET^- is imparted due to the construction of these variables making RET^- zero (negative) whenever RET^+ is positive (zero). Their correlation is inconsequential since both of them are never non-zero simultaneously. L_ME is strongly correlated with several other variables. Its correlation with L_PRC is mechanical; with $NUMEST$ is due to correlation between analyst following and firm size. $ESURP$ and $EVOL$ have stock price in the denominator and that might be driving the negative correlation with L_ME. The largest correlations occur between L_PRC, L_ME and $NUMEST$. Out of three pairs of variables, L_PRC and $NUMEST$ has the least correlation. Moreover, L_ME is also strongly correlated with $ESURP$ and $EVOL$. Hence, to avoid multicollinearity concerns, I drop L_ME from the regression design. The other significant correlation is between $ESURP$ and $EVOL$, which may also be driven by their common denominator. Disagreement (STD_DEV) is positively correlated with LEV, $ESURP$, and $EVOL$. It is higher for smaller (L_ME), younger (L_FAGE) and low priced (L_PRC) firms. Lastly, STD_DEV mildly correlates with forecast dispersion ($FDISP$), which is also a measure of disagreement. This is desirable in establishing STD_DEV as a measure of investor disagreement

[19]The list of industry codes is available at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_48_ind_port.html

and is not merely a different proxy for analyst dispersion. Interestingly, disagreement is much more correlated with negative returns than it is for positive returns.

Table 2 Panel B presents summary statistics of the explanatory variables outlined above. At each month time-series averages of all variables are computed and their descriptive statistics are presented in Panel B. L_FAGE, truncated variables (RET^+ and RET^-), and earnings-related variables ($ESURP$ and $EVOL$) have high skewness and kurtosis. Except for L_FAGE and RET^-, all are positively skewed. The disagreement measures (STD_DEV) is relatively symmetric.

[Insert Table 2 here.]

In Table 3, I regress[20] log turnover (L_TURN) on several combinations of controls and the disagreement measure (STD_DEV). The first specification is the base regression from Chordia et al. (2006)[21]. The sign and relative importance of coefficients in the specification (1) are in harmony with Chordia et al. (2006) results. The sole difference is that I find the sign on BTM to be negative while Chordia et al. (2006) find it to be insignificant. A negative coefficient on BTM ratio could also be due to a positive association between turnover and market capitalization. I purposefully leave out L_ME from regression specification to avoid multi-collinearity[22] arising due to tightly correlated regressors. Except for RET^-, BTM and L_FAGE, all variables have a positive coefficient. The effect of $EVOL$ on turnover is not statistically significant after including STD_DEV. Inclusion of disagreement (STD_DEV) reduces the coefficient of $FDISP$ by half but not the other way round. Only the impact of $CAPM\_\beta$ and $NUMEST$ remain constant while that of L_PRC is increased as a result of including disagreement. The latter finding can be attributed to the fact that several anomalies that enter disagreement computation have stock price in denominator. This induces mechanical negative correlation with STD_DEV and may be the reason of rise in it's coefficient. The impact of all other variables reduces between 20 to 50 percentage points.

The coefficient on STD_DEV is stable across specifications, and in the full model (specification (4)), R^2 increases by roughly 3.5% (0.439/0.424). I report two R^2 for each regression: adjusted R^2 is for the full model including fixed effects, while Within R^2 is for within-group variation with fixed effects projected out. $\%R^2$ explained is the percentage of unexplained variation explained by including disagreement in the regression[23]. Disagreement roughly explains 5.2% of the unexplained variation of the full model. The coefficient on STD_DEV is 1.464, and its standard deviation is 0.0986, which means that a 1 SD increase in STD_DEV predicts the next month's log turnover to increase by 0.144 (1.464*0.0986). This is equivalent to 15.5% ($e^{0.144}$) increase in turnover. If I exclude earnings (forecast) related variables, the net effect increases to 16% (16.6%)

[20] All regressions reported in this paper are executed on R using the lfe package (https://cran.r-project.org/web/packages/lfe/lfe.pdf) and verified on Stata's reghdfe package (http://scorreia.com/help/reghdfe.html).

[21] Specification (10) in Table 3 of Chordia et al. (2006) is similar to the specification (1) in Table 3.

[22] In unreported results, I additonally include L_ME as one of the regressor and found similar results. The coefficient on BTM is still negative but toned down. The coefficient on L_ME is positive and there is no significant change in the impact of STD_DEV on turnover after including firm size. However, the coefficients on $NUMEST$, L_PRC and $EVOL$ gets significantly reduced.

[23] Let R^2_w ($R^2_{w/o}$) be the R^2 from a regression that includes (excludes) disagreement. Then $\%R^2$ explained is the ratio of $R^2_w - R^2_{w/o}$ and $1 - R^2_{w/o}$.

for 1 SD change in disagreement. On the other hand, a change in 1 SD in the other measure of disagreement, $FDISP$, only predicts 1.1% higher turnover in presence of STD_DEV.

All regressions have industry and year fixed effects, and the standard errors are double clustered by firm and month. Petersen (2009) finds that if there are at least 50 clusters in each dimension, then double clustering of standard errors is closer to true standard errors than Fama-MacBeth regressions with Newey West correction. Since there are many more firms and almost 700 months of observations (1962 to 2019), there will always be enough clusters, even with sub-sample analyses.

[Insert Table 3 here.]

The next set of regressions evaluates the effect of disagreement on more immediate daily and weekly turnover. In specification (4) of Table 3, henceforth called the base regression, next month turnover is predicted using variables constructed at the end of the current month. In Table 4, I explore the relationship between the end of the month disagreement and turnover measured during the first day and the first week[24] of the next month. The six regressions of Table 4 can be grouped into two groups: first three, which doesn't include disagreement and the last three, which includes STD_DEV. Comparisons can be made both across the group (with and without disagreement) and within each group (turnover over daily, weekly, and monthly intervals). The coefficient on STD_DEV and the R^2 gently increases over the regressions 4-6 providing evidence that disagreement affects turnover across different time intervals. RET^+, RET^- and $CAPM\_\beta$ are much more impactful in explaining daily turnover as their coefficients falls steeply in magnitude with increase in turnover horizon. Visibility related variables like L_PRC and $NUMEST$ also decrease with turnover horizon. On the other hand, disagreement and dispersion coefficients ($FDISP$ and STD_DEV) increase in magnitude wih increasing turnover horizon. This may be due to the effect of liquidity and visibility being more short termed than disagreement and fundamental uncertainty. For instance, portfolio rebalancing following substantial return changes should materialize in the next few days and not spread over the entire month. Like the previous regression, inclusion of STD_DEV reduces $FDISP$ coefficient by half. Looking at $\%R^2$ explained, disagreement arising out of return anomalies today explains upto 3.5% of unexplained variation in tomorrow's turnover.

[Insert Table 4 here.]

The strong disagreement-volume relationship may be driven by the choice of the dependent or independent variable. To fully convince ourselves about the validity of results, I assess the effect of trying different measures of turnover on the same set of explanatory variables and the disagreement measure STD_DEV. In Table 5, I present regression results from seven different turnover measures where specification (1) is the base regression. The coefficient on STD_DEV vary based on the dependent variable, but the broad consensus is that a higher disagreement of anomaly signals predict a higher turnover next month. However, the positive

[24] I define the first week of the month as the first five trading days immediately following the end of the previous month.

effect of forecast dispersion changes sign when turnover is detrended or adjusted for market level turnover (specifications 4,6 and 7). Results of regression (2) are unsettling since the coefficient on RET^+ and RET^- are reversed, which is completely unexpected and in stark contrast to all other regressions. Coefficients on other variables are also severely attenuated. This is mostly due to the choice of dependent variable wherein L_TURN_{t-1} is subtracted from L_TURN_t. This takes away all interesting variation from L_TURN since it has a highly persistent time-series. It also imposes a coefficient of one on the first autocorrelation parameter, whereas in the sample it is 0.87. In specification (3), the dependent variable is GRT adjusted log turnover, which is the same as used by Chordia et al. (2006). In regression (5), the dependent variable is the residual from regressing L_TURN on Amihud (2002) Illiquidity measure. To substantiate that the channel through which disagreement affects turnover is not related to liquidity, removing any liquidity related variation from turnover should not affect the disagreement-turnover relationship. Lastly, in specifications (6) and (7) the dependent variables are the residuals from regressing L_TURN on value-weighted and equally-weighted log market turnover respectively. This removes any market-wide component in L_TURN.

[Insert Table 5 here.]

Similarly, in the next table, I evaluate the effect of varying the way disagreement is measured. In Table 6, L_TURN is regressed on controls and different sets of disagreement measures whose definition appears in Appendix A.2. For brevity, only the disagreement variables are shown in the regression table. The coefficients are all positive and significant. All measures except the number of flips and divergences explain roughly 5-7% of unexplained variation in turnover with continuous deviation ($CONT\_DEV$) explaining the most. A one SD change in STD_DEV, ABS_DEV, PC_DEV and $CONT\_DEV$ predicts 14.7%, 14.5%, 15% and 17.2% higher trading volume in next month respectively. A one SD change in LO_CORR_SD and HI_CORR_SD, the two disagreement measures from non-overlapping sets of low and high correlation signals, explains 9.8% and 6.5% of next month's volume. The overall explanatory power is similar to other measures, but mildly correlated signals have a relatively higher explanatory power. The evidence in Table 6 provides strong support for the relation between disagreement among anomaly signals and turnover. It doesn't matter how disagreement is measured; the relationship is consistently positive and significant. The number of flips and divergences, as defined in Kandel and Pearson (1995) and applied to anomaly signals rather than analyst forecasts, although significant, explain very little next month turnover.

[Insert Table 6 here.]

My results may be driven by outliers, wherein a small number of observations have a high disagreement as well as large turnover. The standard in empirical literature to account for outliers it to winsorize the variables. At each month, I winsorize the sample based on the independent variables used in the regression at 0.5% on either extreme. To further solidify the results from previous tables, I also perform rank regressions where the respective cross-sectional rank of a given variable is used rather than the variable itself. The details of how ranks are calculated are present in Appendix A.2. Using ranks have some additional advantages: (i) outliers do not drive the hypothesized relationship, (ii) interpreting the results become more intuitive since both the explanatory and dependent variable are ranks, and, (iii) different variables can be compared with each other and also over time. In Table 7, I present rank regressions where log turnover and its rank is regressed on ranks of controls and the rank of disagreement measure. A suffix of $\_R$ represents the rank of that variable. For comparison, I report regression without STD_DEV_R as well.

Comparing specifications (2) and (4) with specifications (1) and (3) respectively shows that the magnitude of almost all control variables declines after including STD_DEV_R. Specification (4) is the most robust regression since all variables are represented as their cross-sectional ranks. Moving from 25th to 75th disagreement percentile predicts next month turnover rank to be higher by 6.5 percentiles ($0.130 \times (0.75 - 0.25)$). The inclusion of disagreement improves the explanatory power of turnover regression by 3.15%. Similarly, from regression (2), moving from 25th to 75th disagreement percentile predicts 27.4% ($e^{0.485*0.5}$) higher turnover in next month. In rank regressions also, inclusion of STD_DEV_R reduces the impact of $FDSIP\_R$ by roughly 20% with disagreement explaining upto three times more turnover than forecast dispersion.

[Insert Table 7 here.]

Finally, to investigate whether the result is driven by the size of the firm or a particular time period within the entire sample duration, I perform robustness check for size terciles and nine non-overlapping 5-year subperiods.[25] The latter also tests the out of sample performance of the hypothesized relation between disagreement and volume. The three size portfolios are made using 70/30 NYSE breakpoints done separately for each month. The $SMALL$, $MEDIUM$ and BIG group contains roughly 75%, 48% and 20% NASDAQ stocks respectively. Thus, most of the NYSE/AMEX stocks are concentrated in the BIG group. For sub-period analysis, the sub-periods start in 1975 and continue until 2019, making nine 5-year periods. Analyst forecast data is available from January 1976 and quarterly earnings data from July 1971; thus, sub-periods can only extend till 1975 using 5-year periods. Even then the first sub-period between 1975 and 1979 has a smaller sample size compared to the rest of the sub-periods.

Table 8 presents the regressions for three size terciles. Specifications 1,3 and 5 are without disagreement, while specifications 2,4 and 6 include disagreement. R^2 is highest for medium sized firms and lowest for small firms. The coefficient on STD_DEV for small and medium firms is considerably higher than the coefficient on large firms. This substantiates the hypothesis that investors in smaller firms tend to use return anomalies more for their investing needs, and hence disagreement arising from these anomalies is reflected strongly in next month's trading volume. Disagreement is significant both economically and statistically across the three size groups. The effect of L_PRC and $NUMEST$, two proxies for visibility, is limited to smaller stocks only. Across the three size groups, a one SD change in disagreement predicts next month's turnover to be higher by 14.4%, 18.6% and 10.5% for small, medium, and big stocks respectively. For forecast dispersion, similar estimates are 1.4%, 0.7%, and 1.8%.

[25] Internet Appendix also contains robustness checks using different stocks splits, across exchanges and over BTM terciles.

[Insert Table 8 here.]

Table 9 has nine regressions for the nine sub-periods. Regression (1) has a smaller sample size due to partial availability of forecast data, and hence the results for this sub-period are less reliable. The coefficient on STD_DEV is significant and positive in all sub-periods and ranges from 0.58 to 1.99. A one SD change in disagreement predicts up to 21.9% higher volume next month for the period 1995-1999. STD_DEV explains almost 4.5% of the unexplained variation for this period. On the contrary, $FDISP$ is insignificant in all but one sub-periods destroying all its explanatory power when subjected to different subsamples.

[Insert Table 9 here.]

7 Portfolio Sorts

Linear regression imposes a linearity assumption on the population regression function. In general, the relation may be nonlinear. Portfolio sorts can further discern any nonlinearities. Apart from univariate and bivariate sorts, higher-order sorts are harder to visualize as well as suffer from the curse of dimensionality. As the number of variables to sort on increases, the total number of portfolios rises exponentially. For instance, using terciles (3 portfolios) for each variable, the base regression with 12 variables and one NASDAQ dummy will require more than a million portfolios ($2 \cdot 3^{12}$). Increasing the number of portfolios reduces the number of firms in each portfolio, and subsequently, the inference is weakened. Hence I form portfolio using only univariate and bivariate sorts.

Table 10 shows mean turnover and changes in mean turnover over ten deciles sorted by several control variables and STD_DEV. At the beginning of each month, stocks are assigned portfolio deciles based on their cross-sectional rankings. The variables used to form deciles are lagged by one month. Column 1 is the mean L_TURN in the first decile; column 2 is the amount by which L_TURN increases as we move from the first decile to second decile and so on. RET^{+} and RET^{-} have a near monotonic and positive effect on turnover. The negative coefficients on RET^{-} in the second row are due to $RET^{-} \leq 0$ by construction. The positive effet of $NUMEST$ dies rapidly and turns negative for higher deciles. The effect of $CAPM\_\beta$ and L_PRC (BTM and L_FAGE) is consistent and positive (negative). $ESURP$ and $EVOL$, two variables capturing fundamental uncertainty, surprisingly associate negatively with turnover for half of the decile changes. BTM proxies for both visibility (growth stocks are more visible) and value investing (investors relying heavily on fundamentals). For the smaller BTM deciles, the effect is negative consistent with the visibility interpretation, but for the last decile, turnover increases significantly consistent with the value investing story.

The positive effect of forecast dispersion ($FDISP$) is limited to first three decile changes and thereafter it turns insignificant or negative. For disagreement (STD_DEV), successive decile changes are positive and gradually increase with the last decile change affecting the turnover most. The evidence from how $FDISP$ and STD_DEV deciles affect turnover corroborate the idea that the two disagreement measures capture different type of investor disagreement where forecast dispersion represents the dispersion in private information while disagreement arising from return anomalies as a result of differential interpretation. An increase in $FDISP$ may generate information asymmetry instead of disagreement and may cause a reduction in trading. STD_DEV, on the other hand, arises purely from interpretation (non-informational) reasons and hence always positively impacts turnover.

[Insert Table 10 here.]

When it comes to two-way portfolio sorts, there are two approaches: independent sorts and dependent sorts. For independent sorts, in the first stage, univariate portfolios are formed independently for both sorting variables, and then their intersection gives bivariate portfolios. The order of sorting is irrelevant. However, if the two variables are correlated, then the firms will be disproportionately allocated across portfolios, and inference becomes less reliable. Dependent sorts can solve this problem where one variable is used to form univariate sorts first, and then another variable is used to sort within the portfolios from the first stage. This makes sure that each portfolio has a similar number of firms. The order of sorting is important here. With dependent sorts on variables A and then B, we can study the effect of B while controlling for A. I use dependent sorts for the bivariate portfolio results. Chapter 5 of Bali, Engle, and Murray (2016) discusses portfolio sorts in detail.

In Table 11, I present average L_TURN over 3×3 portfolio terciles sorted first on one of the control variables and then on disagreement measure STD_DEV. For each control variable tercile, I compute the average turnover for first disagreement tercile and then the successive tercile increase in turnover as we move to second and thrid disagreement terciles. The three control variable terciles are grouped as three broad columns; within each column, the average turnover and its changes over disagreement terciles are shown. For instance, $T_{i,j} - T_{i,j'}$ is the difference between disagreement terciles j and j' within control variable tercile i. All 66 tercile changes are positive and significant validating the strong positive impact of disagreement on trading volume after individually controlling for various controls. Moreover, disagreement significantly impacts turnover even after controlling for forecast dispersion further strengthening the view that STD_DEV and $FDISP$ capture two different types of investor disagreement.

[Insert Table 11 here.]

8 Disagreement and Information Environment of Firms

In this section, I evaluate how a firm's information environment affects the disagreement volume relationship. In particular, how does the amount of information available regarding a firm's operations, financials, and management alter the disagreement-volume relationship. I derive several variables measuring readability, complexity, and staleness of the firm's 10-K EDGAR filings. I also consider the impact of analyst following and institutional ownership on the information environment. I begin by studying the impact of EDGAR implementation on disagreement-volume relationship followed by a discussion of variables affecting the firm's information environment. Lastly, I construct a measure of firm by firm information environment.

8.1 EDGAR Implementation

Securities and Exchange Commission (SEC) mandated online filing of corporate disclosures through Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system in a phased manner from April 1993 through May 1996.[26] Prior to this rule mandatory disclosures, like 10-K and 10-Q, were not readily accessible to the general public. The availability of accounting fundamentals present in annual reports is essential for constructing the return anomalies. In lieu of easy access to fundamentals, the use of return anomalies would be limited. Post EDGAR implementation, the use of firm disclosures for investment decisions should increase, subsequently increasing the reliance on return anomalies. With the increasing use of return anomalies for trading decisions, disagreement arising from these anomalies will have an increased impact on trading activity. Hence, we should expect the disagreement volume relationship to be stronger post EDGAR implementation.

Figure 8 gives the annual regression coefficients for disagreement. The average coefficient before 1993 is 0.966, while after 1996, it increases to 1.485. The level of coefficients shifts to a higher value around EDGAR implementation. The percent of unexplained variation (unreported) also show a similar trend. Table 12 gives regression results from two disjoint periods: pre and post EDGAR implementation. As hypothesized, the coefficient of disagreement is much larger after EDGAR implementation. A one SD change in disagreement predicts 10.4% and 15.4% higher trading volume next month for pre and post EDGAR periods, respectively. Disagreement coefficient varies significantly with time, validating the inclusion of year fixed effects in all regression specifications.

[Insert Figure 8 here.]

[Insert Table 12 here.]

8.2 Information Environment

The most important source of a firm's information is the 10-K EDGAR filings, which are disclosed annually by all firms. Annual reports are the mechanism by which the company's manager reduce the information gap between insiders and investors. Moreover, the efficacy of this process depends on the readability and complexity of the report. If the report is convoluted, then it becomes difficult to extract valuation relevant information. Following the release of a complex and difficult to read annual report, investors would find it difficult to gauge the future value of the firm and would resort to return anomalies for their investment decisions. This would increase the disagreement volume relationship where disagreement arises from the dispersion of return anomaly signals.

Analyst forecasts act as a vehicle of information intermediation between firm insiders and external investors, thereby reducing the information asymmetry. Analysts, through their stock research, knowledge of the industry, management interviews, and analysis of firm disclosures, generate value relevant information not otherwise available publicly. Firms not followed or thinly

[26] Appendix A of Gao and Huang (2019) gives the time line of the staggered EDGAR implementation. Companies were required to file electronically in 10 groups.

followed by analysts lack this medium of information dissemination. In the absence of value relevant private information, in the form of earnings forecasts, investors rely on other simpler and readily available forms of assessing the future value of stocks. One class of such information signals is the return anomalies documented in the literature. To the extent that this lack of information distribution by analyst cause investors to rely on trading signals originating from return anomalies, the disagreement arising out of these anomalies should predict an accentuated trading volume response.

Institutional investors, typically holding large portions of a firm, are desired as careful monitors of the firm. Larger the shareholding of an institutional investor, the more effective is their monitoring as the costs and efforts of monitoring are easily balanced by the gains from better information about the firm. The presence of an information specialist, like analysts or institutional investors, reduce investor's need for valuing the firm by themselves. This change in behavior materializes as reduced dependence on return anomalies.

Document Readability: Loughran and McDonald (2014) define readability as "the ability of individual investors and analysts to assimilate value-relevant information from a financial disclosure" and argue that the most popular measure of readability - Fog Index - is poorly specified for financial disclosure and argue that the file size in megabytes of the firm's 10-K filing is a readily available measure of readability. If a firm must disclose bad information, obfuscating it with value irrelevant details is one way to diffuse such information. A less readable report presents more difficulties for investors to extract valuation-related information. In the absence of easy information flow, investors' use of return anomalies increases, giving rise to increases trading volume reaction following disagreement in anomaly signals.

Firm Complexity: Complexity of quarterly and annual reports could also pose difficulties in gauging the performance of firms. These disclosures provide additional information (other than B/S and P/L items) regarding the current state of the firm. A lengthier (wordier) disclosure presents a complex and convoluted snapshot of the firm performance and hence difficult to interpret. For such cases falling back to a simpler anomalies based investment could be favored. We should expect the use of return anomalies to increase for lengthier disclosures, and subsequently, the disagreement volume relationship would be strengthened. Firm complexity can be more directly captured by counting the number of business segments of a firm. It is difficult to accurately value a firm involving multiple segments and their interactions. Loughran and McDonald (2020), in a recent paper, argue that a simple text-based measure is a better, easily available, and reproducible measure of firm complexity. They construct a list of 374 words most commonly attributed to some form of business or information complexity of a firm and define firm complexity as the unique occurrences of these words in a firm's 10-K filing.

Stale Reports: Accounting information is disclosed annually while trading happens continously based on the same information. During annual announcements, investors would use firm relevant information disclosed in the announcement as well as the stock price and fundamentals to assess the future value of the stock. However, once the announcement becomes stale, the only source of new information is the stock price. Since many of the return anomalies (like valuation and momentum) use stock price in conjunction with balance sheet items, the use of return anomalies should increase with the staleness of 10-K reports. A new filing provides value relevant information with which investors can take trading positions. As the report becomes stale, the dependence on 10-K reports is reduced, and this should result in an increased shift on return anomalies. Thus, I hypothesize that as time passes post 10-K filing, the disagreement-volume relationship should get stronger.

Analyst Following: Analysts' recommendations can substitute for anomalies as investment advice. If a firm is not followed by many analysts, then investors would rely more on anomalies from accounting fundamentals and stock price for their trading decisions. Increased reliance on anomalies should increase the trading volume reaction arising from disagreement within these anomalies. Thus we should expect the disagreement-volume relationship to be stronger for firms with small analyst following.

Firm Age: Young firms are generally difficult to value due to limited history, scanty analyst following, and illiquid assets (from private equity). Falling back to a formula based investing using return anomalies presents an easier alternative, and hence I predict the disagreement volume relationship to weaken as the firm matures.

Shareholding Pattern: The importance of value relevant news is higher with dispersed shareholding. In the absence of a large institutional shareholder, who can perform the task of monitoring the firm effectively, common investors would demand information from intermediaries such as analyst forecasts. Hence, the absence of information intermediaries in a firm with dispersed shareholding would tempt investors to use return anomalies for their investment needs. Subsequently, the trading volume reaction following disagreement in anomaly signals would be higher for firms with low institutional ownership.

In Table 13 Panel A, I present correlations among some of the variables identified above. Panel B has descriptive statistics. The file size of the EDGAR 10-K filing measures readability. Firm complexity is gauged by the number of words, unique words, and complex words in 10-K filing as well as the number of firm segments. Staleness is the time (in months) since the last annual filing. There is a very high correlation between number of words, unique words and complex words. This is expected since all these variables are derived from parsing 10-K words. Number of firm segments and unique segments are also highly correlated. Firm size is strongly correlated only with number of analyst following and institutional ownership percent. Firm size's relation with variables extracted from EDGAR 10-K files is positive but small in magnitude. This supports the claim that firm size is only a weak proxy for firm attributes like readability and complexity (see Loughran and McDonald (2020) for a discussion on size and complexity).

