The Effects of Investors' Information Acquisition On Sell-Side Analysts' Forecast Bias

Astaiza-Gómez, José Gabriel

2022

The Effects of Investors' Information Acquisition On Sell-Side Analysts' Forecast Bias

Abstract: In this research I empirically study the effects of information acquisition by investors or traders on analysts' forecast bias. Based on the theoretical literature on sell-side analysts, I argue that forecast bias is correlated to investors' information gathering in two opposite directions. On the one hand, higher levels of reading activities about individual firms by investors induce analysts to issue more optimistic forecasts if the potential for trading is higher. On the other hand, higher levels of reading activities about individual firms by investors help them identify opportunistic behaviors and thus to discipline analysts. I find that investors' information acquisition is positively related to analysts' optimism when the potential for trading is larger, and negatively related to optimism when investors are more likely to identify inflated forecasts. Together, these results suggest that information acquisition is not only correlated to analysts' optimism but also that its effect does not work trivially and solely in one direction but it activates two different incentives in analysts' decisions.

Key words: trading incentives, analyst's credibility, responsive investors, naive investors, experts' preferences, optimism, institutional investors, private information, public signals, firm's prospects.
JEL codes: M21, M37, G11, G14, G23, G24.

1 Introduction

In many situations, decision makers turn to experts for information and advice. However, experts may engage in opportunistic reporting (Meng, 2014) as their preferences or incentives are often not perfectly aligned with those of decision makers. In particular, sell-side analysts tend to produce more inflated forecasts (DeBondt and Thaler, 1990) in order to induce more trading and thus more income for their brokerage houses (Jackson, 2005). In this paper, I empirically test the hypothesis that the amount of firm-level information collected by investors has two opposite potential effects on analyst forecast bias. I argue that, the analyst acts strategically so as to take advantage of investor heterogeneity i.e. of the simultaneous presence of naive and sophisticated investors, but is restrained from systematically inducing a greater number of trades by making over optimistic forecasts because he also has to address reputation effects especially from sophisticated investors.

There are differences in investors' ability to establish the quality of analysts' reports, and thus, investors react rather heterogeneously to the reports issued by analysts. Particularly, small investors trade more in response to optimistic reports than to pessimistic reports (Mikhail, Walther and Willis, 2007), while institutional investors trade more in response to conservative analysts (Hugon and Muslu, 2010) and assign votes to more accurate analysts in the *Institutional Investor*'s All-American Research Team ranking (Groysberg, Healy and Maber, 2011; Stickel, 1992). This heterogeneity in reaction to reports, has been already incorporated in the theoretical literature, showing that sell-side analysts should react strategically to investors' responses to their reports (Fischer and Stocken, 2010; Kartik, Ottaviani and Squintani, 2007). In particular, Kartik, Ottaviani and Squintani (2007) (hereafter KOS) study if an analyst incur in opportunistic behaviors when there exist investors who are heterogeneous in their ability to establish the quality of analysts' reports, and show that the presence of some naive investors induce inflated forecasts.

To study the relation between heterogeneous sophistication and inflated forecasts, KOS propose a cheap-talk model where an analyst has private information about the performance of a firm upon which he issues a report, and the state of nature decides on the true state of the world. Since it is not the main purpose of this model to study the analyst's decision on the quality of his private information but to formalize the idea that the presence of some naive investors induce inflated forecasts, it is assumed in the model that the private information of the analyst, equals the true state of the world regarding the firm's performance and that there is not a public signal. The analyst is interested in the average response of a pool of investors and there is a fraction of naive investors who cannot formulate equilibrium beliefs about the true state of the world from the analyst report, but they formulate a dis-equilibrium estimate of the true state of the world. Meanwhile, the fraction of strategic investors, formulate an equilibrium estimate and take action accordingly.

Lemma 1 and Theorem 1 of KOS state that there are conditions for which there exists an equilibrium where the analyst sends an inflated report (higher than the true state of the world) in order to induce a larger response from naive investors, even when among investors there are sophisticated ones. In other words, it is feasible a situation in which there are optimistic reports which "still reveal precise information to strategic receivers, while deceiving naive receivers." When investors are more responsive to forecasts, this model tells us that analysts will issue more optimistic reports, and thus, I argue that greater information acquisition on stocks with greater trading potential leads to higher analyst optimism.

The cheap talk literature has studied how concerns about the future, affect communication, including reputational and career concerns (e.g. Ottaviani and Sørensen, 2006; Ely and Välimäki, 2003; Jullien and Park, 2014). However, here the question is if information acquisition by investors plays a significant role in limiting opportunistic reporting. Fischer and Stocken (2010) (hereafter FS) show that, when the quality or precision of the analyst report is common knowledge, higher investors' information acquisition induces more precise analysts' reports. Unfortunately, investors do not have a precise idea about analyst's forecast precision, but they can establish when analysts' reports are consistent with the firm's performance. Therefore, greater information acquisition on stocks for which investors are able to recognize forecast quality, should be associated to more precise - less biased analyst forecasts.

Letting the quality of the report being of common knowledge in the model of FS, provides results on how communication incentives affect the analyst's decisions on the quality of his private information upon which he issues a report. In the cheap-talk model of investors' information acquisition and forecast precision with truthful communication proposed by FS, there is an analyst who issues a report with a binary outcome, *bad* or *good*, upon his private information about the performance of a firm. Then, an investor takes action, based on both the analyst' report and a public signal (e.g. a corporate report) about the state of the world regarding the firm's performance. In this setting, the authors show that an increase in the informativeness of investors' signals, requires that the analyst augments the quality of his private information in order to maintain the investor responsive to the analyst's report[1].

In this paper, I empirically test the hypothesis that the amount of firm-level information collected by investors has two opposite potential effects on analyst forecast bias. First, analysts may issue more optimistic (inflated) forecasts for stocks to which investors are paying more attention when there is greater potential to generate trading. Second, analysts may issue less inflated forecasts to investors who have sought for more firm-level information when they have a greater ability to establish whether the analysts'

[1] This is Corollary 6 of FS.

forecasts are consistent with the firm's prospects or consistent with opportunistic behavior, so that the higher the acquisition of information at the firm level by investors, the lower the optimism in analysts' forecasts. In other words, the amount of firm-level information collected by investors is important to discipline analysts. I find that forecast bias is higher when firm-level information acquisition by investors increases, for stocks with more potential for trading businesses. Moreover, forecast bias decreases when investors acquire more firm-level information, for stocks followed by investors with a greater potential of identifying inflated forecasts. Thus, this research highlights that higher investors' reading activities are not related to forecast bias in one exclusive direction, and may deteriorate or improve the decision-making process of naive investors. Also, this study shows that the effects of investors' information acquisition is not trivial as one would expect: investor sophistication and information acquisition are not equivalent concepts, and the mechanism through which the differences in investor sophistication induce greater forecast bias, is information acquisition by less sophisticated investors.

This paper adds to the empirical literature studying analyst's trade-offs of career concerns and short-term benefits (e.g. Fang and Yasuda, 2009) and is also related with research on private investor's information acquisition (e.g. Ben-Rephael, Da and Israelsen, 2017; Chi and Shanthikumar, 2018). Furthermore, this paper adds to research studying the role of institutional investors on forecast accuracy (e.g. Ljungqvist et al., 2007). To study investors with a greater ability to establish when analysts' reports are consistent with the firm's performance, and following the research showing that institutional investors are more sophisticated (Hilary and Hsu, 2013; Boehmer and Kelley, 2009), I use stocks in the holdings of investment managers with more than 100 million USD in equity under management. Furthermore, I use firms in the Financials sector, which I argue, issue larger amounts of hard information. Also, I measure information acquisition by investors based on the level of activity at Bloomberg Terminals, calculating the quarterly changes of daily averages the "News Heat - Daily Max Readership" index of Bloomberg. In order to analyze stocks with a greater potential of trading, I use firms in the Consumer Goods sector, which I argue, are more likely to call the attention of a wider public and have characteristics that match a story of undervaluation in analysts' reports, thus stimulating the interest of a large number of investors. In addition, as non-robot EDGAR users tend to be retail investors (Chi and Shanthikumar, 2018; Loughran and McDonald, 2017; Asthana, Balsam and Sankaraguruswamy, 2004), and the FINRA Foundation's survey results show that many users of open access sources of financial information are likely to be naive, which is in line with the fact that small investors are more responsive to optimistic reports (Mikhail, Walther and Willis, 2007), studying the changes in searches at EDGAR for these firms, allows me to analyze the differential influence of investors' information acquisition on analyst forecast bias, for investors with a higher potential to generate trading. Therefore, I also calculate the quarterly changes of the number of non-robot downloads of EDGAR filings for each stock.

There are six sections in this paper including the introduction. In section two I expose the theoretical and empirical literature related to inflated forecasts, sophistication and investor reaction to analyst reports. Afterwards, in section three, I describe the data and the variables. In sections four and five I explain the empirical strategy and the results respectively. Finally, in section six I conclude.