Table 14 gives a summary of regression results across portfolios made using variables related to information environment. I look at monthly terciles of firm age, document size, report length, number of words, number of unique words and complex words. For others (number of analysts, number of segments, ownership percentage and months since filing), I use subsamples[27]. Detailed regression tables for all these variables are present in Internet Appendix. For brevity, only $FDISP$ and STD_DEV coefficients are shown. Each row corresponds to two regressions: base regression without disagreement (Table 3, spec 1) and base regression with disagreement (Table 3, spec 4). The portfolio formation criterion is present in first column while the last column gives the

[27]Subsamples are made on entire sample for some variables due to their limited range. For instance, $NUMEST$ doesn't vary much across firms. This results in vague portfolio boundaries. I do not get reasonable sample sizes when using terciles of such variables. Hence, for these variables, I choose portfolio boundaries using the entire sample.

percentage increase in R^2 after including disagreement in the regression.

For all variables splits, inclusion of disagreement either reduces or eliminates forecast dispersion coeffcient. $FDISP$ coefficient loses statistical significance in 20 out of 29 portfolios. STD_DEV remains both ecomically and statistically significant in all regressions. Except for firm segments, disagreement coefficient increases in the hypothesized direction validating the increased use of return anomalies as firm's information dissemination channel becomes more opaque and inefficient. % R^2 explained is also usually higher when STD_DEV is higher.

[Insert Table 13 here.]

[Insert Table 14 here.]

8.3 Firm-level measure of information environment

In the previous section, I argue that the magnitude of the disagreement-volume coefficient captures the underlying information environment of the firm. Firms with broken and complex channels for information dissemination undergo increased trading volume reactions owing to the disagreement of fundamentals. I show evidence of several firm specific attributes like firm age, number of analyst following, institutional ownership, and characteristics of 10-K filings like report length, readability, document size, and complexity affect the disagreement-volume relationship.

The above findings show the wedge in the disagreement coefficient, a meaaure of the firm's information environment (w.r.t. outside investors), through a single panel regression. On the other hand, a firm by firm measure can be useful in studying the properties of firms having a higher disagreement coefficient. It can also be used as a proxy for a firm's information flow efficiency. However, the nature and construction of disagreement measure is essentially cross-sectional[28] making a firm by firm time-series regression not very useful. To get a measure of firm by firm estimate of a firm's information environment measure, I hypothesize that the disagreement coefficient is a linear function of the below firm specific attributes. Writing β from Eq. 6.1 as,

$$\begin{aligned}\beta = \lambda_0 + \lambda_1 \cdot NUMEST_{i,t-1} + \lambda_2 \cdot L\_FAGE_{i,t-1} + \lambda_3 \cdot DOC\_SIZE_{i,t-1} \\ + \lambda_4 \cdot NUM\_WORDS_{i,t-1} + \lambda_5 \cdot NUM\_TOT\_SEG_{i,t-1} + \lambda_6 \cdot COMPLEX\_WORDS_{i,t-1} \\ + \lambda_7 \cdot OWN\_PERC_{i,t-1} + \lambda_8 \cdot SINCE\_10K_{i,t-1}\end{aligned} \tag{8.1}$$

Substituting Eq. 8.1 in Eq. 6.1 gives the below regression specification,

$$\begin{aligned}L\_TURN_{i,t} = \sum_k \alpha_k \cdot Controls_{k,i,t-1} + \sum_d \gamma_d \cdot Dummies_{i,t-1} + \\ \left(\lambda_0 + \sum_l \lambda_l \cdot FC_{l,i,t-1}\right) \cdot STD\_DEV_{i,t-1} + \left(\sum_l \theta_l \cdot FC_{l,i,t-1}\right) + \epsilon_{i,t}\end{aligned} \tag{8.2}$$

[28]Return anomalies are used to construct a cross-sectional ranking. Since the process is repeated for each time period, a return anomaly signal can remain constant for a firm over time. This will impart very little time-series variation in disagreement arising out of these signals.

Since, we are interested in how the firm characteristics ($FC_l, l \in 1, \ldots, 8$) impact turnover by interacting with disagreement, I also include firm characteristics without their interaction with disagreement to control for any direct impact of these firm specific variables on turnover[29]. β varies with firm and over time with variation in FC_l across firms and time. After estimating $\lambda_0, \ldots, \lambda_8$ from Eq. 8.2, I compute information environment, $INF\_ENV_{i,t-1}$, as $\lambda_0 + \sum_l \lambda_l \cdot FC_{l,i,t-1}$. I estimate Eq. 8.2 using variable levels as well as ranks. I call the latter INF_ENV_R.

To validate that my measure of information environment indeed captures the efficiency with which firm specific information flow to investors, I check whether INF_ENV can explain next period return volatility. If the channel of information dissemination is disrupted or inefficient, then value relevant information would flow in a staggered manner. The non-contagious flow of information would cause a non-contagious impact on stock prices. As stock price gradually reflects information, it gives rise to an increase in return volatility. Another variable strongly related to the firm's price informativeness is bid-ask spread, which represents information asymmetry among market participants. Incomplete and incongruous information flow reduces investor's confidence in firm valuation and increases the wedge between buyers and sellers. As a result, the bid-ask gap widens. To the extent that INF_ENV captures the inefficiency of a firm's communication to investors, a higher value should predict a higher bid-ask spread.

I estimate a simple OLS regression model of return volatility and bid-ask spread on information environment. I only include two other variables as controls: disagreement and firm size. Disagreement is a natural choice for a control since by definition, it is the standard deviation of return anomalies and volatility is the standard deviation of realized returns. Deviation in a set of variables known to predict return should also predict deviation in returns.

Size is included for three reasons. First, size is a catch-all proxy for several other firm features which we do not include. Second, smaller firms are usually more volatile and also exhibit large bid-ask spreads. Hence, size should negatively relate to both volatility and spread. Third, size also captures a firm's information environment since smaller firms differ substantially from larger firms in their information dissemination structure. I exclude other variables that can also predict volatility to simplify the discussion. Table 15 shows the results, and I find the firm by firm measure of information environment to significantly and positively predict next period return volatility and bid-ask spread. Using INF_ENV_R, the coefficient is significant for predicting bid-ask spread but not for return volatility.

[Insert Table 15 here.]

9 Conclusion

I construct a new measure of investor disagreement using 36 return anomalies where disagreement is the standard deviation of trading signals emanating from these anomalies. Return anomalies enter as likelihood function in investor's pricing model, and differences in likelihood function give rise to dispersion in beliefs about asset's payoff giving rise to increased trading. I consider all

[29] Khan and Watts (2009) use a similar method to estimate firm by firm estimates of asymmetric timeliness of earnings coefficient.

anomalies mentioned in Linnainmaa and Roberts (2018) and top it with momentum anomalies from McLean and Pontiff (2016).

Disagreement is higher for small, growth, and riskier stocks, which exhibit high fundamental uncertainty. High disagreement firms experience volatile returns, have high earnings uncertainty, have low BTM ratio and low institutional ownership, experience high distress risk, and illiquidity when compared to low disagreement firms. Industries that possess fundamental uncertainty concentrate at high disagreement. Increased fundamental uncertainty makes valuation difficult, which is reflected in higher disagreement originating from return anomalies.

Although fundamental disagreement positively correlates with forecast dispersion, the two measures are different and capture different aspects of investor behavior. Forecast dispersion represents the dispersion in the private information content of analysts, and hence it is a proxy for differential content of private information possessed by analysts. Disagreement arising from return anomalies originates as a result of differential interpretation of commonly available return predicting signals. Trading due to disagreement arises purely from non-informational reasons.

After controlling for prior determinants of trading volume, a move from 25th to 75th disagreement percentile predicts 27.4% higher turnover next period. Disagreement explains more than 5% of the variation in turnover that could not be explained by previous determinants of turnover. The presence of disagreement greatly reduces the impact of forecast dispersion on turnover but not the other way round reaffirming that the two measures capture different types of disagreement. Similar results also hold if instead of monthly turnover, I use the next day or next week's turnover. The positive and significant relationship is robust to different specifications, different measures of turnover, different measures of disagreement, across size groups, over different time periods, using rank regressions and portfolio sorts.

The coefficient of disagreement in turnover regressions measures the information dissemination inefficiency of firms. Using variables derived by parsing firm's 10-K reports, I find that disagreement coefficient is higher when a firm's disclosure is lengthier, has more complex words and when the disclosure becomes stale. Low analyst following, younger firms, and firms with low institutional ownership also exhibit a larger disagreement coefficient. The findings suggest that investors fall back to using return anomalies if the flow of value relevant information from firms is convoluted, opaque, disrupted, or delayed. Increased use of return anomalies materializes as the strengthening of disagreement volume relationship.

References

Abarbanell, J. S., & Bushee, B. J. (1998). Abnormal returns to a fundamental analysis strategy. *Accounting Review*, 19–45.

Akerlof, G. A. (1978). The market for "lemons": Quality uncertainty and the market mechanism. In *Uncertainty in economics* (pp. 235–251). Elsevier.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589–609.

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

Atiase, R. K., & Bamber, L. S. (1994). Trading volume reactions to annual accounting earnings announcements: The incremental role of predisclosure information asymmetry. *Journal of Accounting and Economics*, *17*(3), 309–329.

Atkins, A. B., & Dyl, E. A. (1997). Market structure and reported trading volume: Nasdaq versus the nyse. *Journal of Financial Research*, *20*(3), 291–304.

Bali, T. G., Engle, R. F., & Murray, S. (2016). *Empirical asset pricing: The cross section of stock returns*. John Wiley & Sons.

Bamber, L. S., Barron, O. E., & Stevens, D. E. (2010). Trading volume around earnings announcements and other financial reports: Theory, research design, empirical evidence, and directions for future research. *Research Design, Empirical Evidence, and Directions for Future Research (February 27, 2010)*.

Bamber, L. S., Barron, O. E., & Stober, T. L. (1997). Trading volume and different aspects of disagreement coincident with earnings announcements. *Accounting Review*, 575–597.

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

Barbee Jr, W. C., Mukherji, S., & Raines, G. A. (1996). Do sales–price and debt–equity explain stock returns better than book–market and firm size? *Financial Analysts Journal*, *52*(2), 56–60.

Barron, O. E., Kim, O., Lim, S. C., & Stevens, D. E. (1998). Using analysts' forecasts to measure properties of analysts' information environment. *Accounting review*, 421–433.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance*, *32*(3), 663–682.

Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of Accounting Research*, 67–92.

Beyer, A., & Guttman, I. (2011). The effect of trading volume on analysts' forecast bias. *The Accounting Review*, *86*(2), 451–481.

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *The Journal of Finance*, *43*(2), 507–528.

Bradshaw, M. T., Richardson, S. A., & Sloan, R. G. (2006). The relation between corporate financing activities, analysts' forecasts and stock returns. *Journal of Accounting and Economics*, *42*(1-2), 53–85.

Carlin, B. I., Longstaff, F. A., & Matoba, K. (2014). Disagreement and asset prices. *Journal of Financial Economics*, *114*(2), 226–238.

Chordia, T., Huh, S.-W., & Subrahmanyam, A. (2006). The cross-section of expected trading activity. *The Review of Financial Studies*, *20*(3), 709–740.

Cooper, M. J., Gulen, H., & Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *The Journal of Finance*, *63*(4), 1609–1651.

Daniel, K., & Titman, S. (2006). Market reactions to tangible and intangible information. *The Journal of Finance*, *61*(4), 1605–1643.

De Bondt, W. F., & Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, *40*(3), 793–805.

Diamond, D. W., & Verrecchia, R. E. (1981). Information aggregation in a noisy rational expectations economy. *Journal of Financial Economics*, *9*(3), 221–235.

Dichev, I. D. (1998). Is the risk of bankruptcy a systematic risk? *The Journal of Finance*, *53*(3), 1131–1147.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465.

Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, *43*(2), 153–193.

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

Fischer, P. E., Kim, C., & Zhou, F. (2020). Disagreement about fundamentals: Measurement and consequences. *Available at SSRN 3515685*.

Gallant, A. R., Rossi, P. E., & Tauchen, G. (1992). Stock prices and volume. *The Review of Financial Studies*, *5*(2), 199–242.

Gao, M., & Huang, J. (2019). Informing the market: The effect of modern information technologies on information production. *Review of Financial Studies, Forthcoming*.

Graham, B. (2006). *The intelligent investor*. Harper Business.

Harris, M., & Raviv, A. (1993). Differences of opinion make a horse race. *The Review of Financial Studies*, *6*(3), 473–506.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). … and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68.

Haugen, R. A., Baker, N. L., et al. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, *41*(3), 401–439.

Hirshleifer, D., Hou, K., Teoh, S. H., & Zhang, Y. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, *38*, 297–331.

Hong, H., & Stein, J. C. (2007). Disagreement and the stock market. *Journal of Economic perspectives*, *21*(2), 109–128.

Jacobs, H., & Hillert, A. (2015). Alphabetic bias, investor recognition, and trading behavior. *Review of Finance*, *20*(2), 693–723.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, *48*(1), 65–91.

Kandel, E., & Pearson, N. D. (1995). Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, *103*(4), 831–872.

Karpoff, J. M. (1986). A theory of trading volume. *The Journal of Finance*, *41*(5), 1069–1087.

Khan, M., & Watts, R. L. (2009). Estimation and empirical properties of a firm-year measure of accounting conservatism. *Journal of accounting and Economics*, *48*(2-3), 132–150.

Kruger, S. (2020). Disagreement and liquidity. *Available at SSRN 3189188*.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315–1335.

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance*, *49*(5), 1541–1578.

Linnainmaa, J. T., & Roberts, M. R. (2018). The history of the cross-section of stock returns. *The Review of Financial Studies*, *31*(7), 2606–2649.

Lo, A. W., & MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. *The Review of Financial Studies*, *3*(3), 431–467.

Lo, A. W., & Wang, J. (2010). Stock market trading volume. In *Handbook of financial econometrics: Applications* (pp. 241–342). Elsevier.

Lockwood, L., & Prombutr, W. (2010). Sustainable growth and stock returns. *Journal of Financial Research*, *33*(4), 519–538.

Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance*, *69*(4), 1643–1671.

Loughran, T., & McDonald, B. (2020). Measuring firm complexity. *Available at SSRN 3645372*.

Loughran, T., & Wellman, J. W. (2011). New evidence on the relation between the enterprise multiple and average stock returns. *Journal of Financial and Quantitative Analysis*, *46*(6), 1629–1650.

Lyandres, E., Sun, L., & Zhang, L. (2007). The new issues puzzle: Testing the investment-based explanation. *The Review of Financial Studies*, *21*(6), 2825–2855.

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, *71*(1), 5–32.

Milgrom, P., & Stokey, N. (1982). Information, trade and common knowledge. *Journal of Economic theory*, *26*(1), 17–27.

Novy-Marx, R. (2012). Is momentum really momentum? *Journal of Financial Economics*, *103*(3), 429–453.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, *108*(1), 1–28.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 109–131.

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

Pontiff, J., & Woodgate, A. (2008). Share issuance and cross-sectional returns. *The Journal of Finance*, *63*(2), 921–945.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review*, 289–315.

Soliman, M. T. (2008). The use of dupont analysis by market participants. *The Accounting Review*, *83*(3), 823–853.

Spiess, D. K., & Affleck-Graves, J. (1999). The long-run performance of stock returns following debt offerings. *Journal of Financial Economics*, *54*(1), 45–73.

Thomas, J. K., & Zhang, H. (2002). Inventory changes and future returns. *Review of Accounting Studies*, *7*(2-3), 163–187.

Titman, S., Wei, K. J., & Xie, F. (2004). Capital investments and stock returns. *Journal of Financial and Quantitative Analysis*, *39*(4), 677–700.

Varian, H. R. (1989). Differences of opinion in financial markets. In *Financial risk: Theory, evidence and implications* (pp. 3–37). Springer.

Xing, Y. (2007). Interpreting the value effect through the q-theory: An empirical investigation. *The Review of Financial Studies*, *21*(4), 1767–1795.

Appendix

A.1 Construction of Anomalies

All anomalies are constructed using monthly CRSP files and annual fundamentals from COMPUSTAT. I also provide the CRSP/COMPUSTAT variable name in the anomaly definitions. All anomalies are computed for each firm for every period (year-month). Anomalies constructed using only the annual fundamental data are repeated 11 times, i.e., they remain the same over a period of 12 months. If the predicted relationship of an anomaly is negative i.e., a higher value of anomaly predicts lower future returns, then I multiply the anomaly by -1 so that the relationship becomes positive. Subscript t represents the current time period. $\Delta x_t \equiv x_t - x_{t-1}$ and $\overline{x_t} \equiv \frac{x_t + x_{t-1}}{2}$. For momentum anomalies, $ret.\{b\}t\{a\} \equiv \prod_{t=-a}^{t=-b}(1+ret_t)-1$, where $a \leq b$ and ret_t is the return in month t. Book value of a firm's equity is defined as stockholder's equity plus deferred taxes minus preferred shares (Fama and French (1992)). $BE_t = seq_t + txditc_t - pstkrv_t$. If seq is not present, then $(ceq + pstk)$ is used. If either of ceq or $pstk$ is not present, then $(at - lt)$ is used. The market value of equity is the product of shares outstanding and share price: $ME_t = (prc_t/cfacpr_t) * (shrout_t * cfacshr_t)$. If a security's return is not available, then its delisting return is used from the CRSP monthly stock events file. Only firms traded on NYSE, AMEX or NASDAQ ($exchd \in 1,2,3$) having a share code ($shrcd$) of 10 or 11 are considered. Missing return and volume data in an otherwise continuous series are filled with zeros. Accounting data for a firm performing its operations in a year y is matched with trading data of June of year $y+1$ and carried forward 11 months, i.e., the same annual fundamental data is used from June of year $y+1$ to May of year $y+2$ (Fama and French (1992)).

Below is the definition of 36 anomalies used in this study:

1. **Gross Profitability** is constructed as revenues minus cost of goods sold scaled by total assets. $GrProf = \frac{revt_t - cogs_t}{at_t}$. Table 1 in Novy-Marx (2013) predicts a positive relationship between gross profitability and future returns.

2. **Operating Profitability** is revenues divided by the sum of cost of goods sold, interest expense and selling, general and admin (SGA) expenses. $OpProf = \frac{revt_t}{cogs_t + xint_t + xsga_t}$. Fama and French (2015) Table 1 Panel B finds a positive association between operating profitability and returns.

3. **Return on Assets** is defined as income before extraordinary items divided by total assets. $RoA = \frac{ib_t}{at_t}$. Table 3 in Haugen, Baker, et al. (1996) provides evidence of positive relationship between return on assets and returns.

4. **Return on Equity** is income before extraordinary items divided by book value of equity. $RoE = \frac{ib_t}{BE_t}$. Haugen et al. (1996) Table 3 documents a positive relation between RoE and returns.

5. **Profit Margin** is operating income after depreciation scaled by revenues. $PfMg = \frac{oiadp_t}{revt_t}$. Table 4 of Soliman (2008) finds a positive relationship of profit margin with returns.

6. **Change is Asset Turnover** is constructed as annual change in the ratio of revenues to assets. $ChgAssTurn = \Delta(\frac{revt_t}{at_t})$. Soliman (2008) Table 4 gives and evidence of positive relationship between change in asset turnover and returns.

7. **Accrual anomaly** predicts that stocks with lower accruals earn abnormally high returns in future (Table 6 and 7, Sloan (1996)). Accruals are defined as earnings minus the cash components of earnings divided by average assets. $Accr = \frac{(\Delta act_t - \Delta che_t) - (\Delta lct_t - \Delta dlc_t - \Delta txp_t) - dp_t}{at_t}$. Since the predicted relationship is negative, I transform the last equation by taking the negative of RHS as accrual anomaly.

8. **Net Operating Assets** is defined as the difference of operating assets and operating liabilities scaled by lagged total assets. $NOA = \frac{(at_t - che_t) - (at_t - dlc_t - dltt_t - BE_t)}{at_{t-1}}$. Hirshleifer, Hou, Teoh, and Zhang (2004) Table 5 finds a negative relationship of NOA with returns.

9. **Net Working Capital changes** (annual) are negatively associated with future returns as depicted in Table 7 of Soliman (2008). Net working capital is current assets minus current liabilities. $ChgNWC = \frac{\Delta(act_t - che_t) - \Delta(lct_t - dlc_t)}{at_t}$.

10. **Book to Market ratio** is book value equity divided by market value of equity. Table 6 of Fama and French (1992) finds that stocks with high book to market ratio earns higher future returns. $BTM = \frac{BE_t}{ME_t}$.

11. **Cash Flow to Price ratio** as the name suggests is the sum of earnings and depreciation (the cash flow of a firm) divided by its market equity. $CFP = \frac{ib_t + dp_t}{ME_t}$. Lakonishok, Shleifer, and Vishny (1994) Table 4 find a positive relation between future returns and CFP ratio.

12. **Earnings to Price ratio** is positively related to future returns (Table 4, Lakonishok et al. (1994)). $EP_t = \frac{ib_t}{ME_t}$. The anomaly originally appeared in Basu (1977).

13. **Enterprise Multiple** is defined as the ratio of enterprise value to operating cash flow. $EntMult = \frac{ME_t + dlc_t + dltt_t + pstkrv_t - che_t}{oibdp_t}$. Loughran and Wellman (2011) Table 2 depicts a negative relation between enterprise multiple and future returns.

14. **Sales to price ratio** is negatively associated with returns (Table 2, Barbee Jr, Mukherji, and Raines (1996)). $SP = \frac{revt_t}{ME_t}$.

15. **Short-term momentum** ($ret.6t2$) is the cumulative buy and hold return from $t-6$ to $t-2$ where t is the current month. Jegadeesh and Titman (1993) in Table 7 find a positive association between short-term momentum and future returns.

16. **Lagged Momentum** ($ret.12t7$) is cumulative buy and hold return from $t-12$ to $t-7$ where t is the current month. Table 1 Novy-Marx (2012) gives evidence of positive association with returns.

17. **Short-term reversal** ($ret.1t1$) is just the last month return. Table 7 of Jegadeesh and Titman (1993) finds a negative relation with future returns.

18. **Momentum reversal** ($ret.18t13$) is the annual buy and hold return starting 18 months prior. Table 7 of Jegadeesh and Titman (1993) finds a negative relation with future returns.

19. **Long-term reversal** ($ret.60t13$) is the buy and hold return from $t-60$ to $t-13$. De Bondt and Thaler (1985) document a negative association.

20. **Asset Growth** is defined as relative change in total asset compared to last year. $AssGr = \frac{\Delta at_t}{at_{t-1}}$. Cooper, Gulen, and Schill (2008) in Table III present evidence of negative relation between asset growth and future returns.

21. **Inventory Growth** is the ratio of change in inventory scaled by average assets. $ChgInvt = \frac{\Delta invt_t}{\overline{at_t}}$. Thomas and Zhang (2002) in Table 1 document a negative relationship with returns.

22. **Sustainable Growth** is defined as the relative change in book value of equity compared to last year. $SustGr = \frac{\Delta BE_t}{BE_{t-1}}$. Lockwood and Prombutr (2010) Table 6 finds a negative relation between sustainable growth and future returns.

23. **CAPX Growth** is the relative increase in capital expenditure compared to average expenditure two years before. $CapxGr = \frac{capx_t - \overline{capx_{t-1}}}{\overline{capx_{t-1}}}$. Table 2, Panel B in Abarbanell and Bushee (1998) depict a negative association with future returns.

24. **Growth in Sales minus inventory** is the ratio of relative growth in revenues scaled by relative growth in inventories. $SalesGr\_InvtGr = \frac{revt_t - \overline{rev_{t-1}}}{\overline{rev_{t-1}}} / \frac{invt_t - \overline{invt_{t-1}}}{\overline{invt_{t-1}}}$. Abarbanell and Bushee (1998) in Panel B of Table 2 find a positive association with future returns.

25. **Investment Growth** is the relative growth in capital expenditure. $InvstGr = \frac{\Delta capx_t}{capx_{t-1}}$. Xing (2007) Table 4 depicts negative relation with returns.

26. **Abnormal Capital Investment** is the growth in capital expenditure divided by revenues with respect to its previous three-year average. $AbCapInvst = \frac{capx_t / revt_t}{\frac{1}{3} \cdot \sum_{s=t-3}^{s=t-1} capx_s / rev_s}$. Titman, Wei, and Xie (2004) in Table 6, find it to be negatively related to future returns.

27. **Investment to Capital Ratio** is negatively associated with returns (Table 4, Xing (2007)). $IK = \frac{capx_t}{ppent_{t-1}}$.

28. **Investment to Asset Ratio** is negatively associated with returns (Section 4, Lyandres, Sun, and Zhang (2007)). $IA = \frac{\Delta ppent_t + \Delta invt_t}{at_{t-1}}$.

29. **Debt Issued Indicator** is one if a firm issues net debt in year and zero otherwise. $DebtIssueInd = 1_{\Delta dlc_t + \Delta dltt_t > 0}$. Spiess and Affleck-Graves (1999) find it to be negatively associated with future returns.

30. **Leverage** is defined as long term debt to book value of equity. $LEV = \frac{dltt_t}{ME_t}$. Bhandari (1988) in Table 2, Panel A find it to be positively related to future returns.

31. **One-year Share issuance** is the increase in number of shares outstanding with respect to last year. $ShIssue\_1 = \frac{shrout_t \cdot cfacshr_t}{shrout_{t-1} \cdot cfacshr_{t-1}}$. Table 3 in Pontiff and Woodgate (2008) depicts a negative relation with future returns.