2 Empirical Strategy

I test for two possible stories. First, the amount of firm-level information in the hands of investors may help reduce forecast bias, since analysts must issue more precise forecasts to keep investors responsive to their reports, when they have sought for more firm-level information. Then, the higher the acquisition of information at the firm level by investors, the lower the bias in analysts' forecasts. By contrast, if greater amounts of firm-level information sought by investors is interpreted by analysts as investors having more interest on some stocks, then analysts may issue more optimistic (inflated) forecasts in order to generate more commissions for their brokerage houses.

My specification is the following:

$$\begin{aligned} y_{i,t} = \quad & c_i + \textstyle\sum_{j=1}^{4} \rho_j y_{i,t-j} \\ & +\tau_0 \Delta BNH_{i,t-4} \\ & +\tau_1 (D_{i,CG} \times \Delta BNH_{i,t-4}) \\ & +\tau_2 (D_{i,F} \times \Delta BNH_{i,t-4}) \\ & +\boldsymbol{x}_{i,t-5}\boldsymbol{\theta} + \boldsymbol{h}_{i,t-4}\boldsymbol{\omega} + \lambda_t + \varepsilon_{i,t} \end{aligned} \tag{1}$$

where c_i is a stock-level unobserved effect, λ_t are time effects common to all firms and $\varepsilon_{i,t}$ is the error term. My dependent variable is the quarterly forecast bias in terms of optimism in target prices. For each firm i and quarter t, I calculate the forecast bias as

$$y_{i,t} = \frac{TP_{i,t-4} - P_{i,t}}{P_{i,t-4}}$$

where $TP_{i,t-4}$ is the consensus forecast or target price, issued at the end of the quarter $t-4$ for the next 4 quarters on stock i and $P_{i,t}$ is the stock price at the end of the quarter t. Notice that $y_{i,t}$ is a very intuitive measure of optimism since it equals the difference between the projected growth in price $\frac{TP_{i,t-4}}{P_{t-4}}$ and the realized growth $\frac{P_{i,t}}{P_{i,t-4}}$.

As changes in aggregate economic activity may induce changes in forecast optimism, I inspect the behavior of the market forecast bias through time. In figure 1, I show periods of low economic activity (shaded areas), estimated as the quarters during which the Chicago Fed National Activity Index (CFNAI) went from positive to negative[2]. For each quarter, I calculate the market (equally weighted) forecast bias as the cross-sectional average of the forecast bias (black solid line). As we can see from the figure, there are no trends in the aggregate bias or patterns of behavior during quarters of lower economic activity.

2.1 Information Acquisition

My independent variable of interest $\Delta BNH_{i,t-4}$, corresponds to changes in investors' information gathering, and equals the dynamic change in the quarterly average of the "News Heat - Daily Max Readership" index of Bloomberg:

$$\Delta BNH_{i,t-4} = BNH_{i,t-4} - BNH_{i,t-5}$$

where $BNH_{i,t-4}$ is the daily average of the Bloomberg's index during quarter $t-4$. I argue that activities at Bloomberg terminals capture the information gathering by investors. The "News Heat - Daily Max Readership" index (BNH) is constructed by Bloomberg based upon the "number of times each article is read by its users, as well as the number of times users search for news for a specific stock" (Ben-Rephael, Da and Israelsen, 2017) and takes higher values for higher levels of readers activity going from 0 to 4. As documented by Ben-Rephael, Da, and Israelsen (2017), as of August 26, 2016, around 80% of Blommberg Terminal users worked in financial industries (including banking, asset management, and institutional financial services) with 32% of the job titles being portfolio managers or traders, 19% presidents or directors, and only 17% being analysts, including buy-side (who use sell-side analysts valuations and advise portfolio managers privately only) and sell-side ones. I show in tables 1 and 2 that this index is not correlated to analysts forecast revisions. In table 1 I show the tests of correlation between forecast updates ($TP_{i,t} - TP_{i,t-1}$) and ΔBNH, from which we cannot find evidence of correlation. In addition, in table 2 I show the results of a dynamic linear probability model, in which the dependent variable is a dummy $D_{i,t}^{update}$ that takes the value of one whenever the changes in stock price forecasts are less than zero. While the relationship between the probability of a downward forecast revision and $\Delta eps_{i,t-1}$ is negative, from this regression is not possible to reject the null hypothesis of zero correlation between forecast updates and $\Delta BNH_{i,t}$ or $\Delta BNH_{i,t-1}$. Added to the fact that most of Bloomberg Terminal users are in jobs related to the buy side (portfolio managers, traders, etc.), these statistics reinforce the idea that the " News Heat" index captures the information gathering by investors.

I also look at the behavior of the BNH through time in figure 2. Specifically, for each day of the available data, I calculate the cross-sectional average of the BNH or the market (equally weighted) BNH

[2]The CFNAI is a "monthly index designed to gauge overall economic activity and related inflationary pressure" in the U.S. It is constructed to have an average value of zero and a standard deviation of one. A negative index reading corresponds to growth below trend.

index. Also, in figure 2 I show the Cboe Volatility Index (VIX), which is designed to produce a measure of constant, 30-day expected volatility of the U.S. stock market. Since the index for news seeking could be capturing changes in the uncertainty and volatility of the stock market, it is interesting to analyze whether these two variables hold an obvious relation. As we can see from the figure, there are no suggestive patterns between the stock market volatility and the BNH index. Moreover, for each time series, the Augmented-Dickey Fuller test rejects the hypothesis that these are non-stationary at the 1% level of significance, and the Granger test does not reject the hypothesis that there is no Granger-causality with lags of 30 and 120 days.

I include among my dependent variables $\Delta edgar_{i,t-4}$, which is the quarterly change of the number of non-robot downloads of EDGAR filings from SEC.gov for each stock. More precisely, for each stock, I add up the number of downloads registered in one day. With the daily data, I use the average of downloads of the respective quarter for the corresponding stock. I identify the tickers of the stocks from their CIK codes[3] and name this variariable as $edgar_{i,t}$, thus, the quarterly change I use is

$$\Delta edgar_{i,t-4} = edgar_{i,t-4} - edgar_{i,t-5}$$

In the following table I show as before, the results of a dynamic linear probability model, in which the dependent variable is a dummy $D_{i,t}^{update}$ that takes the value of one whenever the changes in stock price forecasts are less than zero. Similar to the regression in table 2, the estimate on $\Delta eps_{i,t-1}$ is statistically negative, and the estimates on $\Delta edgar_{i,t}$ and $\Delta edgar_{i,t-1}$ are not statistically different than zero.

I use $\Delta edgar_{i,t-4}$ in order to capture information acquisition by non-robot non-institutional investors who are likely to be less sophisticated than those captured by Bloomberg searches[4]. The Investor Survey of the FINRA Foundation's 2018 National Financial Capability Study shows that while it is true that free online services provide many investors (44% of respondents) with information potentially useful for their decisions, many of the users of these open access sources of financial information are likely to be more naive or less sophisticated, as 13.4% of them think they do not pay any kind of fee for investing, 9.75% do not know how much they pay and 28,7% do not know whether any of their investment accounts allow them to make purchases on margin. The corresponding numbers, for those who reported using paid subscription services diminish to 10.1%, 5.2% and 12.7%. Da, Engelberg and Gao (2011), argue that institutional investors access information services that are more sophisticated, such as Reuters or Bloomberg terminals, whereas less sophisticated investors are more likely to obtain financial information from free sources, which is in line with Ben-Rephael, Da and Israelsen (2017) who find that their measure of insti-

[3]https://www.sec.gov/include/ticker.txt

[4]I do not claim that all Bloomberg Terminal users are sophisticated according to an absolute measure, or that all EDGAR users are naive according to an absolute rule, but that non-robot EDGAR users tend to be less sophisticated relative to Bloomberg users.

tutional investor attention, based on Bloomberg searches, leads retail attention but not vice versa. The analysis of the traffic statistics at the EDGAR system carried out by Loughran and McDonald (2017), shows that non-robot investors mostly request filings of widely followed companies such as Facebook. This is consistent with research showing that retail trading is correlated to EDGAR activity (Chi and Shanthikumar, 2018; Asthana, Balsam and Sankaraguruswamy, 2004) and suggests that the counting of non-robot downloads of EDGAR filings captures information acquisition by less sophisticated investors, relative to Bloomberg Terminal users.

In figure 3 I present daily percentages of robot downloads of EDGAR filings from 2006-05-11 to 2015-12-31 for the aggregate of stocks. Interestingly, this percentage is high and as we move forward in time, not only the percentage increases but it becomes less volatile[5]. In table 4 I show the summary statistics for this same percentage of robot downloads from which we can see that most of the traffic in the SEC.gov web page corresponds to non-human downloads. This shows the importance of filtering out the data on robot downloads.

2.2 Investors' Recognition of Inflated Forecasts

In my specification, I use a sector dummy for financial firms ($D_{i,F}$), according to the sector indexes developed by the CRSP, which are based on the Industry Classification Benchmark (ICB) of FTSE International to assign companies to sectors. I define the dummy as

$$D_{i,F} = \begin{cases} 1 & if \quad stock\ i \in CRSP\ Financials\ Index \\ 0 & otherwise \end{cases}$$

Financial firms are heavily regulated and supervised (Goldsmith-Pinkham, 2016; Hugonnier and Morellec, 2017; Gunther and Moore, 2003), among others, by the Federal Reserve, Federal Deposit Insurance Corporation (FDIC), the Office of the Comptroller of the Currency (OCC) and the National Credit Union Administration (NCUA), who asses firms' financial health (safety and soundness)[6] and activities related to (anti)money laundering and consumer protection legislation. Since financial institutions are required to issue larger amounts of hard information on their performance, it may be easier for investors to identify those sell-side analysts issuing inflated forecasts for these firms.

Moreover, following the empirical literature showing that institutional investors are more sophisticated (Hilary and Hsu, 2013; Boehmer and Kelley, 2009), as an additional proxy to capture the presence of in-

[5]Including a dummy that takes the value of one for periods after 2012, interacted with human downloads at EDGAR, do not change the results.

[6]The supervisory authorities ask for information related to capital adequacy (available capital v.s risk-weighted credit exposures), asset quality (loan's quality), management's ability to ensure the safe operation, earnings, liquidity and sensitivity to particular risk exposures. See CAMELS ratings.

vestors with a higher ability to identify opportunistic behavior, I use the dummy $D_{i,13F}$ which equals one whenever stock i is included in the list of section 13(f) securities reported in 2017Q4. The list of section 13(f) securities reports those securities in the holdings of investment managers with more than 100 million USD in equity under management[7].

$$D_{i,13F} = \begin{cases} 1 & if \quad stock\ i \in list\ of\ section\ 13(f)\ securities \\ 0 & otherwise \end{cases}$$

2.3 Stocks With Trading Potential

Furthermore, stocks issued by firms in the Consumer Goods sector ($D_{i,CG}$) are more likely to call the attention and be held by a wider public or have a major part in investors' portfolios, and thus are more likely to have a higher trading potential. As I show in table 5 Panel A, market capitalization of firms in the Consumer Goods sector is larger than the market capitalization of firms in Financials and the overall sample. Larger firms are more widely held and stimulate the interest of a large number of investors with more potential transactions business (Atiase, 1985; Bushan, 1989). Also, larger firms receive more coverage in the business media (Fang and Peress, 2009), and media coverage catches investors attention (Solomon, Soltes and Sosyura, 2014; Engelberg and Parsons, 2011; Tetlock, Saar-Tsechansky and Macskassy, 2008) impacting financial decision making, independently of the information conveyed (Kaniel and Parham, 2017)[8]. In addition, analysts can more easily justify inflated forecasts on stock prices issued by firms in the Consumer Goods sector. In table 5 Panels B and C we can observe that firms in the Consumer Goods sector not only have better fundamentals (greater return on assets) but also their stocks register, on average, lower Price-to-Earnings ratios which fit in a story of undervaluation in analysts' reports.

$$D_{i,CG} = \begin{cases} 1 & if \quad stock\ i \in CRSP\ US\ Consumer\ Goods\ Index \\ 0 & otherwise \end{cases}$$

2.4 Controls

In my specification I control for variables of firm performance such as changes in earnings per share and return on assets. More explicitly, in equation 1, the vectors $\boldsymbol{x}_{i,t-5}$ and $\boldsymbol{h}_{i,t-4}$ contain information on earnings, return on assets, stock returns, volatilities and firm size. I denote these variables using lower-case letters as $\Delta eps_{i,t-4}$ and $roa_{i,t-4}$, which correspond to changes in Earnings Per Share scaled by the stock price, i.e. $\frac{EPS_{i,t-4}-EPS_{i,t-5}}{P_{i,t-5}}$ and the firms' Return on Assets ($\frac{Earnings_{i,t-4}}{Assets_{i,t-4}}$) respectively. I additionally

[7] See https://www.sec.gov/fast-answers/answers-form13fhtm.html

[8] Also, firms covered by the media show stronger momentum (Hillert, Jacobs and Müller, 2014) and firm size is related to favorable prospects (Ramnath, Rock and Shane, 2008; Hayes, 1998; McNichols and OBrien, 1997).

include firm size as conventionally measured in the financial literature, i.e. as the log of market capitalization ($size_{i,t-4}$), also past stock returns ($return_{i,t-4}$) and volatility on returns ($sd_{i,t-4}$) estimated as the quarterly standard deviation of daily returns. Notice that $\Delta eps_{i,t-4}$ and $roa_{i,t-4}$ are variables related to fundamental or intrinsic value; $size_{i,t-4}$ is related to trading potential at the firm level; $return_{i,t-4}$ is a variable related to momentum; and $sd_{i,t-4}$ is related to stock risk. Thus, $\boldsymbol{x}_{i,t-5}$ contains either $\Delta eps_{i,t-5}$ or $roa_{i,t-5}$; $\boldsymbol{h}_{i,t-4}$ contains $return_{i,t-4}$, $sd_{i,t-4}$ and $size_{i,t-4}$; and $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ are vectors of parameters on regressors[9].

The term $\Delta BNH_{i,t-4}$ is endogenous, since $y_{i,t}$ is a function of $TP_{i,t-4}$ and higher target prices issued by analysts at $t-4$ may influence investors' attention at $t-4$. Thus, $\Delta BNH_{i,t-4}$ is correlated with the error term ($\varepsilon_{i,t}$). In addition, because of the unobserved fixed effect, I estimate the model in differences, in which, by construction, $\Delta\varepsilon_{i,t}$ and $\Delta y_{i,t-1}$ are correlated. As is standard in dynamic panel data, I deal with the endogeneity problems using instrumental variables and Arellano and Bond's (1991) 2SGMM estimators. In particular, the set of potential valid instruments for $y_{i,t-1}$ is composed of its first lag $y_{i,t-2}$ and higher and the set for $\Delta BNH_{i,t-4}$ are $\Delta BNH_{i,t-5}$ and higher orders, whose validity I test using Arellano and Bond's (1991) m-statistic. Also notice that other regressors are exogenous, since quarterly reported earnings and assets are not determined by stock forecasts but by the accounting revenues, costs and purchases of firms, and thus earnings per share and return on assets are exogenous to forecast bias. Similarly, it is not likely that the amount of information gathered by investors about a firm, during quarter t, affects the financial statements of that firm at the same quarter t.

3 Data

For 2542 firms included in the Center for Research in Security Prices (CRSP) stock index from the first quarter of 2010 to the fourth quarter of 2017, I observe the quarterly series of Earnings Per Share (EPS) and Return On Assets (ROA) of each firm, as well as daily data on its stock price and market capitalization. Also, I identify Financials and Consumer Goods firms, according to the Industry Classification Benchmark (ICB) of FTSE International which counts with 10 sector indexes in total. In the sector of Consumer Goods sector there are firms producing household goods (clothing, electronics, etc.), automobiles, foods and beverages; among Financials, there are banks and firms related to insurance, asset management and real estate investment trusts (REITs). In addition, I observe daily data on the consensus target price, which is the average forecast of the stock price for the next 12 months from the analysts

[9] Analysts' optimism is linked to past signals of fundamental value such as changes in Earnings Per Share (Da, Hong and Lee, 2016; Bradshaw, 2002; Easterwood and Nutt, 1999; Abarbanell and Bernard, 1992; Ali, Klein and Rosenfeld, 1992) and also to market outcomes such as past stock returns (Ali, Klein and Rosenfeld, 1992; Abarbanell, 1991) and volatility on returns (Aslan and Kumar, 2017; Lim, 2001). In addition, the empirical literature shows that analysts' decide to report forecasts selectively, based on whether the firm has a favorable prospect.

who cover that stock, and excludes forecasts older than three months when it is calculated. Forecasts on stock prices express analysts' opinions about the stock market in the most direct and intuitive manner[10], without the statistical problems that raise from earnings management[11] when using earnings forecasts or operating cash flows[12] to capture optimism.