32. **Five-year Share issuance** is the increase in number of shares outstanding with respect to shares outstanding five years ago. $ShIssue\_5 = \frac{shrout_t \cdot cfacshr_t}{shrout_{t-5} \cdot cfacshr_{t-5}}$. Table 3 in Daniel

and Titman (2006) depicts a negative relation with future returns. The anomaly originally appeared in Daniel and Titman (2006).

33. **Total External Finance** is the net financing raised in a year including both equity and debt. Bradshaw, Richardson, and Sloan (2006) in Table 5 find it to be negatively related to future returns. There are two versions of external finance measure:

    a. $ExtFin = \frac{(ShIssue\_1_t - 1) \cdot ME_t + \Delta dlc_t + \Delta dltt_t - dvc_t}{at_t}$
    b. $ExtFin2 = \frac{sstk_t - prstkc_t + dltis_t - dltr_t + dlcch_t - dv_t}{at_t}$

34. **O-Score** is a measure of distress by Ohlson (1980). It is the fitted value of a logistic regression. $O\_Score = -1.32 - 0.407 \cdot \log(\frac{at_t}{cpiind_t}) + 6.03 \cdot \frac{lt_t}{at_t} - 1.43 \cdot \frac{act_t - lct_t}{at_t} + 0.076 \cdot \frac{lct_t}{act_t} - 1.72 \cdot 1_{lt_t - at_t > 0} - 2.37 \cdot \frac{ib_t}{at_t} - 1.83 \cdot \frac{oiadp_t}{lt_t} + 0.285 \cdot 1_{ib_t < 0,\ ib_{t-1} < 0} - 0.521 \cdot \frac{\Delta ib_t}{|ib_t| + |ib_{t-1}|}$. Dichev (1998) in Table 4 find O-score to negatively predict future returns.

35. **Z-score** is a distress measure by Altman (1968). $Z\_Score = 1.2 \cdot \frac{act_t - lct_t}{at_t} + 1.4 \cdot \frac{re_t}{at_t} + 3.3 \cdot \frac{ni_t + xint_t + txp_t}{at_t} + 0.6 \cdot \frac{ME_t}{lt_t} + 1.0 \cdot \frac{revt_t}{at_t}$. Table 3 in Dichev (1998) finds a positive relation with returns.

A.2 Variable Definitions

The number of analysts following a firm ($NUMEST$) and dispersion in analyst forecast ($FDISP$) uses the I/B/E/S data available from Thomson Reuters. I use the EPS summary file for the US companies and restrict the sample to annual forecasts (having $fpi == 1$). To account for missing data due to analyst unfollowing a firm, both $NUMEST$ and $FDISP$ are repeated forward until forecast data is available. Earnings surprise ($ESURP$) and volatility of earnings ($EVOL$) are constructed using quarterly fundamentals from COMPUSTAT. Since earnings are reported once in three months, both $ESURP$ and $EVOL$ are repeated for two months to get a monthly measure.

The number of words, unique words, complex words, and document size is computed using EDGAR 10-K files. Bill McDonald has provided parsed EDGAR filings for the period 1994-2018. He has also compiled a summary file that directly gives the number of total words, unique words, and document size for each filing. For finding the number of unique occurrences of complex words, I search 374 complex words in the parsed 10-K files. I only consider 10-K, 10-K405, 10-KSB, and 10-KSB40 form types. The list of complex words is taken from Loughran and McDonald (2020). Variables derived from parsing 10-K files are merged with Compustat using CIK firm identifier and fiscal yearnends. Like annual fundamentals, EDGAR related variables from parsing 10-K of fiscal year y are merged with CRSP data from June of year $y+1$ to May of year $y+2$. Since 10-K are typically available only once in a calendar year, the variables are carried forward until a new filing is available.

Firm segments are computed from historical segments data. I only consider the most recent reported segment by letting report date of segments equal fiscal yearend date. I only consider business and operating segments and leave out geographic segments. For each firm and fiscal year, the number of entries gives the total number of segments. Additionally, the number of unique segments is when all segments within an industry are counted as one segment. To be included in the count, each segment must have non-missing sales data. Percent of shares held by institutional investors is available at 13-F Thomson Reuters Institutional (13-F) holdings data. Short interest is the number of shares, as a percentage of total shares outstanding, held short as of settlement date from Compustat. The variables computed from segments, ownership, and short interest databases are repeated forward until a new entry is found. This is done to match the frequency with monthly CRSP data.

A preceding $L\_$ before a variable name implies the logarithm of that variable. A trailing $\_D$ implies deterministic linear time detrending of the concerned variable[30]. Suffixing a variable with $\_R$ represents the cross-sectional ranking of that variable[31]. Lastly, a subscript t represents the current period (month), $t-1$ the previous period, and so on. For instance, ME_t is the current market value of equity, L_ME_{t-1} is the natural log of previous month's market equity, ME_D_{t-2} is the detrended market equity lagged by two periods, and $L\_ME\_R_t$ is the cross-sectional rank of current log market equity. Note that detrending and ranking are performed on the logarithmized variable and not the other way round. For finding ranks, its irrelevant whether we use the variable or its logarithm. At times it is convenient to work with quantiles of variables like portfolio sorts and dummy variable regressions. A trailing $\_Dec$ and $\_Ter$ represents deciles and terciles of the concerned variable respectively[32]. Table 16 gives the definitions of variables used in the regressions:

[Insert Table 16 here.]

[30]Detrending is performed by the following regression: $X_t = \alpha + \beta \cdot t + \epsilon_t$. X_D then equals ϵ_t

[31]At each month t the concerned variable is sorted and ranks are assigned. A smaller value gets a smaller rank. To ease comparison of different variables across different time periods, the ranks are scaled (separately for each month) to fall between 0 and 1.

[32]The boundaries for deciles are $10, 20, \ldots, 90$ percentiles while for terciles the boundaries are 30 and 70 percentiles. Both deciles and terciles are computed cross-sectionally for each time period.

Plots and Tables

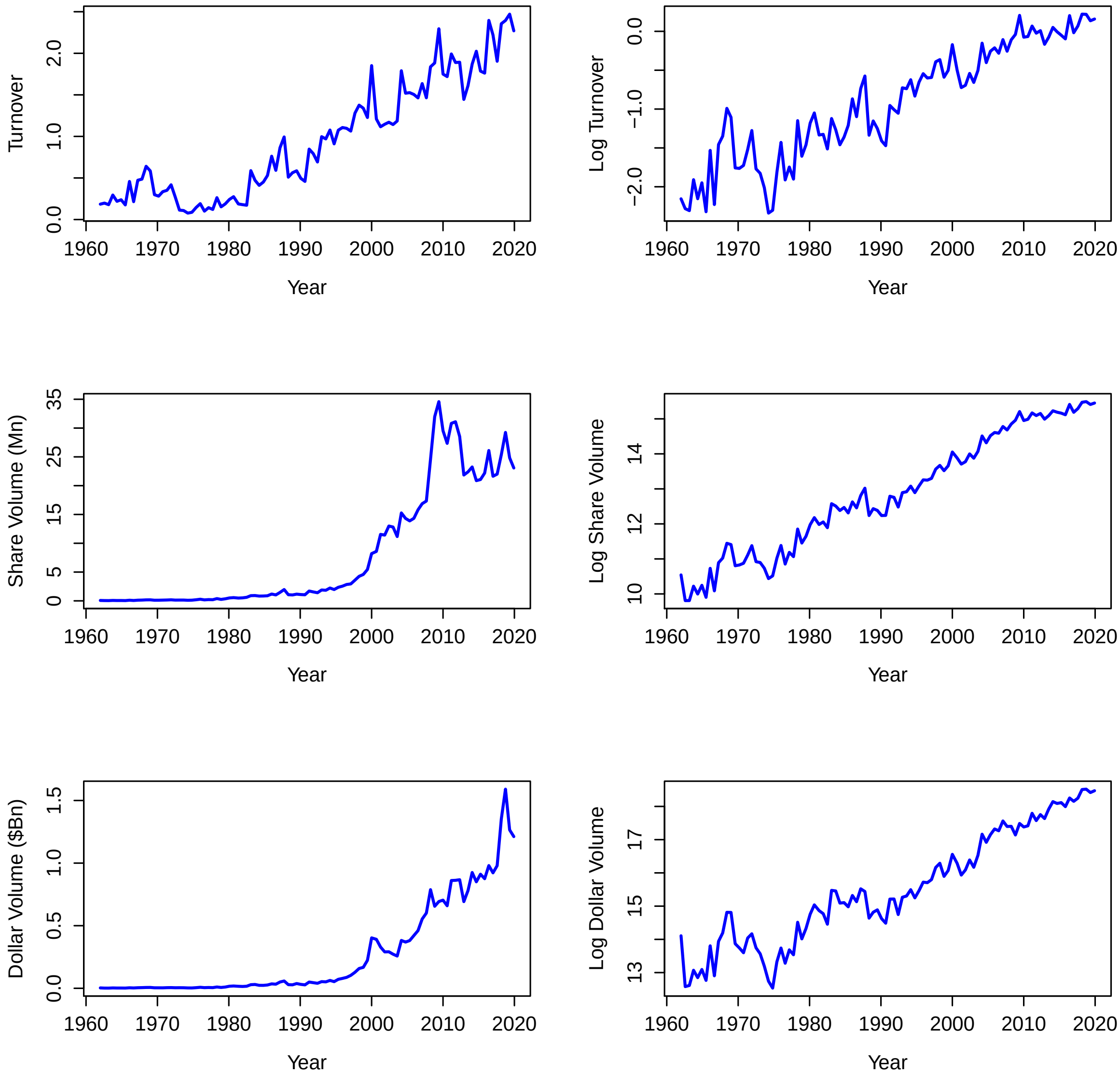

Figure 1: For each month, cross-sectional averages of three trading activity measures across all NYSE stocks are plotted. Averaging is done each month over the period 1962-2019. Turnover is the ratio of dollar volume to dollar market capitalization. Share volume is the number of shares traded. Top three plots use raw variables while bottom plots use logs of corresponding variables.

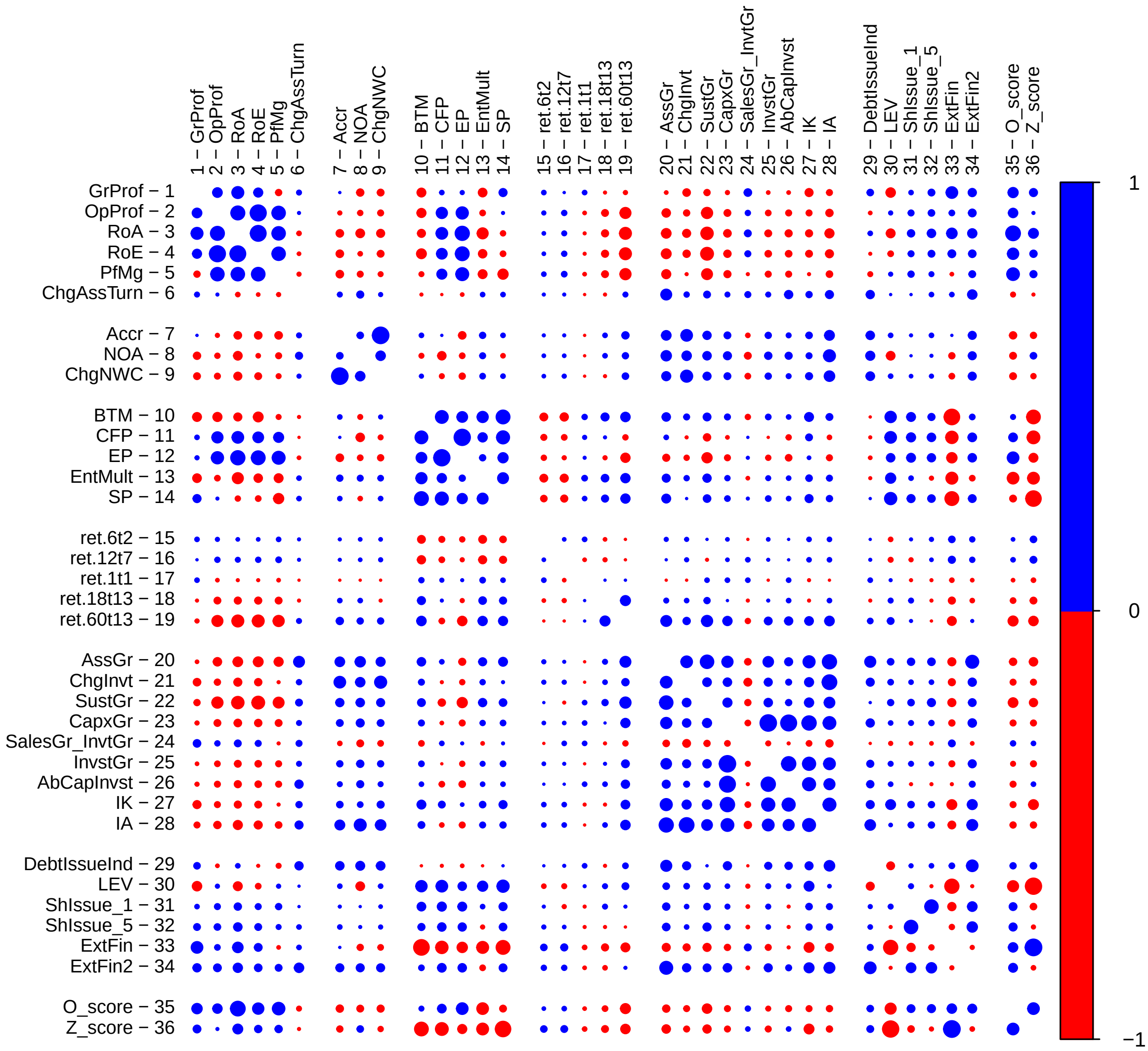

Figure 2: A matrix heat map of pairwise correlations among the 36 anomaly signals. Blue circles represent positive correlation while red circles are negative correlations. A bigger circle represents higher magnitude of correlations. Lower half is symmetric to the upper half. All correlations are calculated for the entire sample.

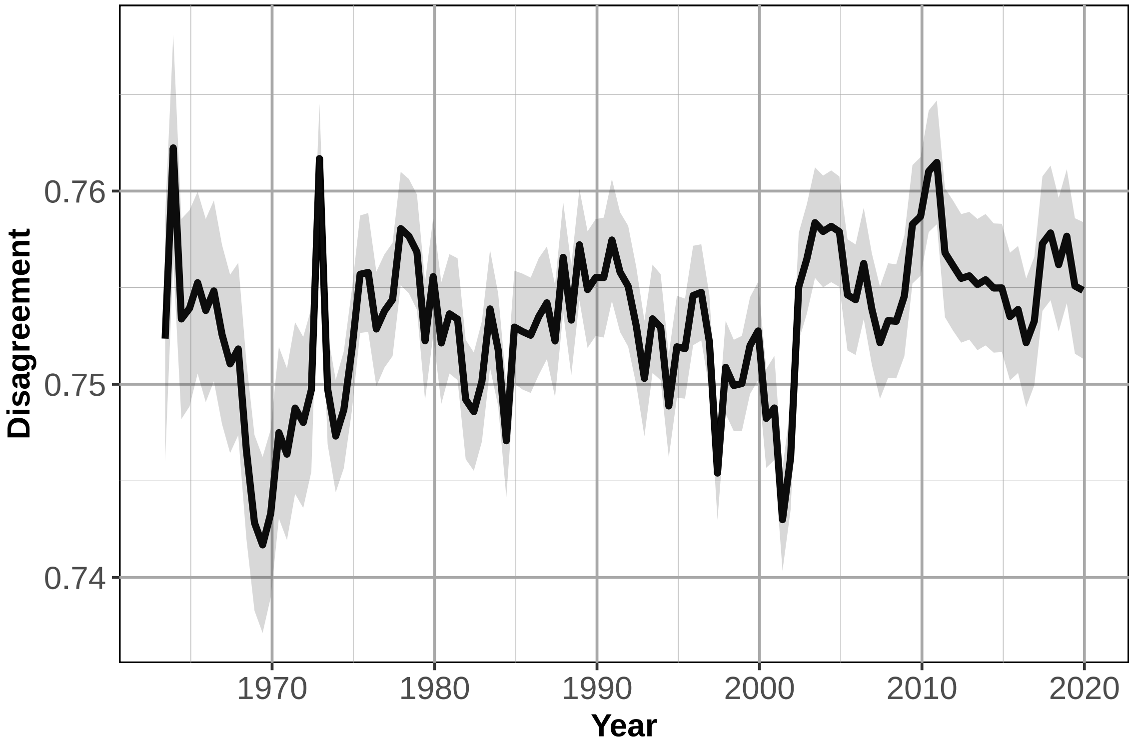

Figure 3: Monthly cross-sectional mean and standard error of the disagreement measure from year 1966. The confidence interval is set to two standard errors.

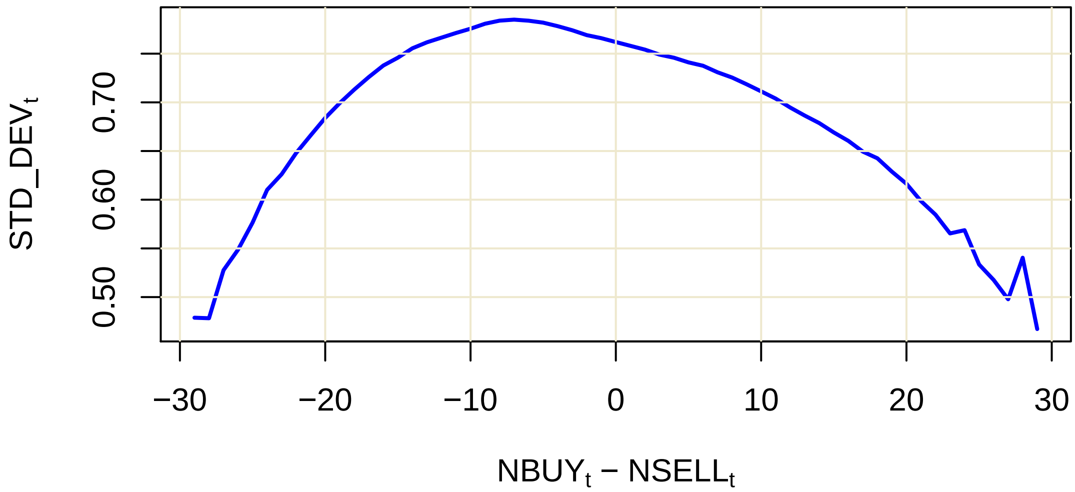

Figure 4: Disagreement averaged over entore NYSE stocks (1962-2019) against the difference between number of buy and sell signals.

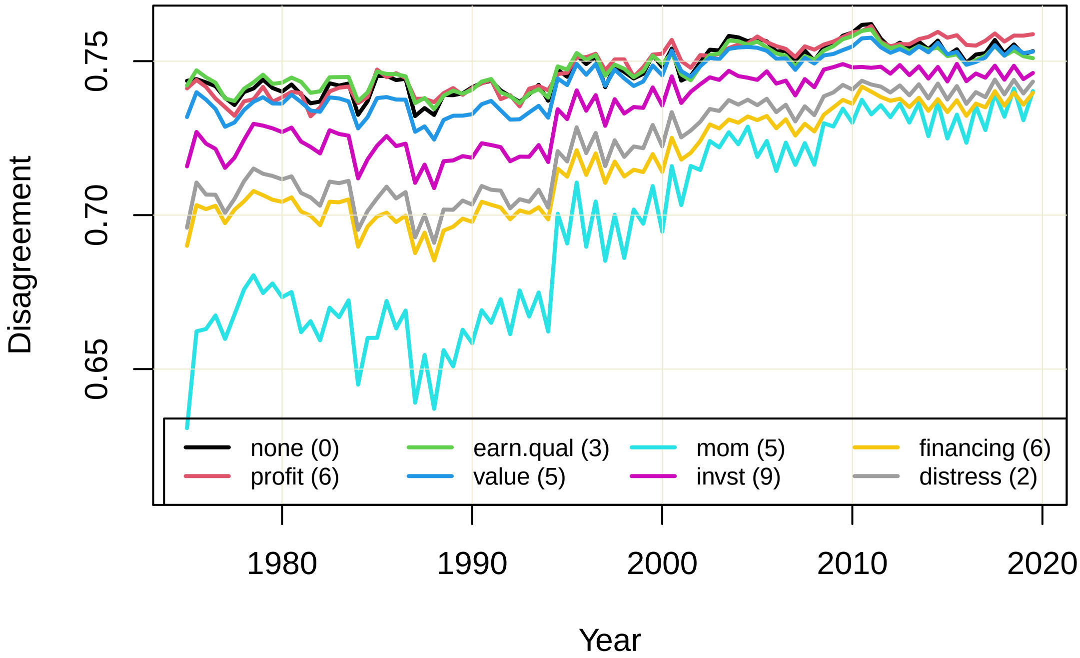

Figure 5: Average (semi-annual) Disagreement, since 1975, using all but one group of anomalies (see Table 1) over time. The requirement of having atleast 10 signals to construct disagreement is relaxed for various disagreement measures in this figure. Disagreement is the standard deviation of trading signals generated from anomalies in equation 3.2.

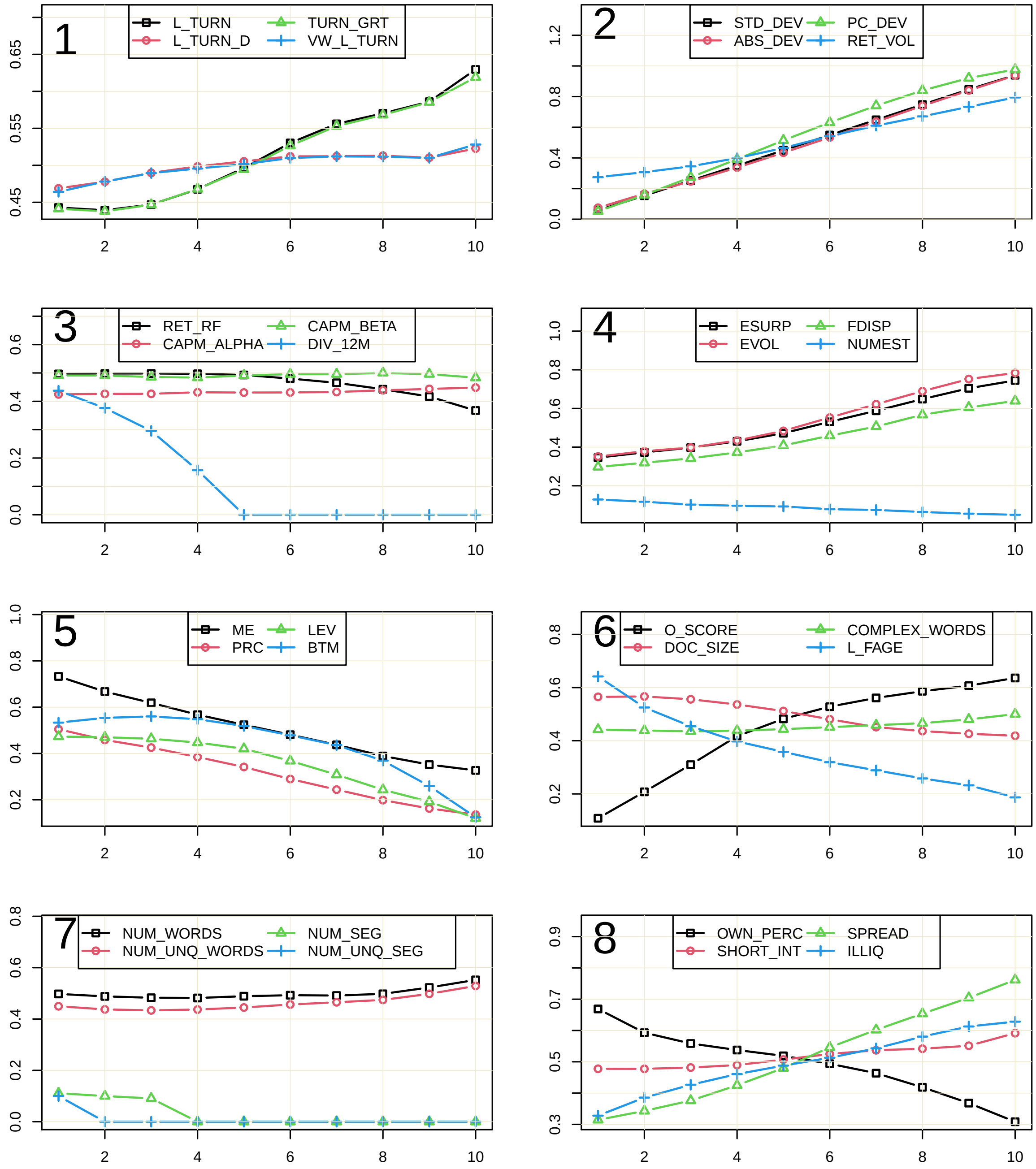

Figure 6: Cross-sectional ranks of firm characteristics across disagreement deciles. Ranks appear on vertical axis and disagreement decile on horizontal axis. Both ranks and disagreement deciles are constructed each month. Out of 5 subplots, first 3 have ranks of firm characteristics, next is the ranks of different disagreement measure and the last subplot has different turnover measures. Construction details of all variables appear in Appendix A.2