Moreover, I use information on the number of non-robot downloads of EDGAR filings through SEC.gov. The Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system is a public database with free access used at the U.S. Securities and Exchange Commission (SEC) which allow users to search for financial information and operations of public companies, mutual funds and exchange-traded funds among others. The Division of Economic and Risk Analysis (DERA) constructed the EDGAR log file data set containing statistics on user access to the SEC.gov website and is "intended to provide insight into the usage of publicly accessible EDGAR company filings". Among other variables, the EDGAR log file data, which counts with available information from 2003, includes the IP addresses, dates, times, browser type, Central Index Key (CIK) codes and whether the user self-identified as a crawler. [13] I use a filtered version of the data from The Software Repository for Accounting and Finance[14] of the University of Notre Dame, with information availability from 2006 to 2015 which is an extension of the information used in the paper of Loughran and McDonald (2017). This data have been filtered to eliminate damaged files[15], irrelevant entries, or those with missing CIK, accession number, IP, or date. Also, robot downloads (those with more than 49 downloads from a single IP within a single day or self identified as a web crawler), or with a server code larger or equal than 300, or records of traffic on the index page of a set of documents (e.g. index.htm) have been filtered out. The information set I use in this paper have a size of 8.55 GB and includes daily data, from 2010 to 2015, of the number of (exclusively) non-robot downloads for each stock, identified with the CIK number, with their respective dates. The advantage of using the EDGAR log file data set from The Software Repository for Accounting and Finance is that robot downloads of financial information have been filtered out, filtering that is not possible to carry out when analyzing data related to traffic on other free sources of information.

[10] This is supported by Asquith, Mikhail and Au (2005) who find that "the market reaction to price target revisions is stronger than that of an equal percentage change in earnings forecasts."

[11] Earnings management refers to the fact that "[m]anagement can improve or impair the quality of financial statements through the exercise of discretion over accounting numbers" (Beaver, 2002), e.g. estimation of accruals. Therefore, "some 'errors' in the distribution of [analyst] forecast errors may arise only because the forecast was inappropriately benchmarked with reported [manipulated] earnings, when in fact the analyst had targeted a different earnings number" (Abarbanell and Lehavy, 2003).

[12] Givoly, Hayn and Lehavy (2009) find that "cash flow forecasts appear to be a naïve extension of analysts' earnings forecasts."

[13] In the SEC website, www.sec.gov/dera/data/edgar-log-file-data-set.html, there are 2,880 zip folders in the EDGAR log file data set for the period 2010 - 2017, one for each day, and each folder contains a "csv" file with the statistics on SEC.gov website traffic and a "README" file documenting the variables. Each file (day) downloaded directly from the SEC website with the statistics on internet search traffic, can include more than a million entries. The uncompressed data set for the period 2003-2015 consisting of 4,839 daily files, takes 1.73 terabytes.

[14] https://sraf.nd.edu/

[15] All files from 2005-09-24 to 2006-05-10 were labeled by the SEC as "lost or damaged" (Loughran and McDonald, 2017).

I also gather data on institutional investors from the list of section 13(f) securities, issued by the U.S Securities and Exchange Commission[16]. These are quarterly reports available in pdf format, from which I use the report of 2017Q4. I identify the 504 stocks in my sample that are included in this report, which counts with more than 17,000 individual securities.

I do not make use of market measures of informed trading such as the Probability of Informed Trading (PIN) of Easley et al. (1996) or those based on price impact (Sadka, 2006) since nowadays, market measures of informed trading convey information about the high-frequency-trading world, where traders are silicon, not human (O'Hara, 2015). Sell-side analysts, who rely on firm fundamentals or accounting measures to value stocks, do not make forecasts to be used in high frequency trading or even in intra-day trading, which is mostly done by computers. They make forecasts to be used in investment decisions of longer horizons, carried out by humans looking for the true intrinsic value of stocks. As O'Hara (2015) points out, over the time intervals of high frequency trading, the information in market measures is not just asset-related (investment-related[17]) but mostly order-related (speculation-related[18]) and the basic unit of market information is orders.

4 Results

In table 6, I show the results of dynamic models, using $D_{i,CG} \times \Delta BNH_{i,t-4}$ and $D_{i,F} \times \Delta BNH_{i,t-4}$ to capture differentials in analysts incentives. For stocks issued by firms in the Consumer Goods sector ($D_{i,CG} = 1$), the change in analyst optimism ($\partial y_{i,t}$) given a change in information acquisition ($\partial \Delta BNH_{i,t-4}$) is statistically positive ($0.036 = 0.051 - 0.015$), suggesting that, when firm information gathering by the public increases, analysts are more likely to increase their bias for larger firms (higher potential for trading) which at the same time have a lower PE ratio (see table 5 Panel B). Interestingly, the estimates on the parameters on $D_{i,F} \times \Delta BNH_{i,t-4}$, although not statistically significant, are negative. This is consistent with the idea that, when firm information gathering by the public increases, analysts are more likely to reduce their bias for firms that issue larger amounts of hard information on their performance which makes it easier for investors to identify those sell-side analysts incurring in opportunistic behavior.

Tables 7 and 8 show results on the within estimation and the Arellano-Bond estimation, without interactions with dummies, in order to analyze the average effect of investor information acquisition on

[16] https://www.sec.gov/divisions/investment/13flists.htm

[17] As defined by Fisher (1930), an investment is the buying of future income streams. More precisely, "the value of any property, or rights to wealth, is its value as a source of income and is found by discounting that expected income."

[18] Speculation is the trading of an asset for reasons not related to its fundamentals or to its ability to generate future income (see e.g. Zhang and Yao, 2016). As defined in Tirole (1982), people "exhibit speculative behavior if the right to resell [an] asset makes them willing to pay more for it than they would pay if obliged to hold it forever."

the whole sample, and to analyze the importance of including lagged dependent variable among the regressors. While in both models, the panel data with fixed effects and the dynamic panel data, the estimates on $\Delta BNH_{i,t-4}$ are negative, only in the dynamic model are statistically significant. These results point out that, when investors increase their gathering or acquisition of financial information at the firm level, analysts reduce their positive bias. In addition, the Sargan Test as well as the tests for first and second order autocorrelation developed in Arellano and Bond (1991), show the validity of the instruments used in the dynamic model. These results also show the importance of including past values of forecast bias in the specification: as the estimates on $y_{i,t-1}$ are positive, analyst bias has some persistence.

The results on the dynamic panel data models (column (1) in tables 7 and 8) show that return on assets is a better performance variable than changes in earnings per share to explain analysts forecasts, and that performance values before $t-4$ (the date forecasts are issued) are better than values at $t-4$, which is sensible since financial statements take time to be elaborated and are not reported instantly by firms. Furthermore, the results show that stock returns, return volatility and firm sizes are important explanatory variables of analyst forecast bias.

Now, in table 9, I show the results of using the stocks in the 13(f) list ($D_{i,13F}$) in order to capture those stocks in the holdings of institutional investors, and the non-robot downloads of EDGAR filings to capture those stocks grabbing the attention of investors with a lower ability to identify opportunistic behaviors. The positive and significant estimates on $D_{i,CG} \times \Delta edgar_{i,t-4}$ suggest that analysts tend to be more optimistic for stocks with a higher trading potential. In addition, for stocks issued by financial firms in the holdings of institutional investors ($D_{i,F} = D_{i,13F} = 1$) , the change in forecast bias ($\partial y_{i,t}$) given a change in information acquisition ($\partial \Delta BNH_{i,t-4}$) is statistically negative ($-0.002 = -0.067 + 0.066 + 0.077 - 0.078$), consistent with the idea that analysts are more likely to reduce their bias for stocks followed by investors with a higher ability to identify inflated forecasts. These results indicate that information acquisition activate two different analysts' incentives. In addition, the estimates on firm size indicate that larger firms are likely to be more popular and to catch more investors attention, similar to Atiase (1985) and Fang and Peress (2009), thus inducing analysts to issue more optimistic forecasts.

As a value of 0.0003 seems very small, I also run a regression using standardized variables of $\Delta edgar_{i,t-4}$ and $\Delta BNH_{i,t-4}$ which I denote as $Z_{i,t-4}^{\Delta edgar}$ and $Z_{i,t-4}^{\Delta BNH}$. As these results, shown in table 10, indicate, an increase in one standard deviation of $\Delta edgar_{i,t-4}$ (23.2984), is related to a bias increase of 0.5% for stocks in the Consumer Goods sector.

4.1 Only Banks Instead of All Financials

Till now, I have argued that financial firms provide more information than firms in other sectors. While all publicly traded companies disclose detailed information on quarterly financial statements (Form 10-Q), audited annual financial performance (Form 10-K), and unscheduled events at a firm that could be of importance to the shareholders such as the hiring of a new director (Form 8-K), many financial firms such as banks, report additional hard information about measures of available capital v.s risk-weighted credit exposures, asset quality (loan's quality), management's ability to ensure the safe operation and sensitivity to particular risk exposures. Nevertheless, within financials, there is heterogeneity on the type and quantity of information that they release. Hedge funds or close ended funds, to give two examples, provide much less information than commercial banks to the general public.

In my sample of financials there are 832 firms. In this section, I show the results of selecting those firms in the financial sector that are banks, and using only these to capture those stocks for which investors can better establish the quality of the reports. In order to do so, I use a list of 5185 banks that appear at usbanklocations.com which provides information on banks in the United States. I carry out a web scraping of the web page and select the 38 banks of my original sample of financials that are in the list of banks in the US.