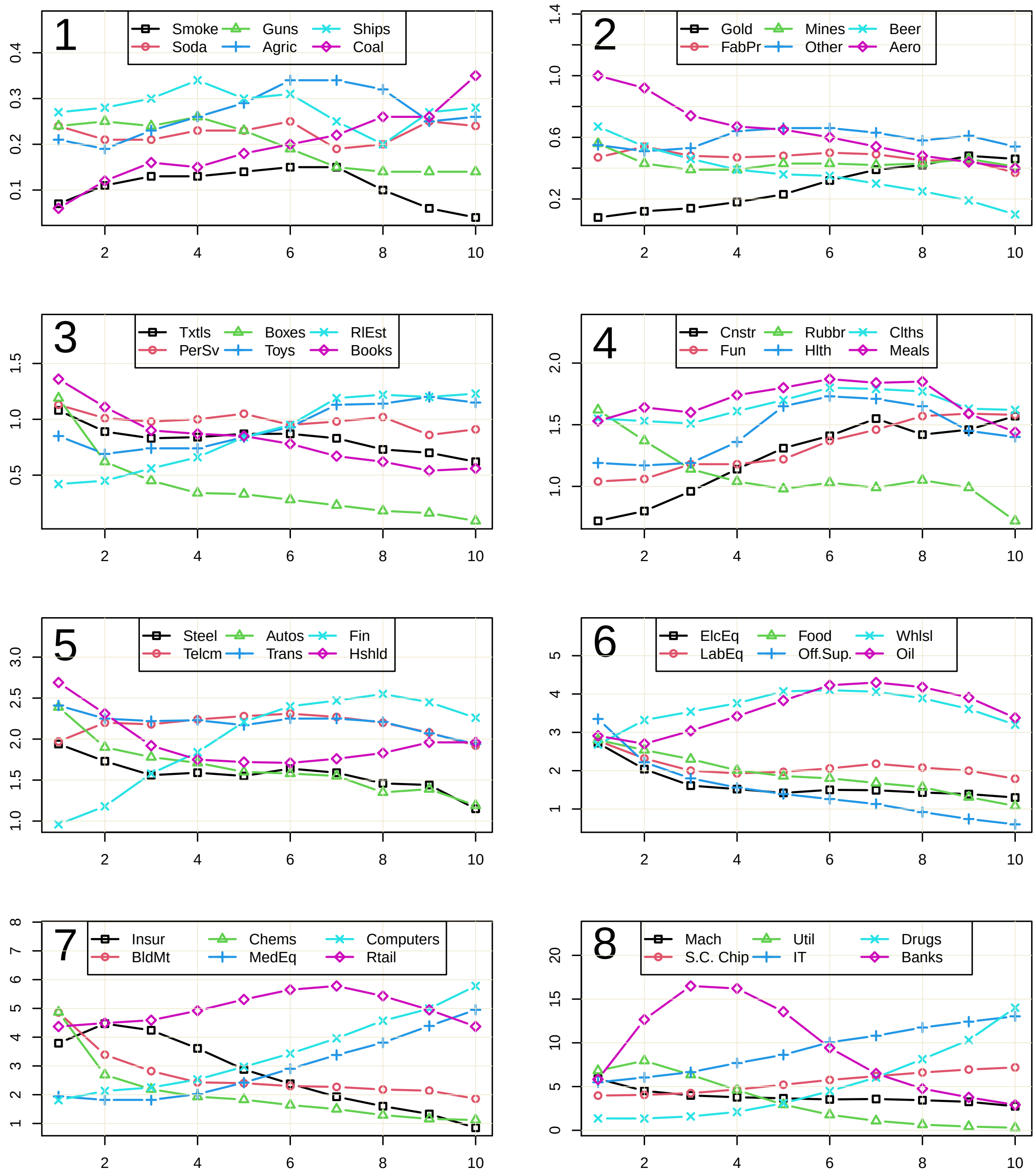

Figure 7: Relative industry concentration across disagreement deciles. For each disagreement decile (horizontal axis) the percentage of firm-month observations belonging to a particular industry appear on vertical axis. Both percentage of firm-month observations and disagreement deciles are calculated cross-sectionally for each month. 48 Fama and French (1997) industries are presented in 8 subplots of 6 industries each. Industries are sorted on their relative concentration (vertical axis) and then the smallest group is presented in subplot 1 and the largest in subplot 8.

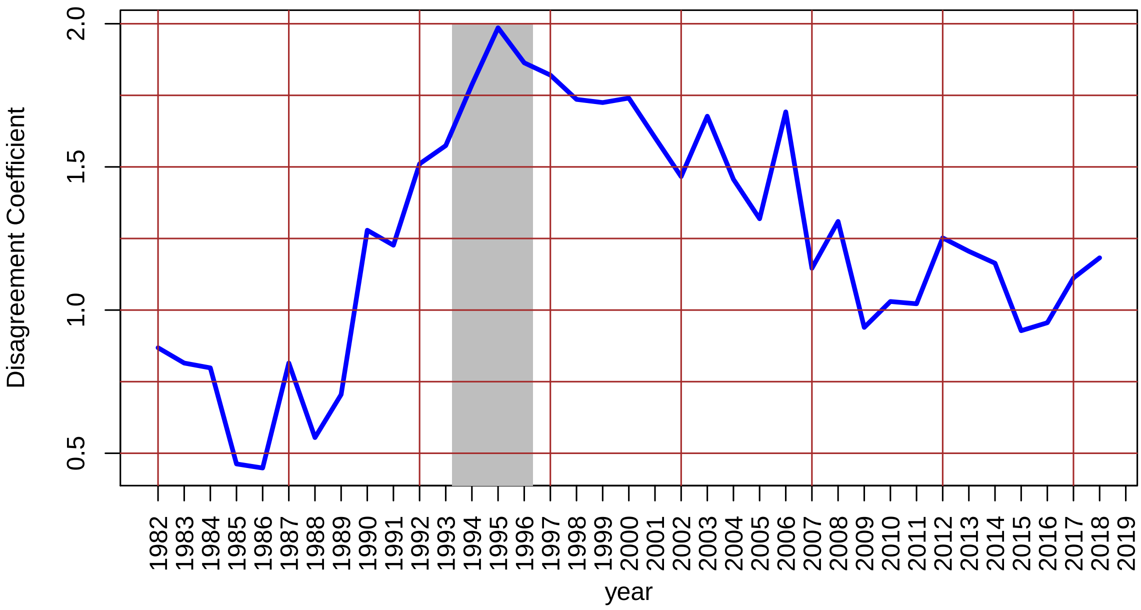

Figure 8: Disagreement coefficient from annual regressions. The EDGAR implementation period (April 1993 to May 1996) is shown in shaded region.

Table 1: **Anomalies List**

Category / S.No.	Anomaly	Predicted Relationship	Source
Profitability			
1	Gross Profitability	Positive	Novy-Marx (2013)
2	Operating Profitability	Positive	Fama and French (2015)
3	Return on Assets	Positive	Haugen et al. (1996)
4	Return on Equity	Positive	Haugen et al. (1996)
5	Profit Margin	Positive	Soliman (2008)
6	Change in Asset Turnover	Positive	Soliman (2008)
Earnings Quality			
7	Accruals	Negative	Sloan (1996)
8	Net Operating Assets	Negative	Hirshleifer et al. (2004)
9	Changes in Net Working Capital	Negative	Soliman (2008)
Valuation			
10	Book to market	Positive	Fama and French (1992)
11	Cash flow to price	Positive	Lakonishok et al. (1994)
12	Earnings to Price	Positive	Basu (1977)
13	Enterprise Multiple	Negative	Loughran and Wellman (2011)
14	Sales to price	Positive	Barbee Jr et al. (1996)
Momentum			
15	Short term momentum	Positive	Jegadeesh and Titman (1993)
16	Lagged Momentum	Positive	Novy-Marx (2012)
17	Short-term reversal	Negative	Jegadeesh and Titman (1993)
18	Medium-term reversal	Negative	Jegadeesh and Titman (1993)
19	Long-term reversal	Negative	De Bondt and Thaler (1985)
Investment			
20	Asset Growth	Negative	Cooper et al. (2008)
21	Inventory Growth	Negative	Thomas and Zhang (2002)
22	Sustainable Growth	Negative	Lockwood and Prombutr (2010)
23	CAPX Growth	Negative	Abarbanell and Bushee (1998)
24	Growth in Sales minus growth in Inventory	Positive	Abarbanell and Bushee (1998)
25	Investment growth	Negative	Xing (2007)
26	Abnormal CAPX	Negative	Titman et al. (2004)
27	Investment to Capital Ratio	Negative	Xing (2007)
28	Investment to Asset Ratio	Negative	Lyandres et al. (2007)
Financing			
29	Increase in Debt Issuance	Negative	Spiess and Affleck-Graves (1999)
30	Leverage	Positive	Bhandari (1988)
31	One year Share Issuance	Negative	Pontiff and Woodgate (2008)
32	Five year Share Issuance	Negative	Daniel and Titman (2006)
33	External Financing – I	Negative	Bradshaw et al. (2006)
34	External Financing – II	Negative	Bradshaw et al. (2006)
Distress			
35	O-Score	Negative	Dichev (1998)
36	Z-Score	Positive	Dichev (1998)

List of 36 anomalies used to construct the disagreement measure. 31 anomalies are from Linnainmaa and Roberts (2018) and 5 momentum anomalies from McLean and Pontiff (2016). Predicted relationship is from the original study findings.

Table 2: **Correlations and Descriptive Statistics**

Panel A: Correlations

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)	$NASDAQ$		0.01	−0.05	−0.23	$0.00^{\#}$	−0.09	−0.30	−0.48	−0.41	0.03	$0.00^{\#}$	−0.28	0.06	0.20
(2)	RET^{+}	0.04		0.53	−0.05	$-0.00^{\#}$	−0.10	0.04	−0.03	0.02	0.06	0.04	−0.03	0.03	0.07
(3)	RET^{-}	−0.06	0.26		$-0.00^{\#}$	$0.00^{\#}$	−0.04	0.21	0.08	0.14	−0.08	−0.09	0.06	−0.11	−0.17
(4)	LEV	−0.07	−0.01	−0.05		0.01	0.41	−0.07	0.16	−0.01	0.18	0.28	0.01	0.11	−0.09
(5)	$CAPM\_\beta$	$-0.00^{\#}$	−0.01	$0.00^{\#}$	$-0.00^{\#}$		0.01	$0.00^{\#}$	0.01	$0.00^{\#}$	$-0.00^{\#}$	$-0.00^{\#}$	0.01	−0.02	−0.01
(6)	BTM	−0.04	−0.04	−0.04	0.38	0.00		−0.21	0.17	−0.25	0.25	0.36	−0.14	0.21	−0.17
(7)	L_PRC	−0.28	−0.04	0.24	−0.11	$0.00^{\#}$	−0.12		0.30	0.77	−0.43	−0.55	0.48	−0.36	−0.35
(8)	L_FAGE	−0.38	−0.04	0.08	0.03	0.01	0.08	0.21		0.33	−0.01	−0.01	0.26	−0.01	−0.27
(9)	L_ME	−0.42	−0.05	0.16	−0.09	$0.01^{\#}$	−0.16	0.76	0.26		−0.11	−0.12	0.75	−0.11	−0.30
(10)	$ESURP$	−0.00	0.03	−0.04	0.12	−0.02	0.05	−0.15	−0.07	−0.38		0.67	−0.17	0.38	0.28
(11)	$EVOL$	−0.01	0.03	−0.03	0.14	−0.02	0.02	−0.16	−0.04	−0.47	0.61		−0.21	0.42	0.31
(12)	$NUMEST$	−0.28	−0.06	0.06	−0.04	0.01	−0.10	0.45	0.29	0.65	−0.06	−0.06		−0.14	−0.13
(13)	$FDISP$	0.02	0.02	−0.05	0.05	$-0.00^{\#}$	0.10	−0.16	−0.08	−0.27	0.10	0.11	−0.06		0.20
(14)	STD_DEV	0.21	0.10	−0.18	0.10	−0.01	0.00	−0.37	−0.26	−0.30	0.13	0.13	−0.13	0.09	

Panel B: Descriptive Statistics

	Mean	**SD**	**Min**	**p25**	**Median**	**p75**	**Max**	**IQR***	**Range***	**Skew**	**Kurt**
$NASDAQ$	0.48	0.24	0.00	0.50	0.59	0.63	0.68	0.55	2.87	-1.45	3.32
RET^{+}	0.06	0.03	0.00	0.03	0.05	0.07	0.28	1.14	8.14	1.59	8.20
RET^{-}	-0.04	0.03	-0.27	-0.05	-0.03	-0.02	0.00	0.96	9.30	-2.45	13.33
LEV	0.06	0.02	0.03	0.04	0.06	0.07	0.16	1.35	5.91	1.05	4.27
$CAPM\_\beta$	1.28	0.47	0.20	1.03	1.25	1.47	2.60	0.94	5.09	0.45	3.28
BTM	0.09	0.03	0.05	0.06	0.08	0.10	0.25	1.09	6.52	1.53	6.22
L_PRC	2.55	0.33	1.71	2.32	2.48	2.81	3.63	1.52	5.90	0.46	2.92
L_FAGE	4.27	0.65	0.00	4.13	4.32	4.72	5.00	0.91	7.73	-2.58	13.29
L_ME	11.56	1.17	9.43	10.66	11.22	12.68	13.74	1.72	3.68	0.33	1.91
$ESURP$	0.14	0.48	0.01	0.03	0.04	0.09	5.00	0.12	10.34	8.39	76.67
$EVOL$	0.14	0.25	0.01	0.03	0.05	0.14	1.80	0.41	7.31	4.18	22.87
$NUMEST$	7.97	0.98	6.41	7.31	7.65	8.92	10.01	1.64	3.67	0.35	1.82
$FDISP$	0.20	0.06	0.06	0.16	0.20	0.25	0.50	1.41	6.76	0.71	4.35
STD_DEV	0.75	0.00	0.74	0.75	0.75	0.76	0.76	1.17	6.11	-0.63	3.72

Panel A reports the time-series average of cross-sectional correlation coefficients. Correlations marked with ($^{\#}$) are not significant at the 5 % level. At each month, cross-sectional correlation among variables is computed and then their time-series average over the duration of sample is reported. Lower triangle represents variable correlation while the upper triangle consists of rank correlations. Panel B presents descriptive Statistics of explanatory variables. Interquartile range (IQR^{*}) and variable range ($Range^{*}$) are in multiples of standard deviation. At each month the average of all the variables is computed and their time-series descriptive statistics are reported. Variable definitions are present in Appendix A.2

Table 3: **Monthly cross-sectional regression: different specifications**

	L_TURN_t					
	(1)	(2)	(3)	(4)	(5)	(6)
$NASDAQ$	0.130***	0.128***	0.058**	0.109***	0.107***	0.037
	(0.024)	(0.024)	(0.026)	(0.024)	(0.024)	(0.025)
RET^+_{t-1}	1.722***	1.801***	1.667***	1.539***	1.584***	1.470***
	(0.069)	(0.068)	(0.066)	(0.062)	(0.061)	(0.058)
RET^-_{t-1}	-2.569***	-2.575***	-2.734***	-2.318***	-2.316***	-2.462***
	(0.091)	(0.093)	(0.092)	(0.084)	(0.085)	(0.085)
LEV_{t-1}	0.787***	0.928***	0.853***	0.597***	0.676***	0.648***
	(0.095)	(0.094)	(0.100)	(0.092)	(0.091)	(0.097)
$CAPM\_\beta_{t-1}$	0.078**	0.078**	0.078**	0.078**	0.078**	0.079**
	(0.038)	(0.038)	(0.038)	(0.036)	(0.036)	(0.036)
BTM_{t-1}	-0.936***	-0.781***	-1.004***	-0.631***	-0.517***	-0.684***
	(0.167)	(0.166)	(0.176)	(0.167)	(0.163)	(0.175)
L_PRC_{t-1}	0.176***	0.163***	0.297***	0.205***	0.198***	0.327***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)
L_FAGE_{t-1}	-0.145***	-0.146***	-0.119***	-0.114***	-0.114***	-0.087***
	(0.014)	(0.015)	(0.015)	(0.014)	(0.014)	(0.015)
$ESURP_{t-1}$	0.504***		0.594***	0.440***		0.519***
	(0.071)		(0.076)	(0.061)		(0.064)
$EVOL_{t-1}$	0.371***		0.448***	0.134		0.188**
	(0.091)		(0.098)	(0.084)		(0.087)
$NUMEST_{t-1}$	0.033***	0.034***		0.033***	0.033***	
	(0.002)	(0.002)		(0.002)	(0.002)	
$FDISP_{t-1}$	0.041***	0.047***		0.022***	0.026***	
	(0.008)	(0.008)		(0.008)	(0.008)	
STD_DEV_{t-1}				1.464***	1.507***	1.559***
				(0.065)	(0.065)	(0.070)
Within R^2	0.165	0.161	0.107	0.185	0.184	0.131
Adj. R^2	0.424	0.422	0.384	0.439	0.438	0.401
% R^2 Explained				5.15	5.35	5.86
Observations	672,385	672,385	672,385	672,385	672,385	672,385

Log turnover regressed on different set of explanatory variables. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table 4: **Monthly cross-sectional regression: daily, weekly and monthly turnover**

	$L\_TURN\_1d_t$	$L\_TURN\_5d_t$	L_TURN_t	$L\_TURN\_1d_t$	$L\_TURN\_5d_t$	L_TURN_t
	(1)	(2)	(3)	(4)	(5)	(6)
$NASDAQ$	0.079***	0.114***	0.138***	0.057**	0.092***	0.116***
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
RET^+_{t-1}	2.558***	2.203***	1.685***	2.379***	2.021***	1.503***
	(0.101)	(0.082)	(0.068)	(0.094)	(0.075)	(0.061)
RET^-_{t-1}	-3.562***	-3.111***	-2.525***	-3.316***	-2.860***	-2.273***
	(0.105)	(0.092)	(0.090)	(0.100)	(0.086)	(0.083)
LEV_{t-1}	0.769***	0.783***	0.798***	0.581***	0.591***	0.606***
	(0.100)	(0.097)	(0.094)	(0.097)	(0.094)	(0.091)
$CAPM\_\beta_{t-1}$	0.112*	0.086**	0.078**	0.112*	0.086**	0.078**
	(0.060)	(0.042)	(0.038)	(0.059)	(0.041)	(0.036)
BTM_{t-1}	-1.124***	-0.959***	-0.855***	-0.822***	-0.650***	-0.546***
	(0.176)	(0.169)	(0.162)	(0.175)	(0.166)	(0.159)
L_PRC_{t-1}	0.235***	0.200***	0.174***	0.264***	0.230***	0.204***
	(0.014)	(0.014)	(0.013)	(0.014)	(0.014)	(0.013)
L_FAGE_{t-1}	-0.115***	-0.141***	-0.148***	-0.085***	-0.110***	-0.117***
	(0.015)	(0.015)	(0.014)	(0.015)	(0.014)	(0.014)
$ESURP_{t-1}$	0.466***	0.522***	0.490***	0.403***	0.457***	0.426***
	(0.085)	(0.079)	(0.071)	(0.076)	(0.068)	(0.061)
$EVOL_{t-1}$	0.299***	0.292***	0.365***	0.067	0.054	0.127
	(0.097)	(0.095)	(0.090)	(0.091)	(0.089)	(0.082)
$NUMEST_{t-1}$	0.037***	0.034***	0.033***	0.036***	0.034***	0.032***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$FDISP_{t-1}$	0.032***	0.035***	0.040***	0.014*	0.016**	0.021***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)
STD_DEV_{t-1}				1.432***	1.466***	1.466***
				(0.068)	(0.066)	(0.065)
Within R^2	0.154	0.167	0.163	0.166	0.184	0.184
Adj. R^2	0.394	0.408	0.421	0.403	0.420	0.436
$\% R^2$ Explained				3.48	4.52	5.24
Observations	665,705	665,705	665,705	665,705	665,705	665,705

Log turnover, measured over daily, weekly and monthly intervals, is regressed on a set of explanatory variables and the disagreement measure STD_DEV. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table 5: **Monthly cross-sectional regression: different measures of turnover**

	L_TURN_t	ΔL_TURN_t	$L\_TURN\_GRT_t$	$L\_TURN\_D_t$	$L\_TURN\_ILLIQ_t$	$VW\_L\_TURN_t$	$EW\_L\_TURN_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$NASDAQ$	0.111***	0.002	0.140***	0.047***	0.001	0.068***	0.047***
	(0.024)	(0.002)	(0.022)	(0.008)	(0.009)	(0.009)	(0.010)
RET^+_{t-1}	1.536***	-0.774***	0.749***	0.948***	0.635***	0.968***	0.925***
	(0.062)	(0.043)	(0.041)	(0.046)	(0.030)	(0.048)	(0.043)
RET^-_{t-1}	-2.316***	0.735***	-1.041***	-1.412***	-1.402***	-1.446***	-1.579***
	(0.084)	(0.065)	(0.075)	(0.067)	(0.050)	(0.059)	(0.052)
LEV_{t-1}	0.601***	0.028***	0.499***	0.094**	0.290***	0.086**	0.129***
	(0.092)	(0.008)	(0.088)	(0.038)	(0.045)	(0.039)	(0.044)
$CAPM\_\beta_{t-1}$	0.078**	0.037	0.002	0.058*	0.033	0.049*	0.033
	(0.036)	(0.047)	(0.038)	(0.031)	(0.039)	(0.029)	(0.025)
BTM_{t-1}	-0.635***	0.015	0.077	-0.535***	0.521***	-0.545***	-0.511***
	(0.167)	(0.027)	(0.144)	(0.078)	(0.060)	(0.077)	(0.081)
L_PRC_{t-1}	0.206***	-0.010***	0.099***	0.112***	0.002	0.109***	0.120***
	(0.013)	(0.003)	(0.011)	(0.008)	(0.005)	(0.007)	(0.008)
L_FAGE_{t-1}	-0.115***	-0.006***	-0.465***	-0.001	0.001	0.020***	0.031***
	(0.014)	(0.001)	(0.016)	(0.005)	(0.005)	(0.006)	(0.006)
$ESURP_{t-1}$	0.441***	0.045**	0.145***	0.240***	0.227***	0.183***	0.221***
	(0.061)	(0.019)	(0.050)	(0.030)	(0.033)	(0.034)	(0.032)
$EVOL_{t-1}$	0.132	0.070***	0.463***	-0.217***	0.193***	-0.126**	-0.176***
	(0.084)	(0.027)	(0.085)	(0.066)	(0.048)	(0.056)	(0.061)
$NUMEST_{t-1}$	0.033***	-0.000	0.032***	-0.002***	0.003***	-0.003***	-0.001
	(0.002)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
$FDISP_{t-1}$	0.023***	-0.003**	0.032***	-0.011**	0.008**	-0.014***	-0.012***
	(0.008)	(0.001)	(0.008)	(0.004)	(0.003)	(0.004)	(0.004)
STD_DEV_{t-1}	1.464***	0.018**	0.833***	0.450***	0.212***	0.509***	0.498***
	(0.066)	(0.008)	(0.062)	(0.030)	(0.025)	(0.032)	(0.033)
Within R^2	0.186	0.017	0.118	0.052	0.062	0.051	0.055
Adj. R^2	0.438	0.018	0.267	0.110	0.140	0.091	0.095
$\%R^2$ Explained	5.15	0.00	1.61	0.43	-0.03	0.63	0.54
Observations	670,922	670,922	670,922	670,922	670,922	670,922	670,922

Several measures of tunover are regressed on a set of explanatory variables and the disagreement measure STD_DEV. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. $\% R^2$ explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table 6: **Monthly cross-sectional regression: different measures of disagreement**

	L_TURN_t					
	(1)	(2)	(3)	(4)	(5)	(6)
STD_DEV_{t-1}	1.347***					
	(0.075)					
ABS_DEV_{t-1}		1.107***				
		(0.061)				
PC_DEV_{t-1}			1.411***			
			(0.076)			
$CONT\_DEV_{t-1}$				3.482***		
				(0.191)		
$LO\_CORR\_SD_{t-1}$					0.730***	
					(0.054)	
$HI\_CORR\_SD_{t-1}$					0.508***	
					(0.040)	
NUM_FLIPS_{t-1}						0.008***
						(0.002)
NUM_DIV_{t-1}						0.001**
						(0.000)
Within R^2	0.179	0.178	0.179	0.183	0.177	0.159
Adj. R^2	0.450	0.449	0.450	0.453	0.449	0.437
$\%R^2$ Explained	5.10	4.85	5.10	6.52	4.71	0.08
Observations	434,661	434,661	434,661	434,661	434,661	434,661

Log turnover regressed on set of controls and several measures of disagreement. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table 7: **Monthly cross-sectional regression: rank regression**