I calculate a dummy $D_{i,Banks}$ that takes the value of one for firms that are exclusively banks and estimate a dynamic panel model whose results I show in table 11, indicating that forecast bias is reduced in 36.3% for those stocks issued by banks in the holdings of institutional investors, when information acquisition increases in one standard deviation.

5 Conclusions

In this paper I estimate the effects of investors' information gathering on analyst forecast bias, and find evidence in favor of the hypothesis that investors' information acquisition is positively related to analysts' optimism when the potential for trading is larger, and negatively related to optimism when investors are more likely to identify inflated forecasts. These results suggest that information acquisition is not only correlated to analysts' optimism but also that its effect does not work trivially and solely in one direction but it activates two different incentives in analysts' decisions.

These results highlight the importance of information acquisition: being more informed in the stock market may increase or decrease analyst forecast bias which deteriorates or improves the decision-making process of naive investors. These also help explaining why there are analysts who are systematically biased in the market: even with widespread access to relevant information, not all investors have the same potential to identify inflated forecasts and to penalize analysts who, driven by their trading incentives, incur in an opportunistic behavior. In the words of Lipman (1991): "knowing a fact does not mean that one knows all the logical implications of that fact...the agent can fail to recognize the appropriate action because this requires him to process his information." People, including investors, can be persuaded to pay for even blatantly useless predictions, as we may irrationally act, for instance, to feel in "control" over a random situation, to avoid the regret of not acquiring the forecast if it turns out to be "correct", or to avoid blaming ourselves if the decision outcome goes wrong (Powdthavee and Riyanto, 2015).

These results also highlight the importance of an unanswered question: what drives more and less sophisticated investors to follow a set of stocks and avoid others. The answer is not trivial since it may be related to information access, technology usage in the processing of information and and cognitive efforts. For instance, before deciding how much information is optimal to acquire (i.e. how much effort to make) on a stock, an investor must decide on which stock to gather information, and then the investor may first compute or understand the implications of following a stock instead of another. If finding the best set of stocks to follow is costly, then the investor may first construct a decision procedure in order to decide which stocks to follow, which involves trading off the benefits to improving the choice with the costs of improving the decision-making. But if the construction of this procedure is costly and there are several alternative procedures, then the investor may first also compute an algorithm to decide which procedure to apply, and so on. This is the infinite regress problem of bounded rationality well explained by Lipman (1991) and is related to the idea that "[c]ognitive resources should be allocated just like other scarce resources" (Gabaix and Leibson, 2005).

The difference between more sophisticated and less sophisticated investors may also be related to government regulation. Thus, for further research it would also be worth studying the effects of regulation on analysts behavior. Regulatory agencies have addressed analysts' conflicts of interests by regulating information issuers, with rules such as the Securities Act of 1933 and the Reg FD of 2000 which have the purpose of increasing the amount of information that investors can access directly from firms, or rules that affect analysts compensations such as the Global Research Analyst Settlement and the Sarbanes-Oxley Act of 2002. An interesting question is whether laws that focus on investors i.e. financial information consumers, would improve the quality of the information issued by sell-side analysts. For instance, Choi (2000) proposes limiting the investments of the less informed investors to passive index mutual funds, while requiring from other investors that they make their transactions through brokerage firms with research departments certified as highly respected.

References

Abarbanell, J. S. (1991). Do analysts' earnings forecasts incorporate information in prior stock price changes?. *Journal of Accounting and Economics*, 14(2), 147-165.

Abarbanell, J. S. and Bernard, V. L. (1992). Tests of analysts' overreaction/underreaction to earnings information as an Explanation for Anomalous Stock Price Behavior. *The Journal of Finance*, 47(3), 1181 - 1207.

Abarbanell, J. and Lehavy, R. (2003). Biased Forecasts or Biased Earnings? The Role of Reported Earnings in Explaining Apparent Bias and Over/Underreaction in Analysts' Earnings Forecasts. *Journal of Accounting and Economics*, 36, 105 - 146.

Ali, A., Klein, A. and Rosenfeld J. (1992). Analysts' Use of Information about Permanent and Transitory Earnings Components in Forecasting Annual EPS. *Accounting Review*, 67(1), 183 - 198.

Atiase, R. K. (1985). Predisclosure Information, Firm Capitalization, and Security Price Behavior Around Earnings Announcements. *Journal of Accounting Research*, 21-36.

Asquith, P., M. B. Mikhail and A. S. Au. (2005). Information Content of Equity Analyst Reports. *Journal of Financial Economics*, 75(2), 245 - 82.

Banerjee, S., Breon-Drish, B., and Engelberg, J. (2020). Discussion of "disclosure processing costs, investors' information choice, and equity market outcomes: A review". Journal of Accounting and Economics, 101337.

Beaver, W. H. (2002). Perspectives on Recent Capital Market Research. *The Accounting Review*, 77(2), 453 - 474.

Ben-Rephael, A., Da, Z. and Israelsen, R. (2017). It Depends on Where You Search: Institutional Investor Attention and Underreaction to News. *Review of Financial Studies*, 30(9), 3009-3047.

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

Bhushan, R. (1989). Firm characteristics and analyst following. *Journal of Accounting and Economics*, 11(2-3), 255-274.

Boehmer, E. and Kelley, E. K. (2009). Institutional investors and the informational efficiency of prices. *The Review of Financial Studies*, 22(9), 3563-3594.

Bonini, S., Zanetti, L., Bianchini, R. and Salvi, A. (2010). Target price accuracy in equity research. *Journal of Business Finance & Accounting*, 37(9-10), 1177-1217.

Chang, J. W. and Choi, H. M. (2017). Analyst Optimism and Incentives under Market Uncertainty. *Financial Review*, 52(3), 307-345.

Bradshaw, M. T., Brown, L. D., and Huang, K. (2013). Do sell-side analysts exhibit differential target price forecasting ability?. Review of Accounting Studies, 18(4), 930-955.

Choi, S. (2000). Regulating Investors Not Issuers: A Market-Based Proposal. *California Law Review*, 88(2), 279-334.

Cowen, A., Groysberg, B. and Healy, P. (2006). Which Types of Analyst Firms Are More Optimistic? *Journal of Accounting and Economics*, 41, 119 - 146.

Da, Z., J. Engelberg, and P. Gao. (2011). In search of attention. *Journal of Finance*, 66, 1461-99.

Da, Z., Hong, K. P. and Lee, S. (2016). What drives target price forecasts and their investment value?. *Journal of Business Finance & Accounting*, 43(3-4), 487-510.

Dahiya, S., Iannotta, G. and Navone, M. (2017). Firm opacity lies in the eye of the beholder. *Financial Management*, 46(3), 553-592.

Das, S., Levine, C. B. and Sivaramakrishnan, K. (1998). Earnings predictability and bias in analysts' earnings forecasts. *Accounting Review*, 277-294.

Das, S., Guo, R. J. and Zhang, H. (2006). Analysts' selective coverage and subsequent performance of newly public firms. *The Journal of Finance*, 61(3), 1159-1185.

DeBondt, W. F. M. and Thaler, R. (1990). Do Security Analysts Overreact? *The American Economic Review*, 80(2), 52 - 27.

Ely, J. C., and Välimäki, J. (2003). Bad reputation. The Quarterly Journal of Economics, 118(3), 785-814.

Engelberg, J. E. and Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, 66(1), 67-97

Easley, D., Kiefer, N. M., O'hara, M., and Paperman, J. B. (1996). Liquidity, information, and infrequently traded stocks. The Journal of Finance, 51(4), 1405-1436.

Easterwood, J. C. and Nutt, S. R. (1999). Inefficiency in Analysts' Earnings Forecasts: Systematic Misreaction or SystematicOptimism? *The Journal of Finance*, 54(5), 1777 - 1797.

Fama, E. (1965). Random Walks in Stock Market Prices. Financial Analysts Journal, 51(1), 75-80

Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, 64(5), 2023-2052.

Fang, L., and Yasuda, A. (2009). The effectiveness of reputation as a disciplinary mechanism in sell-side research. The Review of Financial Studies, 22(9), 3735-3777.

Fisher, I. (1930). The theory of interest: as determined by impatience to spend income and opportunity to invest it. Augustusm Kelly Publishers, Clifton.

Fischer, P. E. and Stocken, P. C. (2010). Analyst Information Acquisition and Communication. *The Accounting Review*, 85(6), 1985 - 2009.

Gabaix, X., and Laibson, D. (2005). Bounded rationality and directed cognition. Harvard University.

Gabaix, X., Laibson, D., Moloche, G., and Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. American Economic Review, 96(4), 1043-1068.

Galperti, S. (2019). Persuasion: The art of changing worldviews. American Economic Review, 109(3), 996-1031.

Groysberg, B., Healy, P. M. and Maber, D. A. (2011). What drives sell-side analyst compensation at high-status investment banks?. *Journal of Accounting Research*, 49(4), 969-1000.

Givoly, D., Hayn, C. and Lehavy, R. (2009). The Quality of Analysts' Cash Flow Forecasts. *The Accounting Review*, 84(6), 877 - 1911.