	L_TURN_t		$L\_TURN\_R_t$	
	(1)	(2)	(3)	(4)
$NASDAQ$	0.135***	0.126***	0.033***	0.031***
	(0.024)	(0.024)	(0.006)	(0.006)
$RET^{+}\_R_{t-1}$	0.472***	0.424***	0.124***	0.112***
	(0.011)	(0.011)	(0.003)	(0.003)
$RET^{-}\_R_{t-1}$	-0.601***	-0.539***	-0.165***	-0.148***
	(0.014)	(0.013)	(0.003)	(0.003)
LEV_R_{t-1}	0.215***	0.224***	0.064***	0.066***
	(0.030)	(0.030)	(0.008)	(0.008)
$CAPM\_\beta\_R_{t-1}$	0.122***	0.125***	0.001	0.002
	(0.040)	(0.040)	(0.007)	(0.007)
BTM_R_{t-1}	-0.479***	-0.353***	-0.128***	-0.095***
	(0.036)	(0.037)	(0.009)	(0.010)
$L\_PRC\_R_{t-1}$	0.724***	0.791***	0.192***	0.210***
	(0.047)	(0.047)	(0.012)	(0.012)
$L\_FAGE\_R_{t-1}$	-0.339***	-0.290***	-0.099***	-0.085***
	(0.035)	(0.035)	(0.010)	(0.010)
$ESURP\_R_{t-1}$	0.278***	0.248***	0.076***	0.068***
	(0.013)	(0.013)	(0.004)	(0.003)
$EVOL\_R_{t-1}$	0.316***	0.225***	0.091***	0.067***
	(0.029)	(0.030)	(0.008)	(0.008)
$NUMEST\_R_{t-1}$	1.304***	1.287***	0.366***	0.362***
	(0.058)	(0.058)	(0.016)	(0.016)
$FDISP\_R_{t-1}$	0.186***	0.153***	0.053***	0.044***
	(0.024)	(0.024)	(0.006)	(0.006)
$STD\_DEV\_R_{t-1}$		0.485***		0.130***
		(0.028)		(0.007)
Within R^2	0.183	0.195	0.189	0.200
Adj. R^2	0.437	0.445	0.296	0.306
% R^2 Explained		3.47		3.15
Observations	672,385	672,385	672,385	672,385

Cross-sectional rank of log turnover regressed on cross-sectional ranks of set of controls and several measures of disagreement. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table 8: **Monthly cross-sectional regression: size terciles**

	L_TURN_t					
	SMALL		*MEDIUM*		*BIG*	
	(1)	(2)	(3)	(4)	(5)	(6)
$NASDAQ$	0.092***	0.088**	0.198***	0.174***	0.335***	0.304***
	(0.035)	(0.035)	(0.026)	(0.025)	(0.039)	(0.038)
RET^+_{t-1}	1.608***	1.483***	1.721***	1.494***	1.687***	1.522***
	(0.081)	(0.075)	(0.077)	(0.069)	(0.081)	(0.074)
RET^-_{t-1}	-2.242***	-2.071***	-2.402***	-2.079***	-2.546***	-2.324***
	(0.077)	(0.076)	(0.112)	(0.101)	(0.141)	(0.132)
LEV_{t-1}	0.519***	0.402***	0.810***	0.498***	1.169***	0.961***
	(0.099)	(0.099)	(0.149)	(0.140)	(0.252)	(0.224)
$CAPM\_\beta_{t-1}$	0.027	0.026	0.084**	0.085**	0.028	0.026
	(0.046)	(0.043)	(0.042)	(0.039)	(0.042)	(0.041)
BTM_{t-1}	-0.716***	-0.553***	-0.339	0.158	0.369	0.877***
	(0.182)	(0.188)	(0.264)	(0.246)	(0.308)	(0.300)
L_PRC_{t-1}	0.201***	0.249***	0.058***	0.084***	-0.007	-0.006
	(0.018)	(0.019)	(0.018)	(0.018)	(0.020)	(0.020)
L_FAGE_{t-1}	-0.078***	-0.063***	-0.093***	-0.057***	-0.191***	-0.159***
	(0.022)	(0.022)	(0.017)	(0.016)	(0.022)	(0.021)
$ESURP_{t-1}$	0.439***	0.401***	0.505***	0.409***	0.629***	0.563***
	(0.065)	(0.059)	(0.115)	(0.093)	(0.141)	(0.133)
$EVOL_{t-1}$	0.526***	0.358***	0.855***	0.508***	0.796**	0.416
	(0.089)	(0.084)	(0.162)	(0.129)	(0.345)	(0.319)
$NUMEST_{t-1}$	0.127***	0.126***	0.059***	0.058***	0.014***	0.013***
	(0.005)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)
$FDISP_{t-1}$	0.030***	0.020**	0.044***	0.016*	0.080***	0.060***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.016)	(0.015)
STD_DEV_{t-1}		1.439***		1.748***		1.057***
		(0.105)		(0.076)		(0.099)
Within R^2	0.157	0.173	0.185	0.219	0.163	0.181
Adj. R^2	0.335	0.348	0.534	0.554	0.602	0.611
$\%R^2$ Explained		1.84		4.27		2.16
Observations	182,404	182,404	298,014	298,014	191,967	191,967

Log turnover regressed on set of controls and several measures of disagreement across three size terciles based on 70/30 NYSE breakpoints. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table 9: **Monthly cross-sectional regression: 5-year subperiods**

	L_TURN_t								
	1975-1979 (1)	*1980-1984* (2)	*1985-1989* (3)	*1990-1994* (4)	*1995-1999* (5)	*2000-2004* (6)	*2005-2009* (7)	*2010-2014* (8)	*2015-2019* (9)
$NASDAQ$	-0.748	0.141*	0.301***	0.420***	0.446***	0.237***	-0.036	-0.174***	-0.097***
	(0.621)	(0.072)	(0.046)	(0.041)	(0.032)	(0.037)	(0.035)	(0.029)	(0.028)
RET^{+}_{t-1}	2.290***	1.510***	1.753***	1.952***	1.324***	1.186***	1.029***	1.460***	1.254***
	(0.249)	(0.227)	(0.159)	(0.105)	(0.129)	(0.092)	(0.119)	(0.076)	(0.090)
RET^{-}_{t-1}	-0.809***	-1.618***	-1.120***	-2.167***	-2.058***	-2.289***	-2.050***	-2.450***	-2.183***
	(0.294)	(0.213)	(0.435)	(0.184)	(0.131)	(0.102)	(0.178)	(0.144)	(0.130)
LEV_{t-1}	0.204***	0.105***	0.484***	0.290	0.286*	0.331***	0.069***	0.011	0.165**
	(0.038)	(0.020)	(0.163)	(0.189)	(0.158)	(0.081)	(0.020)	(0.040)	(0.079)
$CAPM\_\beta_{t-1}$	-0.814	-0.111	0.308	0.035	-0.026	0.196	0.085*	-0.164*	0.131**
	(0.541)	(0.208)	(0.429)	(0.192)	(0.085)	(0.159)	(0.050)	(0.090)	(0.056)
BTM_{t-1}	1.310*	-0.001	0.324	0.159	-0.018	-0.324*	0.181	-0.372	0.415**
	(0.716)	(0.342)	(0.293)	(0.304)	(0.326)	(0.174)	(0.116)	(0.248)	(0.181)
L_PRC_{t-1}	0.023	-0.081*	-0.017	0.204***	0.176***	0.388***	0.299***	0.116***	0.096***
	(0.075)	(0.044)	(0.033)	(0.030)	(0.024)	(0.024)	(0.020)	(0.018)	(0.016)
L_FAGE_{t-1}	0.341***	0.180**	0.009	-0.167***	-0.159***	-0.128***	-0.073***	-0.095***	-0.094***
	(0.130)	(0.084)	(0.049)	(0.037)	(0.024)	(0.023)	(0.024)	(0.020)	(0.018)
$ESURP_{t-1}$	0.460	0.081	-0.014	0.257***	0.175*	0.066	0.045	0.003	0.144***
	(0.538)	(0.255)	(0.086)	(0.081)	(0.103)	(0.045)	(0.049)	(0.003)	(0.052)
$EVOL_{t-1}$	1.380	0.435	0.180	0.107	0.062	0.102	0.182***	0.010***	0.056
	(1.611)	(0.506)	(0.192)	(0.103)	(0.264)	(0.081)	(0.047)	(0.003)	(0.061)
$NUMEST_{t-1}$	-0.011	0.030***	0.037***	0.034***	0.039***	0.040***	0.041***	0.039***	0.025***
	(0.007)	(0.004)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
$FDISP_{t-1}$	0.017	0.006	0.010*	0.008	0.008	0.008***	0.012*	0.004	0.007*
	(0.115)	(0.005)	(0.005)	(0.006)	(0.006)	(0.003)	(0.007)	(0.006)	(0.004)
STD_DEV_{t-1}	0.579**	0.778***	0.754***	1.618***	1.998***	1.720***	1.554***	1.337***	1.134***
	(0.233)	(0.180)	(0.133)	(0.144)	(0.118)	(0.109)	(0.121)	(0.105)	(0.106)
Within R^2	0.107	0.106	0.155	0.184	0.230	0.272	0.221	0.231	0.154
Adj. R^2	0.258	0.233	0.255	0.282	0.375	0.404	0.363	0.341	0.300
$\% R^2$ Explained	0.76	1.10	0.79	2.78	4.44	3.12	2.83	2.56	1.90
Observations	7,774	18,574	47,998	76,355	96,469	107,739	112,976	112,296	92,204

Log turnover regressed on set of controls and several measures of disagreement over nine 5-year subperiods. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table 10: **Univariate Sorts: Different explanatory variables**

	L_TURN_t									
	D_1	D_2-D_1	D_3-D_2	D_4-D_3	D_5-D_4	D_6-D_5	D_7-D_6	D_8-D_7	D_9-D_8	$D_{10}-D_9$
RET^{+}_{t-1}	-0.853	-0.116**	0.045**	0.060**	0.077**	0.075**	0.089**	0.109**	0.137**	0.259**
RET^{-}_{t-1}	-0.285	-0.264**	-0.109**	-0.090**	-0.066**	-0.068**	-0.051**	-0.044**	-0.013*	0.220**
LEV_{t-1}	-0.471	-0.068**	-0.090**	-0.049**	-0.022**	-0.037**	-0.054**	-0.063**	-0.066**	0.101**
$CAPM\_\beta_{t-1}$	-1.146	0.040**	0.349**	0.150**	-0.080**	-0.289**	0.179**	0.148**	0.124**	0.187**
BTM_{t-1}	-0.185	-0.021**	-0.173**	-0.164**	-0.155**	-0.129**	-0.102**	-0.066**	-0.048**	0.164**
L_PRC_{t-1}	-1.065	0.117**	0.107**	0.025**	0.019**	0.089**	0.165**	0.211**	0.152**	0.041**
L_FAGE_{t-1}	-0.518	-0.067**	-0.021**	-0.012**	-0.127**	-0.232**	-0.020**	0.090**	0.016**	-0.282**
$ESURP_{t-1}$	-0.606	0.060**	0.001	0.000	-0.012**	0.005	-0.024**	-0.027**	-0.032**	0.048**
$EVOL_{t-1}$	-0.515	0.005	0.011**	-0.018**	-0.018**	-0.042**	-0.051**	-0.040**	-0.019**	0.034**
$NUMEST_{t-1}$	-0.398	0.523**	0.171**	0.023**	-0.009	-0.106**	0.041**	-0.080**	-0.043*	-0.011
$FDISP_{t-1}$	-0.341	0.184**	0.059**	0.041**	0.013**	0.001	-0.002	0.028**	-0.033**	-0.011*
STD_DEV_{t-1}	-0.921	0.041**	0.053**	0.065**	0.073**	0.080**	0.060**	0.047**	0.048**	0.134**

Average log turnover measured over univariate portfolio decile sorts of several controls and disagreement measure. At each month, the cross-section of stocks is assigned to 10 portfolios based on the sorting variable. This procedure is repeated for each month. D_i is the i^{th} decile, $D_j - D_i$ is the difference of average L_TURN in D_j and D_i. Corresponding signficance levels are from a t-test of sample means across corresponding decile pairs. Statistical significance of 5% and 1% are indicated by * and ** respectively

Table 11: **Bivariate Sorts**

	L_TURN_t								
	Control_Var_Ter$_{(\mathbf{1},.)}$			**Control_Var_Ter**$_{(\mathbf{2},.)}$			**Control_Var_Ter**$_{(\mathbf{3},.)}$		
	T_{11}	$T_{12}-T_{11}$	$T_{13}-T_{12}$	T_{21}	$T_{22}-T_{21}$	$T_{23}-T_{22}$	T_{31}	$T_{32}-T_{31}$	$T_{33}-T_{32}$
RET^{+}_{t-1}	-0.949	0.192**	0.205**	-0.805	0.157**	0.188**	-0.432	0.181**	0.119**
RET^{-}_{t-1}	-0.520	0.115**	0.045**	-0.876	0.133**	0.129**	-0.897	0.208**	0.248**
LEV_{t-1}	-0.870	0.414**	0.240**	-0.830	0.120**	0.115**	-0.984	0.148**	0.071**
$CAPM\_\beta_{t-1}$	-1.061	0.170**	0.181**	-0.870	0.200**	0.155**	-0.653	0.191**	0.166**
BTM_{t-1}	-0.388	0.151**	0.093**	-0.868	0.121**	0.179**	-1.085	0.119**	0.112**
L_PRC_{t-1}	-1.123	0.227**	0.214**	-0.957	0.305**	0.570**	-0.332	0.228**	0.365**
L_FAGE_{t-1}	-0.770	0.330**	0.251**	-0.899	0.187**	0.120**	-1.050	0.110**	0.037**
$ESURP_{t-1}$	-0.788	0.202**	0.387**	-0.748	0.185**	0.272**	-0.699	0.080**	0.100**
$EVOL_{t-1}$	-0.774	0.227**	0.520**	-0.745	0.173**	0.294**	-0.752	0.064**	0.110**
$NUMEST_{t-1}$	-0.446	0.292**	0.286**	0.042	0.248**	0.412**	-0.058	0.181**	0.340**
$FDISP_{t-1}$	-0.429	0.216**	0.377**	-0.288	0.287**	0.281**	-0.245	0.255**	0.152**

Average log turnover measured over bivariate 3×3 sorts. The first sorting variable is one of the 11 control variables and the second sorting variable is the disagreement measure. Sorting is done at time $t-1$ and turnover is measured at time t. The bivariate sorts are dependent sorts where first, one of the control variable is used to make portfolios and then within those portfolios, disagreement is used to make sub-portfolios. Both dimensions of sorting are cross-sectional where a fresh sorting is performed each month. 70/30 portfolio breakpoints are used to make terciles. $T_{i,j}$ is the avergae log turnover for i^{th} control variable tercile and j^{th} disagreement tercile. Each row represents a new control variable being used to make terciles. Within a row, each control variable tercile is represented as a group of three values. There are three groups each corresponding to a different control variable tercile. The second and thrid entries in each group presents difference in average log turnover over two successive terciles. $T_{i,j}-T_{i,j'}$ is the difference between disagreement terciles j and j' within control variable tercile i. Corresponding signficance levels are from a t-test of sample means across corresponding tercile pairs. Statistical significance of 5% and 1% are indicated by * and ** respectively

Table 12: **Monthly cross-sectional regression: EDGAR implementation**

	L_TURN_t			
	pre EDGAR		*post EDGAR*	
	(1)	(2)	(3)	(4)
$NASDAQ$	0.325***	0.326***	0.066**	0.039
	(0.040)	(0.040)	(0.026)	(0.026)
RET^{+}_{t-1}	2.030***	1.906***	1.595***	1.420***
	(0.114)	(0.110)	(0.072)	(0.064)
RET^{-}_{t-1}	-1.821***	-1.657***	-2.619***	-2.375***
	(0.278)	(0.261)	(0.090)	(0.083)
LEV_{t-1}	0.880***	0.755***	0.688***	0.511***
	(0.162)	(0.157)	(0.104)	(0.101)
$CAPM\_\beta_{t-1}$	-0.016	-0.010	0.092**	0.092**
	(0.178)	(0.177)	(0.040)	(0.038)
BTM_{t-1}	-0.418	-0.186	-1.029***	-0.740***
	(0.276)	(0.268)	(0.181)	(0.181)
L_PRC_{t-1}	0.053*	0.071***	0.184***	0.213***
	(0.028)	(0.028)	(0.014)	(0.015)
L_FAGE_{t-1}	-0.088**	-0.055	-0.138***	-0.109***
	(0.038)	(0.038)	(0.015)	(0.015)
$ESURP_{t-1}$	0.687***	0.609***	0.478***	0.424***
	(0.132)	(0.127)	(0.070)	(0.060)
$EVOL_{t-1}$	0.150	-0.234	0.393***	0.183**
	(0.283)	(0.281)	(0.090)	(0.084)
$NUMEST_{t-1}$	0.033***	0.033***	0.035***	0.034***
	(0.002)	(0.002)	(0.002)	(0.002)
$FDISP_{t-1}$	0.038***	0.026*	0.030***	0.014*
	(0.014)	(0.013)	(0.008)	(0.008)
STD_DEV_{t-1}		0.966***		1.485***
		(0.114)		(0.071)
Within R^2	0.131	0.142	0.185	0.206
Adj. R^2	0.262	0.271	0.369	0.386
% R^2 Explained		1.23		2.58
Observations	121,389	121,389	496,124	496,124

Log turnover regressed on different set of explanatory variables for two disjoint periods: pre EDGAR from Jan 1976 to Mar 1993 and post EDGAR from Jun 1996 to Dec 2019. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table 13: **Correlations and Descriptive Statistics**

Panel A: Correlations

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	*NUMEST*		0.29	0.20	0.19	0.19	0.18	0.15	0.20	0.39	0.03	0.65
(2)	*L_FAGE*	0.26		0.05	−0.07	−0.10	0.37	0.37	−0.08	0.21	0.07	0.33
(3)	*DOC_SIZE*	0.18	0.07		0.66	0.60	0.28	0.21	0.52	0.16	$-0.01^{\#}$	0.33
(4)	*NUM_WORDS*	0.17	−0.04	0.61		0.94	0.17	0.13	0.80	0.21	−0.01	0.32
(5)	*NUM_UNQ_WORDS*	0.18	−0.13	0.52	0.83		0.13	0.10	0.84	0.21	−0.02	0.29
(6)	*NUM_TOT_SEG*	0.18	0.32	0.25	0.15	0.13		0.85	0.14	0.16	0.04	0.30
(7)	*NUM_UNQ_SEG*	0.15	0.31	0.19	0.11	0.10	0.85		0.09	0.13	0.05	0.26
(8)	*COMPLEX_WORDS*	0.20	−0.11	0.43	0.70	0.85	0.14	0.09		0.25	−0.03	0.29
(9)	*INST_OWN*	0.40	0.20	0.12	0.17	0.21	0.17	0.13	0.25		$0.01^{\#}$	0.66
(10)	*SINCE_10K*	0.02	0.06	$-0.00^{\#}$	−0.01	−0.03	0.04	0.04	−0.04	$-0.01^{\#}$		0.02
(11)	*L_ME*	0.75	0.26	0.29	0.27	0.29	0.32	0.28	0.30	0.63	$0.01^{\#}$	

Panel B: Descriptive Statistics

	Mean	**SD**	**Min**	**p25**	**Median**	**p75**	**Max**	**IQR***	**Range***	**Skew**	**Kurt**
NUMEST	7.97	0.98	6.41	7.31	7.65	8.92	10.01	1.64	3.67	0.35	1.82
L_FAGE	4.27	0.65	0.00	4.13	4.32	4.72	5.00	0.91	7.73	-2.58	13.29
DOC_SIZE	5.91	7.31	0.34	0.42	1.71	16.28	20.90	2.17	2.81	0.94	2.10
NUM_WORDS	46.15	8.74	31.60	37.05	45.59	53.45	59.55	1.88	3.20	-0.15	1.65
NUM_UNQ_WORDS	2.85	0.32	2.31	2.53	2.85	3.13	3.33	1.87	3.19	-0.13	1.63
NUM_TOT_SEG	2.10	0.43	1.38	1.65	2.18	2.48	2.90	1.95	3.55	-0.27	1.64
NUM_UNQ_SEG	1.63	0.23	1.31	1.54	1.57	1.61	2.49	0.30	5.12	1.76	6.27
COMPLEX_WORDS	0.09	0.01	0.07	0.08	0.09	0.09	0.10	1.72	3.30	-0.22	1.81
INST_OWN	0.40	0.17	0.16	0.25	0.34	0.58	0.67	1.92	2.99	0.18	1.48
SINCE_10K	7.31	3.30	0.00	4.96	7.20	9.27	16.55	1.31	5.01	0.31	2.78
L_ME	11.56	1.17	9.43	10.66	11.22	12.68	13.74	1.72	3.68	0.33	1.91

Panel A reports the time-series average of cross-sectional correlation coefficients. Correlations marked with ($^{\#}$) are not significant at the 5 % level. At each month, cross-sectional correlation among variables is computed and then their time-series average over the duration of sample is reported. Lower triangle represents variable correlation while the upper triangle consists of rank correlations. Panel B presents descriptive Statistics of explanatory variables. Interquartile range (IQR^{*}) and variable range ($Range^{*}$) are in multiples of standard deviation. At each month the average of all the variables is computed and their time-series descriptive statistics are reported. *DOC_SIZE* is in megabytes while *NUM_WORDS*, *NUM_UNQ_WORDS* and *COMPLEX_WORDS* is in 1000s of words. Variable definitions are present in Appendix A.2.

Table 14: **Information Environemnt: Summary of Regression Splits**

Portfolio Criterion	w/o STD_DEV	with STD_DEV		% R2 Explained
	$FDISP_{t-1}$	$FDISP_{t-1}$	STD_DEV_{t-1}	
Number of Analysts				
$NUMEST \in \{2,3\}$	0.046***	0.030***	1.652***	2.46
$NUMEST \in \{4\dots10\}$	0.043***	0.023***	1.569***	3.49
$NUMEST \geq 11$	0.061***	0.043***	1.198***	2.93
Firm Age				
$FIRM\_AGE-1$	0.018*	0.003	1.754***	3.01
$FIRM\_AGE-2$	0.045***	0.030***	1.230***	1.73
$FIRM\_AGE-3$	0.053***	0.034**	0.969***	1.66
10-K Document Size				
DOC_SIZE-1	0.016	0.005	1.202***	1.64
DOC_SIZE-2	0.028***	0.013	1.412***	2.51
DOC_SIZE-3	0.039***	0.018	1.719***	3.53
10-K Report Length				
$LENGTH-1$	0.025	0.010	1.175***	1.61
$LENGTH-2$	0.021**	0.008	1.414***	2.43
$LENGTH-3$	0.034***	0.017	1.622***	3.37
10-K Unique Words				
$UNQ\_WORDS-1$	0.022	0.007	1.241***	1.76
$UNQ\_WORDS-2$	0.022**	0.006	1.424***	2.45
$UNQ\_WORDS-3$	0.034***	0.020	1.620***	3.41
Number of Segments				
$TOT\_SEG=1$	0.027**	0.013	1.413***	2.16
$TOT\_SEG>1$	0.044***	0.024***	1.265***	2.29
$UNQ\_SEG=1$	0.029***	0.016*	1.355***	2.07
$UNQ\_SEG>1$	0.045***	0.023**	1.283***	2.38
10-K Complex Words				
$COMPLEX\_WORDS-1$	0.034*	0.015	1.357***	2.02
$COMPLEX\_WORDS-2$	0.021**	0.007	1.366***	2.28
$COMPLEX\_WORDS-3$	0.013	-0.003	1.688***	4.06
Ownership Percentage				
$OWN(\%) \in [0,0.5]$	0.051***	0.036***	1.564***	2.32
$OWN(\%) \in (0.5,0.75)$	0.029***	0.010	1.398***	2.95
$OWN(\%) \in [0.75,100]$	0.037***	0.017*	1.432***	4.08
Months (s) Since 10-K Filing				
$s \in \{1,2,3\}$	0.036***	0.022***	1.441***	2.52
$s \in \{4,5,6\}$	0.032***	0.012	1.459***	2.51
$s \in \{7,8,9\}$	0.019	0.001	1.504***	2.67
$s \geq 10$	0.032**	0.014	1.599***	2.93

The table presents summary of turnover regressions across portfolios made using several variables related to firm's information environment. Two sets of regressions, corresponding to base specifications 1 and 4 of Table 3 are estimated for each portfolio. Forecast dispersion is included in both regressions. Last column gives the percentage increase in unexplained R^2 after including disagreement in the regression. For brevity, other explanatory variables (Table 3), standard errors and R^2 statistics are skipped. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table 15: **Monthly cross-sectional regression: Information environment, return volatility and bid-ask spread**

	RET_VOL_t			$SPREAD_t$		
	(1)	(2)	(3)	(4)	(5)	(6)
STD_DEV_{t-1}	0.034***	0.033***	0.034***	0.080*	0.034	0.077*
	(0.001)	(0.001)	(0.001)	(0.042)	(0.047)	(0.043)
L_ME_{t-1}	-0.003***	-0.003***	-0.003***	-0.052***	-0.048***	-0.049***
	(0.000)	(0.000)	(0.000)	(0.006)	(0.006)	(0.006)
INF_ENV_{t-1}		0.002***			0.081***	
		(0.000)			(0.016)	
$INF\_ENV\_R_{t-1}$			-0.007			0.856***
			(0.004)			(0.320)
Within R^2	0.147	0.149	0.147	0.002	0.003	0.002
Adj. R^2	0.315	0.317	0.315	0.006	0.006	0.006
Observations	407,478	407,478	407,478	407,478	407,478	407,478

Return Volatility and bid-ask spread regressed on disagreement, firm size and information environment. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table 16: **Variable Definitions**

Variable	Definition
$NASDAQ$	Dummy set to 1 if the stock is traded at NASDAQ ($exchcd = 3$)
RET^+ and RET^-	Monthly return is decomposed into two variables based on its sign. $RET^+ = \max(ret, 0)$ and $RET^- = \min(ret, 0)$. ret is adjusted for delisting of firms.
RET_RF	Excess monthly return. Defined as $RET\_RF = ret - R_F$ where R_F is the risk-free rate.
BE, ME and BTM	Book value of equity, market value of equity and the ratio of book value to market value of equity. Construction of book equity is described in Appendix A.1.
LEV	Ratio of long term debt to book value of equity.
$CAPM\_ALPHA$ and $CAPM\_BETA$	The intercept and slope parameters from regressing firm's excess returns on market excess returns. Regression parameters are obtained in a rolling fashion using the past 60 months of returns data (from t to $t-59$). The intercept is $CAPM\_ALPHA$ while the slope is $CAPM\_BETA$. Additionally, at least 24 non-missing return observations are required to estimate the regression.
PRC	Stock price adjusted for splits, rights issues and other corporate events that affect the face value of a share.
L_FAGE	Firm age is the natural log of number of months since the firm first appeared on the CRSP monthly database.
$ESURP$	Absolute earning surprise is the absolute difference between the most recent quarterly earnings per share (EPS_q) and EPS 4 quarters ago (EPS_{q-4}) scaled by quarter end stock price (P_q). EPS and stock price are adjusted for splits. $ESURP = \frac{\lvert EPS_q - EPS_{q-4} \rvert}{P_q}$ for quarter q.
$EVOL$	Volatility of earnings is the standard deviation of eight recent quarterly earnings per share scaled by the quarter end stock price. $EVOL = \frac{1}{7 \cdot P_q} \cdot \sum_{i=0}^{7} (EPS_{q-i} - \overline{EPS_q})^2$, where $\overline{EPS_q}$ is the mean EPS over the same period.
$NUMEST$	Number of analyst following a firm in a given month
$FDISP$	Standard deviation of analyst forecasts following a firm scaled by absolute value of mean forecast estimate. I require that at least two analysts are following the firm ($NUMEST \geq 2$)
STD_DEV	Standard deviation of all signals for a firm in a month. I require that at least 10 signals are present to reliably estimate standard deviation.

Table 16: **Variable Definitions** *(continued)*

Variable	Definition
$NBUY$	Number of anomalies signalling buy (+1) for a firm in a particular month.
$NSELL$	Number of anomalies signalling sell (-1) for a firm in a particular month.
$MEAN\_SIGNAL$	Average of all signals for a firm in a month. A positive (negative) value indicates that $NBUY(NSELL)$ is higher than $NSELL(NBUY)$.
ABS_DEV	Mean absolute deviation of all the signals for a firm in a month. If T_k represents $k^{th}(1 \leq k \leq K)$ trading signal then $ABS\_DEV = \frac{1}{K}\sum_{k=1}^{K}\lvert T_k - MEAN\_SIGNAL\rvert$.
PC_DEV	Standard deviation of all signals projected in the principal directions[a].
$CONT\_DEV$	Standard deviation of anomaly ranks. Let $A_{f,t,s}$ denote s^{th} anomaly for firm f at time t. Anomal rank, $AR_{f,t,s}$ is the ordering of $A_{f,t,s}$ across f. $AR_{f,t,s}$ is scaled to lie between $[0,1]$. $CONT\_DEV_{f,t}$ is then the standard deviation of $AR_{f,t,s}$ across s.
LO_CORR_SD and HI_CORR_SD	Standard deviation of two sets of signals where the two sets differ in their aggregate absolute correlation within the group[b]. At each month, using a brute-force approach over the full set of 36 anomaly signals, I search for a set of 18 signals which produce the smallest aggregate correlation. The standard deviation of signals within this set is LO_CORR_SD and of the remaining set is HI_CORR_SD
NUM_FLIPS	The number of times any two pairs of signals flip. Let s and s' be any two distinct signals for a given firm. A flip occurs at time t if $s_{t-1} \cdot s'_{t-1} = -1, s_{t-1} \cdot s_t = -1$ and $s'_{t-1} \cdot s'_t = -1$. The number of such occurrences is NUM_FLIPS.
NUM_DIV	The number of times any two pairs of signals diverge. Let s and s' be any two distinct signals for a given firm. A divergence occurs at time t if $s_{t-1} = s'_{t-1} = 0$ and $s_t \cdot s'_t = -1$. The number of such occurrences is NUM_DIV.
$TURN$	Monthly share turnover calculated as monthly share volume divided by adjusted shares outstanding.
$TURN\_1d$ and $TURN\_5d$	Share turnover for the first day and first five trading days of the month respectively.
$TURN\_GRT$	Turnover adjusted as per Gallant et al. (1992). Non-stationarity and calendar effects are removed from both the mean and variance of turnover time-series.[c]

Table 16: **Variable Definitions** *(continued)*

Variable	Definition
$ILLIQ$	Amihud (2002) Illiquidity measure constructed using daily returns and volume data from CRSP daily stock files. $ILLIQ = \frac{1}{D} \cdot \sum_{d=1}^{d=D} \frac{\lvert ret_d \rvert}{vol_d}$, where D is the number of trading days in a month. Days with zero trading are excluded from the summation.
L_TURN_ILLIQ	Residuals from regressing L_TURN on an intercept and L_ILLIQ.
VW_L_TURN and EW_L_TURN	Residuals from regressing L_TURN on an intercept and log of value (equal) weighted market turnover[d].
DIV_12M	Return accruing to dividend payments in last 12 months. It is calculated as the ratio of past twelve months return by past twelve months price change. The latter doesn't incorporate dividend returns.
RET_VOL	Monthly return volatility computed as the standard deviation of daily stocks returns.
O_SCORE	Ohlson (1980) distress measure.
NUM_WORDS, NUM_UNQ_WORDS and DOC_SIZE	Total number of words, number of unique words and the file size of EDGAR 10-K filing. All these variables are borrowed from the LM summary file compiled by Bill McDonald at `https://sraf.nd.edu/`
$COMPLEX\_WORDS$	Number of unique occurrences of complex words in firm's 10-K filing. List of complex words is from Loughran and McDonald (2020).
NUM_SEG and NUM_UNQ_SEG	Total number of segments (unique segments). Only business and operating segments of firms reporting sales are considered. Unique segments count all segments within an industry as one segment while total segments count them separately. Only the most recent reported segments are considered by setting fiscal period equal to report period in compustat segment database.
OWN_PERC	Percent of shares held by institutional investors as reported in 13-F Institutional ownership data.
$SPREAD$	Bid-ask spread scaled by the mid point of bid and ask price. I use $bidlo$ and $askhi$ from CRSP to compute the spread.
$FF3\_ALPHA$ and $FF3\_BETA$	Excess returns are regressed on three well known factors (Fama and French (1992)) to predict returns: excess market return, size and book to market[e]. 60 month rolling regressions are used to estimate the coefficients. $FF3\_ALPHA$ is the intercept while $FF3\_BETA$ is the slope parameter for excess market return.

Table 16: **Variable Definitions** *(continued)*

Variable	Definition
$FF5\_ALPHA$ and $FF5\_BETA$	Same procedure to estimate as that of FF-3 regression except that there are two additional factors[f] (Fama and French (2015)).

[a] At time t, let S be the $F_t \times N$ matrix of signals where F_t is the number of firms and each firm has N signals. Let W be an $N \times N$ matrix of principal components of S. STD_DEV is the row-wise standard deviation of S and PC_DEV is the row-wise standard deviation of SW which is the projection of signals on the principal component space.

[b] Let C be the $N \times N$ correlation matrix of the entire set of signals, N being the number of signals. Finding a smaller set of $n (n < N)$ signals is identical to solving the integer program, $\min_v v^T |C| v$, s.t. $v^T \mathbf{1} = n, v \in \{0,1\}^N$, where $\mathbf{1}$ is a vector of all ones and $|C|$ is the absolute correlation matrix with $|C|_{i,j} = |C_{i,j}| \forall i,j \in 1 \ldots N$. The above optimization problem has integer valued range of values v can take. This type of problem is a hard problem and can only solved by iterating over all possible values of v.

[c] GRT adjustments is carried out in two steps. In the first step the variable to be adjusted, X is regressed on linear and quadratic time trends as well as calendar month dummies: $X_t \sim \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \gamma \cdot D_{1\ldots11} + \epsilon_t$. Here $D_{1\ldots11}$ represents 11 month of the year dummies. In the next step squared residuals are regressed on the same set of variables: $\log(\epsilon_t^2) \sim \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \gamma \cdot D_{1\ldots11} + u_t$. Then the GRT adjusted series is defined as, $X\_GRT_t = \exp(u_t/2)$. Finally, X_GRT is linearly transformed so that its mean and variance matches that of X

[d] Value weighted market turnover is $\sum_{i=1}^{D_t} \frac{ME_{i,t}}{\widehat{ME}_t} \cdot TURN_{i,t}$ and equal weighted market turnover is $\frac{1}{D_t} \cdot \sum_{i=1}^{D_t} TURN_{i,t}$, where where $\widehat{ME}_t = \sum_{i=1}^{D_t} ME_{i,t}$ and D_t is number of firms at time t.

[e] The explanatory variables of the regression are obtained from Kenneth French's online data library. The SMB (size factor) and HML (value factor) construction is provided at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html

[f] The two additional factors are profitability (RMW) and investments (CMA). Their definition is provided at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5_factors_2x3.html

Trading Volume and Dispersion of Signals

IA Internet Appendix

IA.1 Disagreement and Correlations

Disagreement and signal correlations are tightly related to each other. If two signals are perfectly corrleated then they have zero disagreement within them while if they are perfectly negatively correlated then they have maximum deviation. Thus correlation and disagreement are negatively related to each other. To see the connection more generally, assume that for each signal a fraction p of the stocks fall in the buy category, q in sell category and the rest $1-p-q$ in the hold category. $T_{f,t,s}$ can then be modelled as a random variable such that,

$$T_{f,t,s} = \begin{cases} 1, & \text{w.p.} \quad p \\ -1, & \text{w.p.} \quad q \\ 0, & \text{w.p.} \quad 1-p-q \end{cases}$$

We can drop the firm identifier f as all the avergaes and correlation will be across all firms. The signals also have a time-varying correlation structure given by $Corr(T_{ts}, T_{ts'}) = \rho_{tss'}$. Its straightforward to show that, $\mathbb{E}[T_{ts}] = p-q$, $\quad \mathbb{E}[T_{ts}^2] = p+q$ and $\quad \text{var}[T_{ts}^2] = p+q-(p-q)^2$. $\mathbb{E}[T_{ts}T_{ts'}] = \mathbb{E}[T_{ts}]\cdot\mathbb{E}[T_{ts'}] + \text{cov}(T_{ts}, T_{ts'}) = (p-q)^2(1-\rho_{tss'}) + (p+q)\rho_{tss'}$. The mean signal, $\overline{T_t} \equiv \frac{1}{N}\sum_s T_{ts}$, where N is the total number of anomaly signals, has a mean of $\frac{p-q}{N}$ and $\mathbb{E}[\overline{T_t}^2] = \frac{1}{N^2}\sum_s\sum_{s'}\mathbb{E}[T_{ts}T_{ts'}] = \frac{1}{N^2}\sum_s\sum_{s'}\left((p-q)^2(1-\rho_{tss'}) + (p+q)\rho_{tss'}\right) = (p-q)^2 + \frac{C_t}{N^2}\left((p+q)-(p-q)^2\right)$ where $C_t \equiv \sum_s\sum_{s'}\rho_{tss'}$ is sum of all N^2 correlations. The disagreement between signals is defined as, $S_t \equiv \frac{1}{N-1}\sum_s(T_{ts} - \overline{T_t})^2 = \frac{1}{N-1}\left(\sum_s T_{ts}^2 - N\overline{T_t}^2\right)$. Taking expectation we get,

$$\mathbb{E}[S_t] = \frac{N}{N-1}\left(1-\frac{C_t}{N^2}\right)\left(p+q-(p-q)^2\right)$$

$\mathbb{E}[S_t]$ is the cross sectional average of disagreement at time t, which is maximized when $p = q = 0.5$. However in the time-series it depends only on signal correlations and the dependence is only through sum of all indivudal correlations, C_t. This is reassuring in terms of the construction of disagreement where its dependence on one particular signal is limited to how the signal is correlated with all other signals. Skipping or adding a few signals shouldn't vastly change the disagreement. The dependence of disagreement on correlation also gives a natural demarcation of low and high disagreement regimes. We can partition the entire set of signals ($\mathscr{S}$) into two disjoint subsets $\mathscr{S}^{lo}$ and $\mathscr{S}^{hi}$ such that $C_t^{hi} - C_t^{lo}$ is maximized. $\mathscr{S}^{lo}$ represents a set of signals which have small correlations among themselves while $\mathscr{S}^{hi}$ is the set of highly correlated signals. Since $\frac{\partial\mathbb{E}[S_t]}{\partial C_t} < 0$, disagreement within $\mathscr{S}^{lo}$ should be higher on average. We are interested in examining which set of signals predicts trading volume more strongly. Signals within $\mathscr{S}^{hi}$ are usually of the same sign becuase of high correlation. Investors would expect these signals to move together and any disagreement would come as a surprise. Thus whenver signals in this set disagree, it should cause a heightened trading response because disagreement within $\mathscr{S}^{hi}$ is less likely and when it occurs it creates more uncertainty about the asset's future performance[33].

IA.2 Correlation between disagreement and signals

I explore the correlations between disagreement and signals. According to Figure 5, momentum anomalies contributes most to the disagreement level but it doesn't tell how does signal variation relate to variation in disagreement. If one particular signal or a class of signals is recommending sell then what does it tells us about the disagreement? Will it be higher, lower or unaffected? Figure IA.1 gives the correlation map of disagreement and signals. Profitability, value, security issuance and distress signals are mostly negatively correlated while earnings quality is positively correlated. Other signals groups doesn't seem to have a clear direction of association with disagreement. Of the 36 signals, Earnings to Price ratio has a correlation of -0.280 and Net Operating Assets has a correlation of 0.158 with disagreement. The average correlation across all 36 signals is -0.028. If we exclude distress anomalies from disagreement this correlation goes upto -0.126. It is an interesting finding that most of the signals correlate negatively with disagreement. The signals are return predicting anomalies and a higher value of signal predicts higher future return[34]. Thus a negative association between anomalies and disagreement hints at a negative disagreement return relationship.

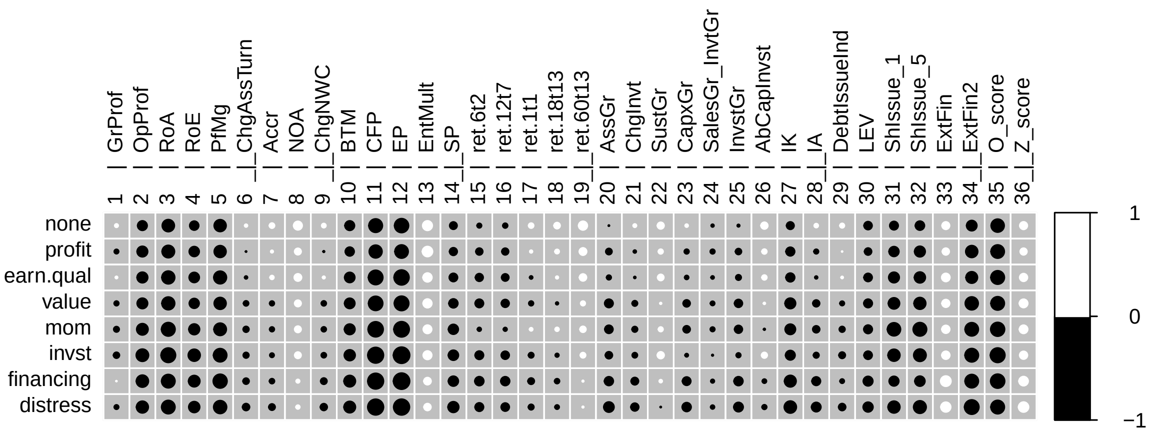

Figure IA.1: Correlation heat map of 36 signals and disagreement. The row names gives the category of signals left out in the disagreement measure (see Table 1). Positive correlation is represented by white circles while negative correlation is shown by black circles. The size of the circle is proportional to the magnitude of the correlation coefficient.

IA.3 Within and Cross Disagreement

Figure 5 reveals that different anomaly groups affect disagreement differently. In particular, momentum anomalies has the most contribution to disagreement measure. Does these differences

[33] Unfortunately, we can't say anything about var$[S_t]$ within the framework of the above model. This requires computation of covariances between signal pairs, i.e. cov$(T_{ts}T_{ts'}, T_{ts''}T_{ts'''})$. Estimating these from data is also not possible since with 36 signals the total number of estimations every period would be ${}^{36}C_4 = 58905$.

[34] All anomalies which predict a negative association with future returns are scaled by -1 so as to make them positively associate with future returns. The predicted association is as per the original source which documented the anomaly. The complete list of anomalies and the predicted relationship is present in Table 1.

in disagreement across anomaly groups also materialize in volume regressions? Conversely, how much of the disagreement volume relationship is explained by disagreement within anomaly groups and disagreement across anomaly groups.

Let there be G anomaly groups with g^{th} group containing n_g anomalies ($g \in 1,2,\ldots,G$). Let $T_{f,t,g}$ be the average trading signal[35] for anomaly group g, firm f at time t. Thus, $T_{f,t,g} = \frac{1}{n_g} \cdot \sum_{i=1}^{n_g} T_{f,t,i}$ where i denotes a particular anomaly within a group. Let $N = \sum_{g=1}^{G} n_g$ equal total number of anomales and, $\overline{T_{f,t}} = \frac{1}{N} \cdot \sum_{i=1}^{N} T_{f,t,i} = \frac{1}{N} \cdot \sum_{g=1}^{G} n_g \cdot T_{f,t,g}$ be the average trading signal across all anomalies. The cross anomaly disagreement captures the extent by which group trading signals differ from mean singal across all groups. We can define the cross groups sum of squares as, $SS^{cross} = \sum_{g=1}^{G} n_g \cdot (T_{f,t,g} - \overline{T_{f,t}})^2$. The total sum of squares is, $SS^{total} = \sum_{g=1}^{G} \sum_{i=1}^{n_g} (T_{f,t,i} - \overline{T_{f,t}})^2$. Using these two squared sums, we can define within sum of squares as, $SS^{within} = SS^{total} - SS^{cross}$. Since different groups have different number of constituent anomalies, we can compute disagreement across anomaly groups by weighing group sum of squares with number of anomalies in that group. $STD\_DEV^{cross}_{f,t} = \sqrt{\frac{1}{G-1} \cdot \sum_{g=1}^{G} \frac{n_g}{N/G} \cdot (T_{f,t,g} - \overline{T_{f,t}})^2}$. An equally weighted computation would use unity in place of $\frac{n_g}{N/G}$ as weights. Overall disagreement across all signals is $STD\_DEV^{total}_{f,t} = \sqrt{\frac{SS^{total}}{N-1}}$. Similarly, the within disagreement is defined as $STD\_DEV^{within}_{f,t} = \sqrt{\frac{SS^{within}}{N-G}}$.

I consider six anomaly groups viz. profitability (P), earnings quality (E), valuation (V), momentum (M), investment (I) and, financing & distress (F). Table 1 gives the list of 36 anomalies and their groups. Table IA.1 gives regression results. Of the six groups, profitability, earnings quality and, disagreement across valuation anomalies doesn't significantly impact volume (specification 1-3 and 7). Momentum group is by far the most impactful anomaly group affecting the disagreement-volume relation. Disagreement within 5 momentum anomalies alone explains upto 5.7% of variation in turnover (specification 4). Investing and financing anomalies are also significant, albeit less than momentum anomalies (specification 5 and 6). Within disagreement explains 6.1% turnover next period while disagreement across anomaly groups explain 1.5% turnover (specification 9 and 10). Both types of disagreement are statistically and economically significant. Decomposing the overall disagreement into within and cross disagreement increases the amount of unexplained variation from 6.5% to nearly 7% (specification 8 and 11). The sample size in this table is considerably smaller than the sample in base specification (Table 3) due to the requirement that all anomalies must be present for computing group-wise disagreement.

[35] For ease of disagreement computation within an anomaly group, I am considering variable ranks rather than -1/0/1 signals. For small anomaly groups (like earnings quality group), the variability of disagreement is severely restricted if discrete valued signals are used.

Table IA.1: **Monthly cross-sectional regression: Within and between group disagreement**

	L_TURN_t											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$STD\_DEV^{P}_{t-1}$	0.208***						-0.013					0.035
	(0.076)						(0.071)					(0.071)
$STD\_DEV^{E}_{t-1}$		0.090					0.045					0.129**
		(0.061)					(0.059)					(0.059)
$STD\_DEV^{V}_{t-1}$			0.092				-0.126					-0.190**
			(0.089)				(0.085)					(0.086)
$STD\_DEV^{M}_{t-1}$				1.963***			1.870***					1.813***
				(0.062)			(0.060)					(0.058)
$STD\_DEV^{I}_{t-1}$					1.245***		1.007***					1.072***
					(0.084)		(0.080)					(0.081)
$STD\_DEV^{F}_{t-1}$						0.782***	0.550***					0.301***
						(0.111)	(0.106)					(0.106)
$STD\_DEV^{ALL}_{t-1}$								3.482***				
								(0.191)				
$STD\_DEV^{WITHIN}_{t-1}$									3.476***		3.199***	
									(0.192)		(0.184)	
$STD\_DEV^{CROSS}_{t-1}$										1.720***	1.342***	1.497***
										(0.137)	(0.130)	(0.135)
$\%R^2$ Explained	0.29	0.12	0.24	5.67	2.38	0.51	7.57	6.52	6.16	1.48	6.96	8.85
Observations	434,661	434,661	434,661	434,661	434,661	434,661	434,661	434,661	434,661	434,661	434,661	434,661

Log turnover regressed on set of controls and disagreement across and within several anomaly groups. 36 anomalies from Table 1 are divided into six groups: profitability (P), earnings quality (E), valuation (V), momentum (M), investment (I) and, financing & distress (F). $STD\_DEV^{g}_{t-1}$ is the disagreement across signals within anomaly group g where $g \in \{P, E, V, M, I, F\}$. $STD\_DEV^{ALL}_{t-1}$, $STD\_DEV^{WITHIN}_{t-1}$ and, $STD\_DEV^{CROSS}_{t-1}$ is the total, within-group and cross-group disagreement respectively. In all regressions control variables are not shown for brevity. The controls used in regressions are $NASDAQ$, RET^{+}, RET^{-}, LEV, $CAPM\_\beta$, BTM, L_PRC, L_FAGE, $ESURP$, $EVOL$, $NUMEST$ and, $FDISP$. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

IA.4 Robustness Checks

As a robustness check, I perform disagreement-volume regression across portfolios partitoned by book to market and leverage terciles. I also look at NASDAQ stocks and different stock splitting rules. Lastly, I present some more results from univariate portfolio sorts.

IA.4.1 Book to market splits

Tables IA.2 gives the regression results for BTM terciles. The coefficient on STD_DEV is highest for high BTM stocks i.e. value stocks. Across the three terciles, a one SD change in STD_DEV predicts 4.9%, 8.9% and 12.5% higher turnover in next month respectively. Adj. R^2 and $\%R^2$ explained also increases with BTM terciles. Thus, disagreement arising from fundamental anomalies has more explanatory power for value stocks. This also provides evidence in favour of the hypothesis that investors in value stocks primarily use return anomalies for their trading decisions and hence disagreement among the anomalies strongly predicts next month's trading volume.

IA.4.2 NASDAQ stocks

NASDAQ stocks are structurally different from NYSE/AMEX stocks. The exchange was constituted in 1971 with electronic stock market. The stocks at NASDAQ exchange tend to be young and small technology firms. As of December 2018, the average NYSE/AMEX firm is 2.6 times bigger and 9 years older than the average NASDAQ firm. The evidence in Tables 8 and IA.9 suggest that small and young stocks have a bigger disagreement coefficient. Since NASDAQ stocks generally have both these characteristics, we should expect to see much larger coefficients as compared to base rgeression (specification (4) of Table 3). Table IA.3 below gives the regression summary for NASDAQ stocks. Not only is the coefficient on disagreement higher than base regression, the explanatory power is also much higher. Disagreement explains upto 4% of unexplined variation in turnover after controlling for other covariates. A one SD change in disagreement predicts 11.8% higher turnover in the full model (Table IA.3 specification (4)).

IA.4.3 Different stock splits

I have performed the analysis in this study using 70/30 stock splits on NYSE/AMEX universe to construct disagreement. To safeguard against the choice of the particular splitting criterion, I perform robustness check using 80/20 and 50/50 splits. Additionally, I also calculate disagreement using all stocks from NYSE, AMEX and NASDAQ exchanges. Table IA.4 gives the results. The biggest regression coefficient appears when all stocks are used in a 80/20 split to form disagreement where the inclusion of disagreement improves unexplained variation by 2.1%. 50/50 split gives insignificant results with no increase in R^2. The base regression used in this study is the NYSE/AMEX 70/30 split of specification (6). A one SD change in ${}^{ALL}STD\_DEV^{80/20}$, ${}^{ALL}STD\_DEV^{70/30}$, ${}^{NYSE}STD\_DEV^{80/20}$ and ${}^{NYSE}STD\_DEV^{70/30}$ predicts next month's turnover to be higher by 10.8%, 8.7%, 9.7% and 7.8% respectively.