Goldsmith-Pinkham, Paul and Hirtle, Beverly and Lucca, David O., Parsing the Content of Bank Supervision (2016-03-01). FRB of NY Staff Report No. 770. Available at SSRN: https://ssrn.com/abstract=2750315

Gunther, J. W., and Moore, R. R. (2003). Loss underreporting and the auditing role of bank exams. Journal of Financial Intermediation, 12(2), 153-177.

Hayes, R. M. (1998). The impact of trading commission incentives on analysts' stock coverage decisions and earnings forecasts. *Journal of Accounting Research*, 36(2), 299-320.

Hayes, R. M. and Levine, C. B. (2000). An approach to adjusting analysts' consensus forecasts for selection bias. *Contemporary Accounting Research*, 17(1), 61-83.

Hilary, G. and Hsu, C. (2013). Analyst forecast consistency. *The Journal of Finance*, 68(1), 271-297.

Hillert, A., Jacobs, H. and Müller, S. (2014). Media makes momentum. *The Review of Financial Studies*, 27(12), 3467-3501.

Hong, H., Lim, T. and Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance*, 55(1), 265-295.

Hugon, A., and Muslu, V. (2010). Market demand for conservative analysts. Journal of Accounting and Economics, 50(1), 42-57.

Hugonnier, J., and Morellec, E. (2017). Bank capital, liquid reserves, and insolvency risk. Journal of Financial Economics, 125(2), 266-285.

Jackson, A. R. (2005). Trade Generation, Reputation, and Sell-Side Analysts. *The Journal of Finance*, 60(2), 673 - 717.

Jullien, B., and Park, I. U. (2014). New, like new, or very good? Reputation and credibility. The Review of Economic Studies, 81(4), 1543-1574.

Kaniel, R. and Parham, R. (2017). WSJ Category Kings–The impact of media attention on consumer and mutual fund investment decisions. *Journal of Financial Economics*, 123(2), 337-356.

Kartik, N., Ottaviani, M. and Squintani, F. (2007). Credulity, Lies, and Costly Talk. *Journal of Economic Theory*, 134, 93 - 116.

Lim, T. (2001). Rationality and analysts' forecast bias. *The Journal of Finance*, 56(1), 369-385.

Lin, J. T., Bumcrot, C., Mottola, G., and Walsh, G. (2019). Investors in the United States: A report of national financial capability study. FINRA: Investor Education Foundation.

Lipman, B. L. (1991). How to decide how to decide how to...: Modeling limited rationality. Econometrica: Journal of the Econometric Society, 1105-1125.

Loughran, T. and McDonald, B. (2017). The use of EDGAR filings by investors. *Journal of Behavioral Finance*, 18(2), 231-248.

Ljungqvist, A., Marston, F., Starks, L. T., Wei, K. D. and Yan, H. (2007). Conflicts of interest in sell-side research and the moderating role of institutional investors. *Journal of Financial Economics*, 85(2), 420-456.

Malmendier, U. and Shanthikumar, D. (2014). Do security analysts speak in two tongues?. *The Review of Financial Studies*, 27(5), 1287-1322.

Mahmud, S., Scarth, N., Shah, N. and Molloy, P. (2016). The Future of Equity Research. Bloomberg Professional Services, available at:
https://www.bloomberg.com/professional/blog/future-equity-research/

McNichols, M. and O'Brien,P. (1997). Self-Selection and Analyst Coverage. *Journal of Accounting Research*, 35,167 - 199.

Mikhail, M. B., Walther, B. R. and Willis, R. H. (2007). When security analysts talk, who listens?. *The Accounting Review*, 82(5), 1227-1253.

Mikhail, M. B., Walther, B. R. and Willis, R. H. (1999). Does Forecast Accuracy Matter to Security Analysts? *The Accounting Review*, 74(2), 185-200.

Millimet, D. L. and Tchernis. (2013). Estimation of treatment effects without an exclusion restriction: with an application to the analysis of the school breakfast program. *Journal of Applied Econometrics*, 28(6), 982-1017.

Niehaus, G. and Zhang, D. (2010). The Impact of Sell-Side Analyst Research Coverage on an Affiliated Broker's Market Share of Trading Volume. *Journal of Banking & Finance*, 34, 776 - 787.

Ottaviani, M., and Sørensen, P. N. (2006). Reputational cheap talk. The Rand journal of economics, 37(1), 155-175.

Oppel Jr, R. A. (2002). Wall Street analysts faulted on Enron. *The New York Times* (Feb. 28).

Peress, J. (2008). Media coverage and investors' attention to earnings announcements, Working Paper. INSEAD.

Pinto, J. E., Robinson, T. R. and Stowe, J. D. (2019). Equity valuation: A survey of professional practice. Review of Financial Economics, 37(2), 219-233.

Powdthavee, N., and Riyanto, Y. E. (2015). Would you pay for transparently useless advice? A test of boundaries of beliefs in the folly of predictions. Review of Economics and Statistics, 97(2), 257-272.

Ramnath, S., Rock, S. and Shane, P. (2008). The financial analyst forecasting literature: A taxonomy with suggestions for further research. *International Journal of Forecasting*, 24(1), 34-75.

Solomon, D. H., Soltes, E. and Sosyura, D. (2014). Winners in the spotlight: Media coverage of fund holdings as a driver of flows. *Journal of Financial Economics*, 113(1), 53-72.

Soltes, E. (2009). News dissemination and the impact of the business press. The University of Chicago.

Stickel, S. E. (1992). Reputation and performance among security analysts. The Journal of Finance, 47(5), 1811-1836.

Tetlock, P. C., Saar-Tsechansky, M. and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3), 1437-1467.

Tirole, J. (1982). On the possibility of speculation under rational expectations. *Econometrica: Journal of the Econometric Society*, 1163-1181.

Womack, K. L. (1996). Do Brokerage Analysts' Recommendations Have Investment Value? *The Journal of Finance*, 51(1), 137 - 167.

Zhang, Y. J. and Yao, T. (2016). Interpreting the movement of oil prices: driven by fundamentals or bubbles?. *Economic Modelling*, 55, 226-240.

Figures and Tables

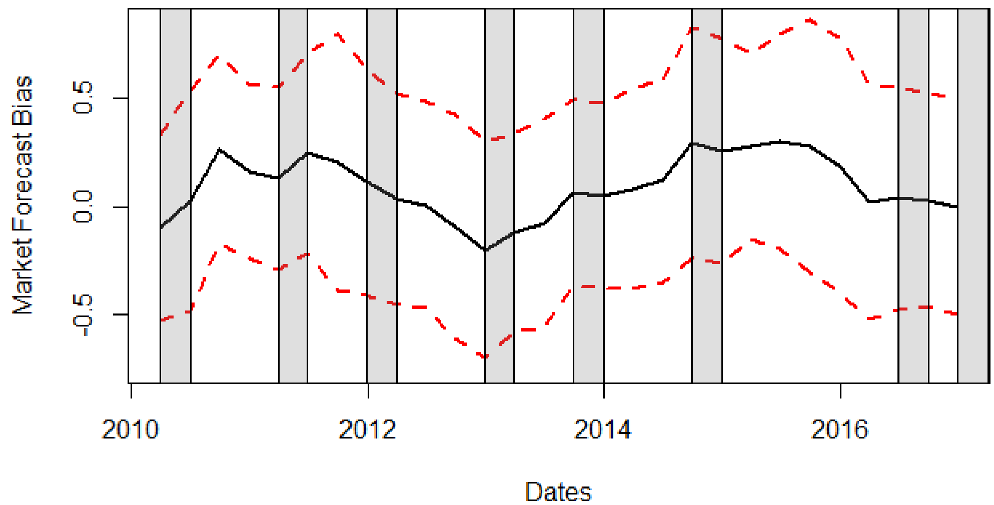

Figure 1: Market forecast bias calculated as the cross-sectional average of the forecast biases. Red dashed lines represent one standard deviation from the mean. Shaded areas are quarters during which the Chicago Fed National Activity Index (CFNAI) went from positive to negative until the first quarter of 2017.

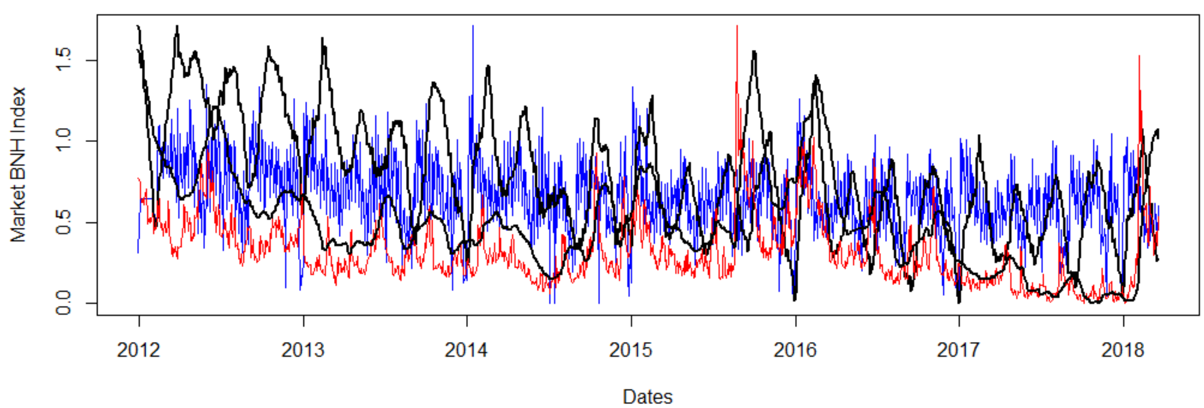

Figure 2: Daily market BNH (blue), Cboe Volatility Index, VIX (red), and 30-day moving averages (black) from 2011-12-30 to 2018-03-20.

Figure 4: Banks in the Sample of Financials

Bank of America	Morgan Stanley	Charles Schwab	State Street	Progressive
Key	Bank of the Ozarks	FNB	South State	Green Dot
BancorpSouth Bank	First Merchants	Trustmark	Renasant	Banner
CenterState Bank	First Ban	Univest	Opus Bank	Cadence Ban
Mercantile Bank	First Financial	Access National	1st Source	Hingham Institution for Savings
Citizens & Northern	Northrim Ban	Paragon Commercial	ACNB	BankFinancial
LCNB	Community Financial	Summit State Bank	Citizens First	Central Federal
Goldman Sachs Group Inc	Elmira Savings Bank	Public Storage		

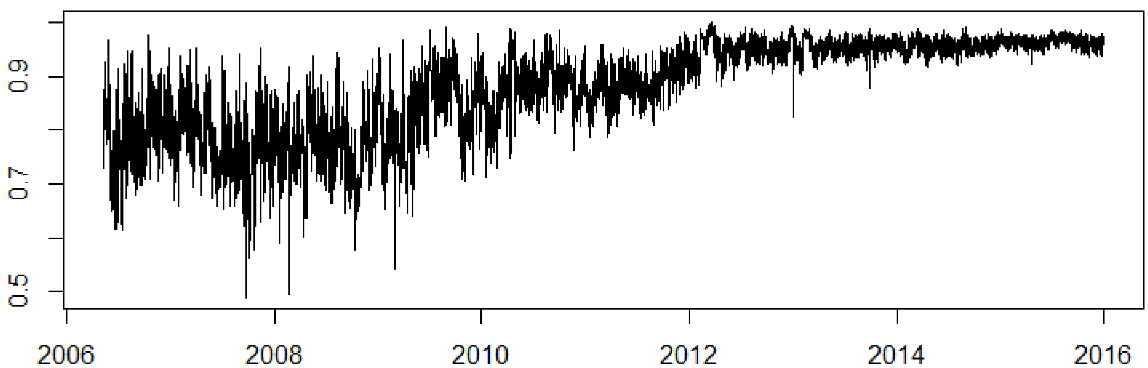

Figure 3: Daily percentages of robot downloads of EDGAR filings from 2006-05-11 to 2015-12-31 for the aggregate of the stocks provided by The Software Repository for Accounting and Finance of the University of Notre Dame. Files from 2005-09-24 to 2006-05-10 were lost or damaged. Robot downloads for specific stocks are not available.

Table 1: Correlations Between Forecast Changes and ΔBNH

Forecast changes are $TP_{i,t} - TP_{i,t-1}$.

	sample estimates	t	p-value	95% percent confidence interval
$\Delta BNH_{i,t}$	0.005910115	1.2643	0.2061	-0.0032 - 0.0151
$\Delta BNH_{i,t-1}$	-0.007173707	-1.5347	0.1249	-0.0163 - 0.001988

Table 2: Linear Probability Model.

Dependent: forecast update $D_{i,t}^{update}$. In the difference equation, the valid instruments for $D_{i,t-1}^{update}$ are $D_{i,t-2}^{update}$ and $D_{i,t-3}^{update}$; for $\Delta BNH_{i,t}$ its lags from $t-2$ to $t-5$. The Arellano-Bond m-statistics for first and second-order autocorrelation indicate a good fit with valid instruments. *Note:* *p<0.1; **p<0.05; ***p<0.01.

	Dependent variable:
	$D_{i,t}^{update} = 1 \quad for \quad TP_{i,t} - TP_{i,t-1} < 0$
$D_{i,t-1}^{update}$	0.128^{***}
	(0.012)
$\Delta eps_{i,t-1}$	-0.106^{**}
	(0.043)
$\Delta BNH_{i,t}$	0.195
	(0.157)
$\Delta BNH_{i,t-1}$	0.039
	(0.037)
A-B m-statistic (1st): -24.6873 (*p-value* $\leq$ 2.22e-16)	
A-B m-statistic (2nd): -0.9461 (*p-value* $=$ 0.34407)	

Table 3: Linear Probability Model.

Dependent: forecast update $D_{i,t}^{update}$. In the difference equation, the valid instruments for $D_{i,t-1}^{update}$ are $D_{i,t-2}^{update}$ and $D_{i,t-3}^{update}$; for $\Delta edgar_{i,t}$ its lags from $t-2$ to $t-5$. The Arellano-Bond m-statistics for first and second-order autocorrelation indicate a good fit with valid instruments. *Note:* *p<0.1; **p<0.05; ***p<0.01.

	Dependent variable:
	$D_{i,t}^{update} = 1 \quad for \quad TP_{i,t} - TP_{i,t-1} < 0$
$D_{i,t-1}^{update}$	0.131^{***}
	(0.009)
$\Delta eps_{i,t-1}$	-0.118^{**}
	(0.057)
$\Delta edgar_{i,t}$	-0.0002
	(0.0004)
$\Delta edgar_{i,t-1}$	-0.0001
	(0.0002)
A-B m-statistic (1st): -32.41349 (*p-value* $\leq$ 2.22e-16)	
A-B m-statistic (2nd): -0.671849 (*p-value* $=$ 0.50168)	

Table 4: Summary Statistics On The Percentage of Robot Downloads of EDGAR Filings. Daily from 2006-05-11 to 2015-12-31.

Calculated as the number of robot downloads divided by the sum of non-robot and robot downloads.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.4899	0.8126	0.9035	0.8763	0.9520	0.9994

Author's calculations with data from The Software Repository for Accounting and Finance of the University of Notre Dame.

Table 5: Summary Statistics for Two Sectors (Not Exhaustive)

Panel A. Market Capitalization.
Calculated as shares outstanding times the last price. Reported in millions.

	Median	Mean	N^o Firms in CRSP
Financials	2,273.26	8,860.91	759
Consumer Goods	2,510.4	13,065.9	322
All	2,266.4	10,502.6	2,542

Panel B. PE ratios. Calculated as $\frac{P_t}{EPS_t}$.

	Median	Mean	N^o Firms in CRSP
Financials	58.14	71.63	759
Consumer Goods	59.22	50.78	322
All	59.10	67.46	2,542

Panel C. Return on assets. Calculated as $\frac{Earnings_t}{Assets_t}$.

	Median	Mean	N^o Firms in CRSP
Financials	0.033432	0.009414	759
Consumer Goods	0.04145	0.03428	322
All	0.03721	0.01321	2,542

Panel D. Forecast bias. Calculated as $\frac{TP_{i,t-4} - P_{i,t}}{P_{i,t-4}}$.

	Median	Mean	N^o Firms in CRSP
Financials	0.03796	0.09962	759
Consumer Goods	0.01851	0.05577	322
All	0.04082	0.09685	2,542

Author's calculations.

Table 6: Results. Arellano-Bond Estimators.
Optimism $y_{i,t}$ is the dependent variable. In the difference equation, the valid instruments for $y_{i,t-1}$ are $y_{i,t-2}$ and $y_{i,t-3}$; for $\Delta BNH_{i,t-4}$ are $\Delta BNH_{i,t-6}$ and $\Delta BNH_{i,t-7}$; for the interactions between $\Delta BNH_{i,t-4}$ and sector dummies, are their lags from $t-6$ to $t-8$. *Note:* *p<0.1; **p<0.05; ***p<0.01.