Table IA.2: **Monthly cross-sectional regression: BTM terciles**

	L_TURN_t					
	BTM-1		*BTM-2*		*BTM-3*	
	(1)	(2)	(3)	(4)	(5)	(6)
$NASDAQ$	0.312***	0.287***	0.109***	0.096***	-0.024	-0.018
	(0.033)	(0.033)	(0.025)	(0.025)	(0.036)	(0.035)
RET^+_{t-1}	1.515***	1.424***	1.655***	1.499***	1.347***	1.207***
	(0.067)	(0.063)	(0.067)	(0.061)	(0.094)	(0.084)
RET^-_{t-1}	-2.363***	-2.198***	-2.537***	-2.277***	-2.225***	-2.083***
	(0.088)	(0.085)	(0.101)	(0.094)	(0.101)	(0.099)
LEV_{t-1}	-0.233	-0.090	1.369***	1.158***	0.531***	0.423***
	(0.169)	(0.161)	(0.153)	(0.149)	(0.111)	(0.111)
$CAPM\_\beta_{t-1}$	0.069*	0.073*	0.050	0.049	0.171***	0.162***
	(0.038)	(0.037)	(0.038)	(0.037)	(0.049)	(0.047)
BTM_{t-1}	-3.613***	-1.733**	-2.567***	-1.560***	0.363	0.325
	(0.676)	(0.673)	(0.392)	(0.387)	(0.244)	(0.242)
L_PRC_{t-1}	0.185***	0.201***	0.133***	0.172***	0.194***	0.244***
	(0.017)	(0.017)	(0.015)	(0.015)	(0.023)	(0.025)
L_FAGE_{t-1}	-0.219***	-0.197***	-0.113***	-0.090***	-0.049**	-0.035
	(0.020)	(0.020)	(0.016)	(0.016)	(0.024)	(0.023)
$ESURP_{t-1}$	0.541***	0.493***	0.781***	0.663***	0.355***	0.342***
	(0.099)	(0.096)	(0.100)	(0.089)	(0.076)	(0.068)
$EVOL_{t-1}$	0.281*	0.184	0.541***	0.174	0.412***	0.258***
	(0.160)	(0.162)	(0.174)	(0.149)	(0.101)	(0.096)
$NUMEST_{t-1}$	0.016***	0.017***	0.037***	0.037***	0.054***	0.053***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
$FDISP_{t-1}$	-0.006	-0.015	0.041***	0.021*	0.057***	0.044***
	(0.017)	(0.017)	(0.011)	(0.011)	(0.009)	(0.009)
STD_DEV_{t-1}		1.164***		1.358***		1.503***
		(0.107)		(0.068)		(0.123)
Within R^2	0.145	0.155	0.163	0.181	0.212	0.226
Adj. R^2	0.422	0.429	0.455	0.466	0.415	0.426
$\%R^2$ Explained		1.17		2.19		1.86
Observations	186,856	186,856	317,010	317,010	168,519	168,519

Log turnover regressed on set of controls and several measures of disagreement across three BTM terciles made on 70/30 BTM splits. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table IA.3: **Monthly cross-sectional regression: Subsamples by exchange**

	L_TURN_t					
	ALL		*NYSE/AMEX*		*NASDAQ*	
	(1)	(2)	(3)	(4)	(5)	(6)
RET^{+}_{t-1}	1.742***	1.551***	1.609***	1.460***	1.571***	1.412***
	(0.071)	(0.063)	(0.067)	(0.062)	(0.086)	(0.078)
RET^{-}_{t-1}	-2.588***	-2.327***	-2.407***	-2.180***	-2.367***	-2.170***
	(0.092)	(0.085)	(0.081)	(0.076)	(0.118)	(0.112)
LEV_{t-1}	0.712***	0.530***	0.546***	0.393***	0.746***	0.596***
	(0.093)	(0.091)	(0.133)	(0.130)	(0.118)	(0.116)
$CAPM\_\beta_{t-1}$	0.078**	0.078**	0.051	0.055	0.105***	0.102***
	(0.038)	(0.036)	(0.042)	(0.039)	(0.038)	(0.037)
BTM_{t-1}	-0.976***	-0.657***	-1.715***	-1.343***	-0.098	0.073
	(0.168)	(0.167)	(0.196)	(0.190)	(0.233)	(0.235)
L_PRC_{t-1}	0.165***	0.197***	0.184***	0.209***	0.130***	0.155***
	(0.013)	(0.013)	(0.016)	(0.016)	(0.019)	(0.019)
L_FAGE_{t-1}	-0.173***	-0.136***	-0.157***	-0.127***	-0.080***	-0.057***
	(0.015)	(0.015)	(0.021)	(0.020)	(0.018)	(0.017)
$ESURP_{t-1}$	0.506***	0.439***	0.551***	0.497***	0.427***	0.374***
	(0.069)	(0.059)	(0.085)	(0.076)	(0.065)	(0.059)
$EVOL_{t-1}$	0.350***	0.110	0.337***	0.075	0.373***	0.203**
	(0.089)	(0.083)	(0.125)	(0.120)	(0.115)	(0.103)
$NUMEST_{t-1}$	0.032***	0.032***	0.049***	0.048***	0.026***	0.026***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
$FDISP_{t-1}$	0.041***	0.022***	0.029***	0.015	0.047***	0.028***
	(0.008)	(0.008)	(0.010)	(0.010)	(0.011)	(0.010)
STD_DEV_{t-1}		1.501***		1.418***		1.174***
		(0.066)		(0.083)		(0.084)
Within R^2	0.161	0.183	0.221	0.237	0.133	0.151
Adj. R^2	0.422	0.437	0.419	0.431	0.511	0.521
$\%R^2$ Explained		2.63		2.04		2.07
Observations	672,385	672,385	316,277	316,277	356,108	356,108

Log turnover regressed on different subsamples based on stock exchange. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table IA.4: **Monthly cross-sectional regression: disagreement by different stock splits**

	L_TURN_t									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
${}^{ALL}STD\_DEV_{t-1}^{80/20}$		1.414***								
		(0.058)								
${}^{ALL}STD\_DEV_{t-1}^{70/30}$			1.534***							
			(0.068)							
${}^{ALL}STD\_DEV_{t-1}^{50/50}$				-0.141*						
				(0.084)						
${}^{NYSE}STD\_DEV_{t-1}^{80/20}$					1.038***					
					(0.074)					
${}^{NYSE}STD\_DEV_{t-1}^{70/30}$						1.096***				
						(0.084)				
${}^{NYSE}STD\_DEV_{t-1}^{50/50}$							-0.308***			
							(0.099)			
${}^{NASDAQ}STD\_DEV_{t-1}^{80/20}$								1.295***		
								(0.071)		
${}^{NASDAQ}STD\_DEV_{t-1}^{70/30}$									1.388***	
									(0.084)	
${}^{NASDAQ}STD\_DEV_{t-1}^{50/50}$										0.102
										(0.122)
Within R^2	0.163	0.191	0.185	0.163	0.154	0.147	0.132	0.238	0.233	0.218
Adj. R^2	0.420	0.440	0.435	0.420	0.519	0.516	0.507	0.430	0.426	0.415
$\% R^2$ Explained		6.79	4.89	0.00	5.54	3.98	0.81	15.95	14.73	10.38
Observations	672,385	672,385	672,385	672,385	355,881	355,881	355,881	316,504	316,504	316,504

Log turnover regressed on set of controls and disagreement constructed using stocks from either NASDAQ, NYSE/AMEX or both (ALL) stock exchanges. It also gives disagreement measure based on different splitting criterion. A 70/30 stock split assigns top 30% stocks to buy category, bottom 30% stocks to sell category and the rest to hold category. Similarly for 80/20 and 50/50 splits. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

IA.4.4 Univariate sorts

In Table IA.5, Panel A gives the changes in average turnover over deciles made on different disagreement measures. Majority of the decile changes are psoitive where out of the total 63 decile changes, 8 are negative and 27 are positive at 1% significance level. Across deciles, the evidence is strong in earlier disagreement deciles and mixed in later deciles.

Panel B present portfolio averages of different turnover measures with portfolios sorted on STD_DEV. Turnover is measured at time t while portfolios are made at time $t-1$. The effect of disagreement is strongest for adjusted turnover measures like L_TURN_GRT and L_TURN_ILLIQ. For other measures, the effect diminishes in later deciles. Overall, the broad majority of decile changes are positive and significant for all turnover measures. There is evidence that illiquidity also rises with disagreement (Figure 6, subplot 5). This can have a negative effect on turnover as disagreement rises.

Table IA.5: **Univariate Sorts: Different disagreement measures**

Panel A: Disagreement Measures

	L_TURN_t									
	D_1	D_2-D_1	D_3-D_2	D_4-D_3	D_5-D_4	D_6-D_5	D_7-D_6	D_8-D_7	D_9-D_8	$D_{10}-D_9$
STD_DEV_{t-1}	-0.921	0.041**	0.053**	0.065**	0.073**	0.080**	0.060**	0.047**	0.048**	0.134**
ABS_DEV_{t-1}	-0.975	0.067**	0.080**	0.066**	0.096**	0.073**	0.066**	0.040**	0.050**	0.129**
PC_DEV_{t-1}	-0.739	0.024**	0.022**	0.023**	0.032**	0.034**	0.033**	0.013*	0.013*	0.072**
$LO\_CORR\_SD_{t-1}$	-1.085	0.081**	0.213**	-0.035**	0.078**	0.028**	0.073**	0.115**	0.061**	-0.041**
$HI\_CORR\_SD_{t-1}$	-0.990	0.066**	0.143**	0.002	0.073**	0.077**	0.043**	0.040**	0.046**	-0.104**
NUM_FLIPS_{t-1}	-0.774	0.264**	-0.036*	-0.061**	-0.001	-0.018	-0.034	0.183**	0.195*	-0.364**
NUM_DIV_{t-1}	-0.837	0.148**	0.070**	0.018**	0.023**	-0.020*	0.019	-0.007	0.016	0.110**

Panel B: Turnover Measures

	Deciles made on STD_DEV_{t-1}									
	D_1	D_2-D_1	D_3-D_2	D_4-D_3	D_5-D_4	D_6-D_5	D_7-D_6	D_8-D_7	D_9-D_8	$D_{10}-D_9$
$L\_TURN\_1d_t$	0.474	0.062**	0.059**	0.063**	0.066**	0.071**	0.051**	0.038**	0.048**	0.130**
$L\_TURN\_5d_t$	2.176	0.038**	0.050**	0.066**	0.074**	0.080**	0.062**	0.045**	0.051**	0.151**
L_TURN_t	-0.921	0.041**	0.053**	0.065**	0.073**	0.080**	0.060**	0.047**	0.048**	0.134**
$L\_TURN\_GRT_t$	-0.932	0.030**	0.049**	0.062**	0.077**	0.088**	0.071**	0.058**	0.057**	0.110**
$L\_TURN\_D_t$	-0.036	0.005**	0.013**	0.010**	0.011**	0.006**	0.004	-0.002	0.002	0.022**
$L\_TURN\_ILLIQ_t$	-0.070	0.007**	0.014**	0.014**	0.016**	0.024**	0.025**	0.025**	0.027**	0.032**
$VW\_L\_TURN_t$	-0.048	0.019**	0.012**	0.008**	0.007**	0.010**	0.006**	-0.003	0.002	0.033**
$EW\_L\_TURN_t$	-0.027	0.006**	0.008**	0.005**	0.007**	0.008**	0.005*	-0.004	-0.002	0.031**

In Panel A, average log turnover is measured over univariate portfolio decile sorts of several measures of lagged disagreement. In Panel B, different measures of turnover are averaged over univariate decile sorts of STD_DEV_{t-1}. At each month, the cross-section of stocks is assigned to 10 portfolios based on the sorting variable. This procedure is repeated for each month. In Panel A, D_i is the average turnover in i^{th} disagreement decile, $D_j - D_i$ is the difference of average L_TURN in D_j and D_i. For Panel B, in the fourth row for instance, D_i is avergae GRT adjusted log turnover in the i^{th} STD_DEV_{t-1} decile, $D_j - D_i$ is the difference of average L_TURN_GRT in D_j and D_i. Corresponding signficance levels are from a t-test of sample means across corresponding decile pairs. Statistical significance of 5% and 1% are indicated by * and ** respectively

IA.5 Asymmetric Effect of Buy and Sell Signals

Does a buy signal affect investor's trading behavior in the same way as a sell signal? Prior research has shown the presence of a disposition effect which predicts that investors tend to keep losses and relaize gains. This phenomenon has two consequences: (i) the trading volume response following gains should be higher than the trading volume response following losses, and, (ii) the price variability would be subdued since investors hold (sell) stocks following negative (positive) shocks.

In previous section, I find contradictory evidence where negative price change has a higher magnitude volume response than positive price changes. From specification (4) in Table 7, the coefficient on RET^- is 35% higher in magnitude than the coefficient on RET^+. The base specification in Table 3 also offer same evidence.

Since positive and negative price change are viewed differently by the market with respect to trading volume, the natural question to ask is whether buy and sell trading signals are viewed differently by the market. In Panel A of Table IA.6, I examine the effect of increasing number of sell signals ($NSELL$) while keeping the number of buy ($NBUY$) signals fixed on trading volume and stock returns. In the data, $NBUY$ and $NSELL$ are positively correlated hence increasing one also increases the other. A portfolio of low $NBUY$ and high $NSELL$ would return very few observations. To overcome this, I perform two way dependent sorts first on $NBUY$ terciles and then on $NSELL$ terciles. This ensures that each portfolio has good number of observations. Panel B has the sorting order reversed and performed first on $NSELL$ and then on $NBUY$. We can compare the tercile changes across the two panels to understand the differential impact of increasing $NSELL$ whilst keeping $NBUY$ constant and vice-versa.

The results support the evidence in contradiction to disposition effect and trading volume changes more in Panel A than in Panel B. Within low $NBUY$ ($NSELL$) tercile, moving from medium to high $NSELL$ ($NBUY$) tercile changes volume by roughly 8 (3) percentiles. The effect on returns is similar and same $NSELL$ ($NBUY$) tercile change within low $NBUY$ ($NSELL$) tercile causes a -2.2 (0.3) percentile change in return rank. Another observation is that while sorting by $NBUYS$ and then $NSELL$ in Panel A, the turnover and return ranks are higher when we move from second to third $NSELL$ tercile. However, when sorting is done using $NSELL$ and then $NBUY$, they are higher when moving from first to second $NBUY$ tercile in Panel B. This means that the effect of increase in sell signals on turnover and returns is reflected more aggressively in moving from second to third tercile while for number of sell signals it happens in moving from first to second tercile.

In Table IA.7, I regress turnover and return ranks on $NBUY$ and $NSELL$ ranks. I do not include other regressors for brevity and the use of variable ranks[36] rather than levels should provide some relief against issues of spurious relationships. In Panel B, both the coeficients and R^2 provide the same evidence and turnover turns out to be twice as more sensitive (specification 3) to $NSELL$ than $NBUY$. In specification 4, the interaction of $NBUY$ and $NSELL$ is also significant since disagreement rises monotonically with $NBUY \times NSELL$ (specification 4 in Panel A). Return regressions in Panel C also point to the same evidence and $NSELL$ affects returns much more sharply than $NBUY$. The coefficient of interaction term in specification 5 is insignificant as

[36]Variable ranks are evaluated separately each month to counter spurious time-series relation between turnover and other variables.

expected since increase in both number of buy and sell signals should give an ambigous signal for returns.

Table IA.6: **Bivariate Sorts: Disagreement, turnover and return over terciles of number of buy/sell signals**

Panel A: 3×3 **dependent sort on** $NBUY$ **and then** $NSELL$

	NBUY_Ter$_{(1,.)}$			**NBUY_Ter$_{(2,.)}$**			**NBUY_Ter$_{(3,.)}$**		
	T_{11}	$T_{12}-T_{11}$	$T_{13}-T_{12}$	T_{21}	$T_{22}-T_{21}$	$T_{23}-T_{22}$	T_{31}	$T_{32}-T_{31}$	$T_{33}-T_{32}$
STD_DEV_R	29.264	18.274**	14.081**	35.440	14.504**	19.987**	NA	NANA	-2.115
$TURN\_R$	47.872	6.580**	8.024**	46.353	3.916**	2.656**	NA	NANA	-0.386
RET_R	48.374	-0.809**	-1.539	49.843	-0.391*	-0.661**	NA	NANA	-0.409
STD_DEV	0.679	0.074**	0.050**	0.716	0.049**	0.075**	NA	NANA	-0.007
$TURN$	-0.788	0.257**	0.260**	-0.793	0.141**	0.082**	NA	NANA	0.015
RET	10.874	1.710**	1.372	14.025	5.219**	7.867**	NA	NANA	-3.746
N	496953	150597	716	17669	1206128	173107	NA	219	89541

Panel B: 3×3 **dependent sort on** $NSELL$ **and then** $NBUY$

	NSELL_Ter$_{(1,.)}$			**NSELL_Ter$_{(2,.)}$**			**NSELL_Ter$_{(3,.)}$**		
	T_{11}	$T_{12}-T_{11}$	$T_{13}-T_{12}$	T_{21}	$T_{22}-T_{21}$	$T_{23}-T_{22}$	T_{31}	$T_{32}-T_{31}$	$T_{33}-T_{32}$
STD_DEV_R	20.411	9.423**	7.208**	35.036	29.488**	20.732**	36.661	29.756**	19.262**
$TURN\_R$	41.009	4.689**	2.449**	51.017	2.790**	-0.781**	60.586	2.092**	0.824**
RET_R	49.786	0.662**	0.206*	47.821	0.741**	-0.009	44.792	0.970**	0.559
STD_DEV	0.631	0.055**	0.029**	0.710	0.113**	0.078**	0.717	0.112**	0.073**
$TURN$	-1.081	0.249**	0.040**	-0.652	0.129**	-0.059**	-0.265	0.131**	-0.012
RET	14.053	4.077**	5.582**	9.185	10.571**	13.002**	1.272	10.308**	9.634**
N	179973	570078	116044	315181	733349	131491	31831	46894	10089

Average ranks of disagreement, turnover and return measured over bivariate 3×3 sorts. The first sorting variable is the one period lagged number of buy signals ($NBUY$) in panel A and lagged number of sell signals ($NSELL$) in panel B. The bivariate sorts are dependent sorts where second sorting is done on the portfolios from the first sort. The second level of sorting is performed using $NSELL$ in panel A and $NBUY$ in panel B respectively. Both dimensions of sorting are cross-sectional where a fresh sorting is performed each month. 70/30 portfolio breakpoints are used to make terciles. In panel A (panel B), $T_{i,j}$ is the average rank for i^{th} $NBUY$ ($NSELL$) tercile and j^{th} $NSELL$ ($NBUY$) tercile. First three row presents the ranks of lagged disagreement, current log turnover and current returns respectively. The next three rows present the levels. Last row gives the number of observations in each portfolio. Ranks are multiplied by 100, returns are multiplied by 1200 so they represent annual percentages. Turnover and disagreement are unscaled. In panel A (panel B), within a row, each $NBUY$ ($NSELL$) tercile is represented as a group of three values. There are three groups each corresponding to a different $NBUY$ ($NSELL$) tercile. The second and thrid entries in each group presents difference in average value over two successive terciles. $T_{i,j} - T_{i,j'}$ is the difference between $NSELL$ ($NBUY$) terciles j and j' within $NBUY$ ($NSELL$) tercile i. Corresponding signficance levels are from a t-test of sample means across corresponding tercile pairs. Statistical significance of 5% and 1% are indicated by * and ** respectively

IA.6 Information Environment Portfolio Splits

Below is a list of turnover regressions estimated on portfolio splits made in several firm characteristics related to the firm's information environment.

Table IA.7: **Number of buy/sell signals, disagreement, turnover and returns**

Panel A: Disagreement Rank

	(1)	(2)	(3)	(4)	(5)
$NBUY\_R_{t-1}$	0.561***		0.873***		-0.276***
	(0.001)		(0.001)		(0.001)
$NSELL\_R_{t-1}$		0.851***	1.068***		-0.174***
		(0.001)	(0.001)		(0.001)
$NBUY\_R_{t-1} \times NSELL\_R_{t-1}$				2.926***	3.468***
				(0.001)	(0.003)
Adj. R^2	0.131	0.353	0.646	0.779	0.787
Observations	2,134,930	2,134,930	2,134,930	2,134,930	2,134,930

Panel B: Turnover Rank

	(1)	(2)	(3)	(4)	(5)
$NBUY\_R_{t-1}$	0.011***		0.107***		0.195***
	(0.001)		(0.001)		(0.003)
$NSELL\_R_{t-1}$		0.303***	0.330***		0.425***
		(0.001)	(0.001)		(0.003)
$NBUY\_R_{t-1} \times NSELL\_R_{t-1}$				0.564***	-0.265***
				(0.003)	(0.007)
Adj. R^2	0.000	0.035	0.039	0.023	0.039
Observations	2,134,930	2,134,930	2,134,930	2,134,930	2,134,930

Panel C: Return Rank

	(1)	(2)	(3)	(4)	(5)
$NBUY\_R_{t-1}$	0.040***		0.018***		0.006**
	(0.001)		(0.001)		(0.003)
$NSELL\_R_{t-1}$		-0.079***	-0.075***		-0.087***
		(0.001)	(0.001)		(0.003)
$NBUY\_R_{t-1} \times NSELL\_R_{t-1}$				-0.088***	0.034***
				(0.003)	(0.007)
Adj. R^2	0.001	0.003	0.003	0.001	0.003
Observations	2,134,930	2,134,930	2,134,930	2,134,930	2,134,930

Lagged disagreement rank ($STD\_DEV\_R_{t-1}$), log turnover rank ($TURN\_R$) and return rank (RET_R) are regressed on all combinations of lagged number of buy and sell signal ranks ($NBUY\_R_{t-1}$ and $NSELL\_R_{t-1}$). Definitions of all the variables appears in Appendix A.2. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively

Table IA.8: **Monthly cross-sectional regression: Analyst following**

	L_TURN_t					
	$NUMEST \in \{2,3\}$		$NUMEST \in \{4\ldots10\}$		$NUMEST \geq 11$	
	(1)	(2)	(3)	(4)	(5)	(6)
$NASDAQ$	0.096***	0.085***	0.165***	0.142***	0.300***	0.270***
	(0.033)	(0.032)	(0.026)	(0.025)	(0.031)	(0.030)
RET^{+}_{t-1}	1.868***	1.688***	1.487***	1.293***	1.440***	1.268***
	(0.077)	(0.072)	(0.082)	(0.073)	(0.077)	(0.069)
RET^{-}_{t-1}	-2.375***	-2.127***	-2.239***	-1.971***	-2.214***	-1.977***
	(0.089)	(0.083)	(0.099)	(0.093)	(0.114)	(0.107)
LEV_{t-1}	0.668***	0.517***	0.872***	0.637***	0.897***	0.649***
	(0.108)	(0.107)	(0.121)	(0.119)	(0.144)	(0.137)
$CAPM\_\beta_{t-1}$	0.047	0.050	0.076**	0.075**	0.054	0.053
	(0.045)	(0.043)	(0.039)	(0.036)	(0.044)	(0.043)
BTM_{t-1}	-1.227***	-0.946***	-0.583**	-0.264	0.698***	1.098***
	(0.186)	(0.182)	(0.247)	(0.263)	(0.265)	(0.256)
L_PRC_{t-1}	0.207***	0.256***	0.151***	0.179***	-0.017	-0.013
	(0.018)	(0.018)	(0.015)	(0.015)	(0.017)	(0.016)
L_FAGE_{t-1}	-0.088***	-0.063***	-0.092***	-0.057***	-0.194***	-0.162***
	(0.023)	(0.023)	(0.016)	(0.015)	(0.018)	(0.017)
$ESURP_{t-1}$	0.509***	0.442***	0.354***	0.293***	0.384***	0.325***
	(0.075)	(0.066)	(0.100)	(0.085)	(0.106)	(0.096)
$EVOL_{t-1}$	0.554***	0.321***	0.628***	0.399***	0.716***	0.329
	(0.112)	(0.099)	(0.118)	(0.104)	(0.231)	(0.216)
$NUMEST_{t-1}$	0.260***	0.259***	0.065***	0.065***	0.007***	0.006***
	(0.015)	(0.015)	(0.003)	(0.003)	(0.002)	(0.002)
$FDISP_{t-1}$	0.046***	0.030***	0.043***	0.023***	0.061***	0.043***
	(0.011)	(0.011)	(0.008)	(0.008)	(0.013)	(0.013)
STD_DEV_{t-1}		1.652***		1.569***		1.198***
		(0.092)		(0.084)		(0.085)
Within R^2	0.105	0.127	0.105	0.137	0.173	0.198
Adj. R^2	0.299	0.316	0.452	0.471	0.591	0.603
$\% R^2$ Explained		2.46		3.49		2.93
Observations	209,863	209,863	261,782	261,782	200,740	200,740

Log turnover regressed on set of controls and several measures of disagreement across three groups based on number of analyst following the stock ($NUMEST$). Observations with $NUMEST = 1$ have been deleted. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table IA.9: **Monthly cross-sectional regression: Firm Age terciles**

	L_TURN_t					
	$FIRM\_AGE-1$		$FIRM\_AGE-2$		$FIRM\_AGE-3$	
	(1)	(2)	(3)	(4)	(5)	(6)
$NASDAQ$	0.252***	0.222***	0.095***	0.083***	-0.055	-0.070
	(0.031)	(0.030)	(0.029)	(0.029)	(0.051)	(0.050)
RET^{+}_{t-1}	1.503***	1.328***	1.734***	1.578***	1.643***	1.502***
	(0.067)	(0.059)	(0.078)	(0.072)	(0.086)	(0.082)
RET^{-}_{t-1}	-2.408***	-2.140***	-2.517***	-2.305***	-2.285***	-2.121***
	(0.077)	(0.070)	(0.103)	(0.098)	(0.148)	(0.141)
LEV_{t-1}	0.418***	0.225**	0.891***	0.741***	1.530***	1.333***
	(0.111)	(0.108)	(0.136)	(0.136)	(0.235)	(0.229)
$CAPM\_\beta_{t-1}$	0.083**	0.085**	0.082**	0.085**	0.057	0.053
	(0.042)	(0.040)	(0.040)	(0.038)	(0.042)	(0.040)
BTM_{t-1}	-0.991***	-0.630***	-1.232***	-0.971***	0.115	0.288
	(0.181)	(0.174)	(0.245)	(0.253)	(0.319)	(0.314)
L_PRC_{t-1}	0.248***	0.281***	0.153***	0.177***	0.093***	0.113***
	(0.016)	(0.016)	(0.017)	(0.017)	(0.026)	(0.025)
L_FAGE_{t-1}	-0.005	0.025	-0.288***	-0.261***	0.034	0.056
	(0.026)	(0.026)	(0.037)	(0.037)	(0.118)	(0.116)
$ESURP_{t-1}$	0.516***	0.438***	0.454***	0.407***	0.551***	0.501***
	(0.071)	(0.063)	(0.089)	(0.079)	(0.108)	(0.102)
$EVOL_{t-1}$	0.532***	0.308***	0.555***	0.337***	-0.117	-0.338
	(0.105)	(0.097)	(0.139)	(0.131)	(0.218)	(0.209)
$NUMEST_{t-1}$	0.047***	0.045***	0.039***	0.038***	0.023***	0.023***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$FDISP_{t-1}$	0.018*	0.003	0.045***	0.030***	0.053***	0.034**
	(0.011)	(0.010)	(0.011)	(0.011)	(0.015)	(0.015)
STD_DEV_{t-1}		1.754***		1.230***		0.969***
		(0.097)		(0.095)		(0.099)
Within R^2	0.217	0.240	0.184	0.198	0.139	0.153
Adj. R^2	0.404	0.422	0.454	0.464	0.512	0.520
$\% R^2$ Explained		3.01		1.73		1.66
Observations	189,863	189,863	288,297	288,297	194,225	194,225

Log turnover regressed on set of controls and several measures of disagreement across three 70/30 terciles made on log of firm age (L_FAGE). All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. $\% R^2$ explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table IA.10: **Monthly cross-sectional regression: Document size terciles**

	L_TURN_t					
	DOC-SIZE-1		*DOC-SIZE-2*		*DOC-SIZE-3*	
	(1)	(2)	(3)	(4)	(5)	(6)
$NASDAQ$	0.096***	0.080**	0.094***	0.063**	0.012	-0.010
	(0.036)	(0.036)	(0.027)	(0.026)	(0.036)	(0.035)
RET^{+}_{t-1}	1.594***	1.455***	1.535***	1.371***	1.590***	1.386***
	(0.075)	(0.069)	(0.073)	(0.066)	(0.082)	(0.071)
RET^{-}_{t-1}	-2.672***	-2.478***	-2.496***	-2.261***	-2.618***	-2.342***
	(0.097)	(0.094)	(0.099)	(0.092)	(0.113)	(0.105)
LEV_{t-1}	0.663***	0.493**	0.445***	0.297**	0.542***	0.341***
	(0.212)	(0.213)	(0.130)	(0.126)	(0.131)	(0.127)
$CAPM\_\beta_{t-1}$	0.080*	0.080*	0.079*	0.080**	0.088**	0.089**
	(0.047)	(0.045)	(0.041)	(0.039)	(0.042)	(0.040)
BTM_{t-1}	-1.911***	-1.646***	-0.982***	-0.669***	-0.383*	-0.163
	(0.322)	(0.334)	(0.213)	(0.207)	(0.227)	(0.221)
L_PRC_{t-1}	0.238***	0.257***	0.182***	0.211***	0.100***	0.135***
	(0.019)	(0.019)	(0.015)	(0.016)	(0.021)	(0.021)
L_FAGE_{t-1}	-0.180***	-0.151***	-0.120***	-0.092***	-0.082***	-0.054***
	(0.025)	(0.025)	(0.017)	(0.016)	(0.021)	(0.020)
$ESURP_{t-1}$	0.702***	0.639***	0.392***	0.344***	0.430***	0.376***
	(0.097)	(0.093)	(0.082)	(0.071)	(0.072)	(0.067)
$EVOL_{t-1}$	0.480***	0.282	0.293**	0.069	0.276**	0.077
	(0.180)	(0.172)	(0.130)	(0.125)	(0.118)	(0.116)
$NUMEST_{t-1}$	0.040***	0.039***	0.032***	0.031***	0.030***	0.030***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
$FDISP_{t-1}$	0.016	0.005	0.028***	0.013	0.039***	0.018
	(0.015)	(0.014)	(0.011)	(0.010)	(0.013)	(0.013)
STD_DEV_{t-1}		1.202***		1.412***		1.719***
		(0.112)		(0.087)		(0.095)
Within R^2	0.229	0.241	0.167	0.188	0.140	0.170
Adj. R^2	0.410	0.420	0.403	0.418	0.380	0.402
$\%R^2$ Explained		1.64		2.51		3.53
Observations	126,386	126,386	192,143	192,143	165,960	165,960

Log turnover regressed on set of controls and several measures of disagreement across three 70/30 terciles made on document size of annual (10-K) reports. Document size is the size in megabytes of the raw EDGAR filing. Document size terciles are made every month. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table IA.11: **Monthly cross-sectional regression: Report length terciles**

	L_TURN_t					
	LENGTH-1		*LENGTH-2*		*LENGTH-3*	
	(1)	(2)	(3)	(4)	(5)	(6)
$NASDAQ$	0.101***	0.078**	0.084***	0.055*	0.083**	0.062**
	(0.035)	(0.035)	(0.029)	(0.029)	(0.032)	(0.031)
RET^{+}_{t-1}	1.674***	1.528***	1.513***	1.353***	1.500***	1.317***
	(0.080)	(0.075)	(0.069)	(0.061)	(0.078)	(0.069)
RET^{-}_{t-1}	-2.852***	-2.656***	-2.514***	-2.286***	-2.390***	-2.131***
	(0.107)	(0.103)	(0.101)	(0.094)	(0.103)	(0.095)
LEV_{t-1}	0.831***	0.650***	0.310**	0.160	0.513***	0.337***
	(0.233)	(0.233)	(0.138)	(0.136)	(0.126)	(0.120)
$CAPM\_\beta_{t-1}$	0.039	0.039	0.093**	0.095**	0.109***	0.106***
	(0.048)	(0.046)	(0.041)	(0.039)	(0.039)	(0.037)
BTM_{t-1}	-2.147***	-1.882***	-1.004***	-0.699***	-0.166	0.066
	(0.346)	(0.360)	(0.212)	(0.206)	(0.219)	(0.212)
L_PRC_{t-1}	0.235***	0.253***	0.187***	0.218***	0.114***	0.150***
	(0.020)	(0.021)	(0.016)	(0.016)	(0.019)	(0.019)
L_FAGE_{t-1}	-0.158***	-0.131***	-0.089***	-0.061***	-0.105***	-0.077***
	(0.026)	(0.026)	(0.017)	(0.017)	(0.020)	(0.019)
$ESURP_{t-1}$	0.657***	0.613***	0.462***	0.397***	0.370***	0.326***
	(0.122)	(0.113)	(0.079)	(0.070)	(0.069)	(0.064)
$EVOL_{t-1}$	0.397*	0.165	0.290**	0.096	0.225**	0.023
	(0.238)	(0.231)	(0.123)	(0.120)	(0.114)	(0.115)
$NUMEST_{t-1}$	0.037***	0.036***	0.030***	0.030***	0.027***	0.027***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
$FDISP_{t-1}$	0.025	0.010	0.021**	0.008	0.034***	0.017
	(0.016)	(0.016)	(0.010)	(0.010)	(0.012)	(0.011)
STD_DEV_{t-1}		1.175***		1.414***		1.622***
		(0.113)		(0.091)		(0.094)
Within R^2	0.218	0.231	0.160	0.180	0.134	0.163
Adj. R^2	0.407	0.417	0.404	0.419	0.374	0.396
$\%R^2$ Explained		1.61		2.43		3.37
Observations	133,882	133,882	195,599	195,599	155,008	155,008

Log turnover regressed on set of controls and several measures of disagreement across three 70/30 terciles made on length of annual (10-K) reports. Length of the report is measured by number of words in the report. Report length terciles are made every month. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table IA.12: **Monthly cross-sectional regression: Unique word count terciles**

	L_TURN_t					
	UNQ-WORDS-1		*UNQ-WORDS-2*		*UNQ-WORDS-3*	
	(1)	(2)	(3)	(4)	(5)	(6)
$NASDAQ$	0.065*	0.042	0.069**	0.044	0.120***	0.091***
	(0.036)	(0.036)	(0.028)	(0.028)	(0.034)	(0.033)
RET^{+}_{t-1}	1.685***	1.531***	1.567***	1.398***	1.425***	1.258***
	(0.084)	(0.077)	(0.068)	(0.060)	(0.081)	(0.073)
RET^{-}_{t-1}	-2.862***	-2.657***	-2.539***	-2.305***	-2.333***	-2.085***
	(0.113)	(0.109)	(0.097)	(0.091)	(0.103)	(0.096)
LEV_{t-1}	0.745***	0.554***	0.514***	0.369***	0.607***	0.416***
	(0.198)	(0.199)	(0.129)	(0.126)	(0.136)	(0.130)
$CAPM\_\beta_{t-1}$	0.061	0.061	0.078*	0.079**	0.119***	0.117***
	(0.050)	(0.048)	(0.040)	(0.038)	(0.040)	(0.038)
BTM_{t-1}	-1.935***	-1.690***	-0.847***	-0.546***	-0.293	-0.049
	(0.348)	(0.361)	(0.197)	(0.192)	(0.235)	(0.229)
L_PRC_{t-1}	0.223***	0.241***	0.176***	0.207***	0.143***	0.180***
	(0.021)	(0.022)	(0.016)	(0.016)	(0.020)	(0.020)
L_FAGE_{t-1}	-0.145***	-0.118***	-0.097***	-0.069***	-0.110***	-0.080***
	(0.027)	(0.027)	(0.017)	(0.016)	(0.021)	(0.020)
$ESURP_{t-1}$	0.684***	0.632***	0.458***	0.399***	0.303***	0.256***
	(0.109)	(0.104)	(0.076)	(0.067)	(0.080)	(0.073)
$EVOL_{t-1}$	0.294	0.083	0.302***	0.094	0.297**	0.112
	(0.182)	(0.184)	(0.112)	(0.108)	(0.121)	(0.120)
$NUMEST_{t-1}$	0.039***	0.038***	0.033***	0.032***	0.023***	0.023***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
$FDISP_{t-1}$	0.022	0.007	0.022**	0.006	0.034***	0.020
	(0.017)	(0.016)	(0.010)	(0.010)	(0.013)	(0.012)
STD_DEV_{t-1}		1.241***		1.424***		1.620***
		(0.118)		(0.084)		(0.104)
Within R^2	0.208	0.222	0.164	0.185	0.129	0.159
Adj. R^2	0.414	0.425	0.387	0.402	0.352	0.374
$\% R^2$ Explained		1.76		2.45		3.41
Observations	124,967	124,967	246,342	246,342	113,180	113,180

Log turnover regressed on set of controls and several measures of disagreement across three 70/30 terciles made on number of unique words in annual (10-K) reports. Terciles on number of unique words are made every month. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table IA.13: **Monthly cross-sectional regression: Number of segments**

	L_TURN_t							
	TOT-SEG = 1		*TOT-SEG > 1*		*UNQ-SEG = 1*		*UNQ-SEG > 1*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$NASDAQ$	0.286***	0.265***	0.030	0.010	0.231***	0.208***	0.028	0.010
	(0.031)	(0.030)	(0.030)	(0.030)	(0.028)	(0.028)	(0.035)	(0.035)
RET^{+}_{t-1}	1.594***	1.444***	1.583***	1.417***	1.574***	1.426***	1.634***	1.459***
	(0.065)	(0.059)	(0.079)	(0.071)	(0.066)	(0.060)	(0.080)	(0.072)
RET^{-}_{t-1}	-2.417***	-2.198***	-2.450***	-2.236***	-2.429***	-2.217***	-2.467***	-2.245***
	(0.084)	(0.078)	(0.107)	(0.101)	(0.080)	(0.075)	(0.121)	(0.113)
LEV_{t-1}	0.609***	0.431***	0.852***	0.669***	0.591***	0.422***	0.970***	0.782***
	(0.129)	(0.127)	(0.140)	(0.137)	(0.113)	(0.111)	(0.171)	(0.166)
$CAPM\_\beta_{t-1}$	0.070*	0.073**	0.085**	0.080**	0.082**	0.085**	0.061	0.055
	(0.039)	(0.037)	(0.039)	(0.038)	(0.036)	(0.035)	(0.043)	(0.041)
BTM_{t-1}	-1.390***	-1.024***	-0.539**	-0.349	-1.306***	-0.957***	-0.545**	-0.368
	(0.213)	(0.207)	(0.236)	(0.236)	(0.197)	(0.191)	(0.275)	(0.276)
L_PRC_{t-1}	0.194***	0.219***	0.179***	0.206***	0.199***	0.226***	0.158***	0.185***
	(0.016)	(0.016)	(0.018)	(0.019)	(0.015)	(0.015)	(0.021)	(0.022)
L_FAGE_{t-1}	-0.199***	-0.169***	-0.136***	-0.114***	-0.191***	-0.163***	-0.133***	-0.110***
	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)	(0.018)	(0.022)	(0.021)
$ESURP_{t-1}$	0.521***	0.446***	0.486***	0.436***	0.515***	0.446***	0.529***	0.476***
	(0.083)	(0.073)	(0.074)	(0.067)	(0.076)	(0.067)	(0.085)	(0.077)
$EVOL_{t-1}$	0.584***	0.393***	0.235*	0.033	0.575***	0.392***	0.093	-0.118
	(0.115)	(0.101)	(0.127)	(0.121)	(0.108)	(0.095)	(0.144)	(0.142)
$NUMEST_{t-1}$	0.038***	0.037***	0.025***	0.025***	0.036***	0.035***	0.024***	0.024***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$FDISP_{t-1}$	0.027**	0.013	0.044***	0.024***	0.029***	0.016*	0.045***	0.023**
	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)
STD_DEV_{t-1}		1.413***		1.265***		1.355***		1.283***
		(0.090)		(0.084)		(0.082)		(0.093)
Within R^2	0.204	0.221	0.139	0.159	0.198	0.215	0.126	0.147
Adj. R^2	0.427	0.439	0.445	0.458	0.418	0.430	0.454	0.467
$\% R^2$ Explained		2.16		2.29		2.07		2.38
Observations	288,449	288,449	313,694	313,694	353,849	353,849	248,294	248,294

Log turnover regressed on set of controls and several measures of disagreement across two groups made using number of segments. For each group, first column is for single segments firms while the second column is for multi segment firms. $UNQ-SEG$ is the total number of unique segments by industry wherein all segments within an industry are counted as one segment. $TOT-SEG$ is the total number of segments irrepsective of how many segments are present in each industry. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. $\% R^2$ explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table IA.14: **Monthly cross-sectional regression: Complex word counts terciles**

	L_TURN_t					
	COMPLEX-WORDS-1		*COMPLEX-WORDS-2*		*COMPLEX-WORDS-3*	
	(1)	(2)	(3)	(4)	(5)	(6)
$NASDAQ$	0.123***	0.109**	0.048*	0.022	0.262***	0.222***
	(0.046)	(0.045)	(0.026)	(0.026)	(0.047)	(0.045)
RET^{+}_{t-1}	1.795***	1.622***	1.539***	1.380***	1.260***	1.097***
	(0.088)	(0.083)	(0.073)	(0.066)	(0.096)	(0.086)
RET^{-}_{t-1}	-2.956***	-2.717***	-2.544***	-2.323***	-1.987***	-1.744***
	(0.136)	(0.129)	(0.098)	(0.092)	(0.116)	(0.106)
LEV_{t-1}	0.667**	0.447	0.589***	0.436***	0.768***	0.573***
	(0.314)	(0.305)	(0.117)	(0.115)	(0.169)	(0.157)
$CAPM\_\beta_{t-1}$	0.059	0.060	0.082**	0.082**	0.080*	0.073*
	(0.051)	(0.048)	(0.041)	(0.039)	(0.042)	(0.040)
BTM_{t-1}	-1.842***	-1.571***	-0.897***	-0.635***	0.347	0.638*
	(0.433)	(0.425)	(0.192)	(0.196)	(0.341)	(0.328)
L_PRC_{t-1}	0.243***	0.254***	0.178***	0.206***	0.147***	0.189***
	(0.032)	(0.032)	(0.015)	(0.015)	(0.027)	(0.027)
L_FAGE_{t-1}	-0.140***	-0.109***	-0.113***	-0.086***	-0.096***	-0.073***
	(0.036)	(0.036)	(0.015)	(0.015)	(0.029)	(0.027)
$ESURP_{t-1}$	0.878***	0.811***	0.432***	0.383***	0.294***	0.230***
	(0.206)	(0.192)	(0.072)	(0.064)	(0.086)	(0.085)
$EVOL_{t-1}$	0.722***	0.418*	0.304***	0.111	0.157	0.023
	(0.254)	(0.254)	(0.098)	(0.096)	(0.141)	(0.141)
$NUMEST_{t-1}$	0.043***	0.043***	0.032***	0.032***	0.015***	0.016***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
$FDISP_{t-1}$	0.034*	0.015	0.021**	0.007	0.013	-0.003
	(0.020)	(0.019)	(0.009)	(0.009)	(0.018)	(0.018)
STD_DEV_{t-1}		1.357***		1.366***		1.688***
		(0.149)		(0.078)		(0.142)
Within R^2	0.213	0.229	0.171	0.190	0.116	0.152
Adj. R^2	0.456	0.467	0.364	0.378	0.326	0.353
$\%R^2$ Explained		2.02		2.28		4.06
Observations	65,530	65,530	359,733	359,733	42,829	42,829

Log turnover regressed on set of controls and several measures of disagreement across three 70/30 terciles made on unique number of complex words in annual (10-K) reports. Complex words are defined by Loughran and McDonald (2020). Terciles on complex word counts are made every month. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table IA.15: **Monthly cross-sectional regression: Institutional ownership terciles**

	L_TURN_t					
	$OWN(\%) \in [0, 0.5]$		$OWN(\%) \in (0.5, 0.75)$		$OWN(\%) \in [0.75, 100]$	
	(1)	(2)	(3)	(4)	(5)	(6)
$NASDAQ$	0.294***	0.286***	0.226***	0.202***	0.110***	0.073***
	(0.029)	(0.029)	(0.030)	(0.029)	(0.024)	(0.023)
RET^+_{t-1}	1.924***	1.748***	1.571***	1.389***	1.251***	1.063***
	(0.076)	(0.070)	(0.075)	(0.068)	(0.086)	(0.074)
RET^-_{t-1}	-2.291***	-2.048***	-2.464***	-2.217***	-2.216***	-1.958***
	(0.110)	(0.103)	(0.101)	(0.094)	(0.125)	(0.116)
LEV_{t-1}	0.656***	0.496***	0.707***	0.516***	0.855***	0.602***
	(0.110)	(0.110)	(0.123)	(0.118)	(0.170)	(0.164)
$CAPM\_\beta_{t-1}$	0.022	0.031	0.048	0.045	0.087*	0.085*
	(0.048)	(0.046)	(0.036)	(0.035)	(0.050)	(0.047)
BTM_{t-1}	-1.299***	-1.043***	-0.965***	-0.671***	-0.712***	-0.322
	(0.200)	(0.205)	(0.220)	(0.213)	(0.224)	(0.214)
L_PRC_{t-1}	0.063***	0.108***	0.147***	0.168***	0.057***	0.072***
	(0.017)	(0.017)	(0.018)	(0.018)	(0.016)	(0.016)
L_FAGE_{t-1}	-0.087***	-0.065***	-0.156***	-0.116***	-0.116***	-0.088***
	(0.022)	(0.022)	(0.018)	(0.018)	(0.016)	(0.015)
$ESURP_{t-1}$	0.600***	0.531***	0.425***	0.378***	0.410***	0.342***
	(0.072)	(0.067)	(0.092)	(0.083)	(0.091)	(0.077)
$EVOL_{t-1}$	0.392***	0.155	0.377***	0.156	0.581***	0.309**
	(0.114)	(0.114)	(0.135)	(0.129)	(0.162)	(0.132)
$NUMEST_{t-1}$	0.038***	0.037***	0.021***	0.020***	0.027***	0.026***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
$FDISP_{t-1}$	0.051***	0.036***	0.029***	0.010	0.037***	0.017*
	(0.010)	(0.010)	(0.011)	(0.011)	(0.010)	(0.009)
STD_DEV_{t-1}		1.564***		1.398***		1.432***
		(0.095)		(0.080)		(0.072)
Within R^2	0.123	0.143	0.140	0.165	0.149	0.184
Adj. R^2	0.262	0.280	0.331	0.350	0.296	0.325
$\% R^2$ Explained		2.32		2.95		4.08
Observations	253,066	253,066	202,264	202,264	206,571	206,571

Log turnover regressed on set of controls and several measures of disagreement across three portfolios of 13-F institutional ownership concentration. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.

Table IA.16: **Monthly cross-sectional regression: Time since 10-K filing**

	L_TURN_t			
	$s \in \{1,2,3\}$	$s \in \{4,5,6\}$	$s \in \{7,8,9\}$	$s \geq 10$
	(1)	(2)	(3)	(4)
$NASDAQ$	0.029	0.020	0.013	0.106***
	(0.027)	(0.028)	(0.028)	(0.032)
RET^{+}_{t-1}	1.330***	1.549***	1.425***	1.375***
	(0.125)	(0.079)	(0.072)	(0.084)
RET^{-}_{t-1}	-2.394***	-2.490***	-2.371***	-2.322***
	(0.096)	(0.104)	(0.152)	(0.091)
LEV_{t-1}	0.629***	0.555***	0.507***	0.504***
	(0.116)	(0.113)	(0.114)	(0.122)
$CAPM\_\beta_{t-1}$	0.152***	0.036	0.195***	0.003
	(0.056)	(0.060)	(0.068)	(0.074)
BTM_{t-1}	-0.761***	-0.838***	-0.643***	-0.606***
	(0.212)	(0.218)	(0.200)	(0.208)
L_PRC_{t-1}	0.208***	0.222***	0.202***	0.237***
	(0.017)	(0.019)	(0.018)	(0.020)
L_FAGE_{t-1}	-0.114***	-0.099***	-0.099***	-0.101***
	(0.016)	(0.016)	(0.016)	(0.017)
$ESURP_{t-1}$	0.305***	0.584***	0.506***	0.420***
	(0.092)	(0.085)	(0.096)	(0.098)
$EVOL_{t-1}$	0.089	0.148*	0.221**	0.175
	(0.104)	(0.078)	(0.113)	(0.149)
$NUMEST_{t-1}$	0.033***	0.034***	0.035***	0.035***
	(0.002)	(0.002)	(0.002)	(0.002)
$FDISP_{t-1}$	0.022***	0.012	0.001	0.014
	(0.008)	(0.010)	(0.013)	(0.012)
STD_DEV_{t-1}	1.441***	1.459***	1.504***	1.599***
	(0.079)	(0.077)	(0.075)	(0.086)
Within R^2	0.203	0.209	0.205	0.209
Adj. R^2	0.393	0.395	0.390	0.394
$\% R^2$ Explained	2.52	2.51	2.67	2.93
Observations	111,971	113,467	114,466	115,177

Log turnover regressed on controls and disagreement across four samples based on time (in months) since the last 10-K filing. s is the number of months since the last filing. All independent variables are one month lagged variables. Definitions of all the variables appears in Appendix A.2. All regression specifications have industry and year fixed effects. Within R^2 is the R^2 for within groups variation with fixed effects projected out while Adj. R^2 is for the entire model. % R^2 explained is the percent of unexplained variation explained by including disagreement in the regression. Standard errors are double clustered by firm and year-month. Statistical significance of 10%, 5% and 1% are indicated by *, ** and *** respectively.