$y_{i,t-1}$	0.490***
	(0.027)
$roa_{i,t-5}$	0.216***
	(0.057)
$return_{i,t-4}$	-0.133^{***}
	(0.021)
$sd_{i,t-4}$	-0.046^{***}
	(0.011)
$size_{i,t-4}$	0.423***
	(0.024)
$\Delta BNH_{i,t-4}$	-0.015
	(0.016)
$D_{i,F} \times \Delta BNH_{i,t-4}$	-0.026
	(0.023)
$D_{i,CG} \times \Delta BNH_{i,t-4}$	0.051*
	(0.029)
Observations	39690
Sargan Test	10.1158
A-B m-statistic (1st)	-13.8195^{***}
A-B m-statistic (2nd)	0.6856

$y_{i,t-1}$	0.496***
	(0.028)
$\Delta eps_{i,t-5}$	0.033
	(0.035)
$return_{i,t-4}$	-0.132^{***}
	(0.021)
$sd_{i,t-4}$	-0.048^{***}
	(0.011)
$size_{i,t-4}$	0.425***
	(0.024)
$\Delta BNH_{i,t-4}$	-0.016
	(0.016)
$D_{i,F} \times \Delta BNH_{i,t-4}$	-0.027
	(0.023)
$D_{i,CG} \times \Delta BNH_{i,t-4}$	0.051*
	(0.029)
Observations	39690
Sargan Test	9.9947
A-B m-statistic (1st)	-13.8228^{***}
A-B m-statistic (2nd)	0.6914

Table 7: Results. Arellano-Bond and Fixed Effects Estimators.

Optimism $y_{i,t}$ is the dependent variable. In both equations there are fixed and time effects. In the difference equation of the dynamic model, the valid instruments for $y_{i,t-1}$ are $y_{i,t-2}$ and $y_{i,t-3}$; for $\Delta BNH_{i,t-4}$ are $\Delta BNH_{i,t-6}$ and $\Delta BNH_{i,t-7}$. *Note:* *p<0.1; **p<0.05; ***p<0.01.

	A-B (2SGMM)	*Fixed Effects (Within)*
$y_{i,t-1}$	0.492^{***}	
	(0.027)	
$roa_{i,t-4}$	0.035	-0.077^{***}
	(0.080)	(0.021)
$roa_{i,t-5}$	0.218^{***}	0.046^{***}
	(0.056)	(0.017)
$return_{i,t-4}$	-0.130^{***}	-0.515^{***}
	(0.020)	(0.009)
$sd_{i,t-4}$	-0.046^{***}	0.040^{***}
	(0.010)	(0.007)
$size_{i,t-4}$	0.421^{***}	0.308^{***}
	(0.023)	(0.005)
$\Delta BNH_{i,t-4}$	-0.025^{**}	-0.001
	(0.011)	(0.004)
Observations	39690	45766
	Sargan Test: 1.5777	R^2: 0.1215
	A-B m-statistic (1st): -13.8266***	Adjusted R^2: 0.0691
	A-B m-statistic (2nd): 0.6616	F Statistic: 995.343*** (df = 6; 43191)

Table 8: Results. Arellano-Bond and Fixed Effects Estimators.

Optimism $y_{i,t}$ is the dependent variable. In both equations there are fixed and time effects. In the difference equation of the dynamic model, the valid instruments for $y_{i,t-1}$ are $y_{i,t-2}$ and $y_{i,t-3}$; for $\Delta BNH_{i,t-4}$ are $\Delta BNH_{i,t-6}$ and $\Delta BNH_{i,t-7}$. *Note:* *p<0.1; **p<0.05; ***p<0.01.

	A-B (2SGMM)	*Fixed Effects (Within)*
	(1)	(2)
$y_{i,t-1}$	0.498^{***}	
	(0.027)	
$\Delta eps_{i,t-4}$	0.024	-0.141^{***}
	(0.043)	(0.026)
$\Delta eps_{i,t-5}$	0.047	-0.062^{**}
	(0.040)	(0.026)
$return_{i,t-4}$	-0.130^{***}	-0.515^{***}
	(0.021)	(0.009)
$sd_{i,t-4}$	-0.048^{***}	0.040^{***}
	(0.011)	(0.007)
$size_{i,t-4}$	0.424^{***}	0.307^{***}
	(0.024)	(0.005)
$\Delta BNH_{i,t-4}$	-0.026^{**}	-0.001
	(0.011)	(0.004)
Observations	39690	45766
	Sargan Test: 1.4667	R^2: 0.1218
	A-B m-statistic (1st): -13.8712^{***}	Adjusted R^2:0.069448
	A-B m-statistic (2nd): 0.6905	F Statistic: 998.251*** (df = 6; 43191)

Table 9: Results. Arellano-Bond Estimators Using EDGAR Filings Downloads and The 13(f) list.

Optimism $y_{i,t}$ is the dependent variable. In the difference equation, the valid instruments for $y_{i,t-1}$ are $y_{i,t-2}$ and $y_{i,t-3}$; for $\Delta BNH_{i,t-4}$ are $\Delta BNH_{i,t-6}$ and $\Delta BNH_{i,t-7}$; for $\Delta edgar_{i,t-4}$ are $\Delta edgar_{i,t-6}$ and $\Delta edgar_{i,t-7}$. *Note:* *p<0.1; **p<0.05; ***p<0.01.

$y_{i,t-1}$	0.498^{***}
	(0.034)
$roa_{i,t-5}$	0.292^{***}
	(0.054)
$return_{i,t-4}$	-0.133^{***}
	(0.025)
$sd_{i,t-4}$	-0.038^{***}
	(0.013)
$size_{i,t-4}$	0.441^{***}
	(0.029)
$\Delta BNH_{i,t-4}$	-0.067^{***}
	(0.024)
$D_{i,F} \times \Delta BNH_{i,t-4}$	0.066^{***}
	(0.024)
$D_{i,13F} \times \Delta BNH_{i,t-4}$	0.077^{***}
	(0.024)
$D_{i,F} \times D_{i,13F} \times \Delta BNH_{i,t-4}$	-0.078^{***}
	(0.025)
$D_{i,CG} \times \Delta edgar_{i,t-4}$	0.0003^{**}
	(0.0001)
Observations	25391
Sargan Test	2.3036
A-B m-statistic (1st)	-11.1240^{***}
A-B m-statistic (2nd)	0.4296

Table 10: Results Using Standardized Variables

Optimism $y_{i,t}$ is the dependent variable. In the difference equation, the valid instruments for $y_{i,t-1}$ are $y_{i,t-2}$ and $y_{i,t-3}$; for $Z^{\Delta BNH}_{i,t-4}$ are $Z^{\Delta BNH}_{i,t-6}$ and $Z^{\Delta BNH}_{i,t-7}$; for $Z^{\Delta edgar}_{i,t-4}$ are $\Delta edgar_{i,t-6}$ and $\Delta edgar_{i,t-7}$. *Note:* *p<0.1; **p<0.05; ***p<0.01.

$y_{i,t-1}$	0.496^{***}
	(0.034)
$roa_{i,t-5}$	0.289^{***}
	(0.053)
$return_{i,t-4}$	-0.134^{***}
	(0.025)
$sd_{i,t-4}$	-0.039^{***}
	(0.013)
$size_{i,t-4}$	0.441^{***}
	(0.029)
$Z^{\Delta BNH}_{i,t-4}$	-0.029^{***}
	(0.010)
$Z^{\Delta edgar}_{i,t-4}$	0.002
	(0.001)
$D_{i,F} \times Z^{\Delta BNH}_{i,t-4}$	0.029^{***}
	(0.010)
$D_{i,13F} \times Z^{\Delta BNH}_{i,t-4}$	0.034^{***}
	(0.010)
$D_{i,F} \times D_{i,13F} \times Z^{\Delta BNH}_{i,t-4}$	-0.034^{***}
	(0.011)
$D_{i,CG} \times Z^{\Delta edgar}_{i,t-4}$	0.005^{*}
	(0.003)
Obs.	25391
Sargan	3.5234
m-stat. (1st)	-11.0857^{***}
m-stat. (2nd)	0.4205

Table 11: Results Using Banks

Optimism $y_{i,t}$ is the dependent variable. In the difference equation, the valid instruments for $y_{i,t-1}$ are $y_{i,t-2}$ and $y_{i,t-3}$; for $Z^{\Delta BNH}_{i,t-4}$ are $Z^{\Delta BNH}_{i,t-6}$ and $Z^{\Delta BNH}_{i,t-7}$; for $Z^{\Delta edgar}_{i,t-4}$ are $\Delta edgar_{i,t-6}$ and $\Delta edgar_{i,t-7}$. *Note:* *p<0.1; **p<0.05; ***p<0.01.

$y_{i,t-1}$	0.495^{***}
	(0.034)
$roa_{i,t-5}$	0.287^{***}
	(0.053)
$return_{i,t-4}$	-0.135^{***}
	(0.025)
$sd_{i,t-4}$	-0.040^{***}
	(0.013)
$size_{i,t-4}$	0.441^{***}
	(0.029)
$Z^{\Delta BNH}_{i,t-4}$	-0.024^{***}
	(0.008)
$Z^{\Delta edgar}_{i,t-4}$	0.005
	(0.004)
$D_{i,Banks} \times Z^{\Delta BNH}_{i,t-4}$	0.025
	(0.019)
$D_{i,13F} \times Z^{\Delta BNH}_{i,t-4}$	0.026^{***}
	(0.008)
$D_{i,Banks} \times D_{i,13F} \times Z^{\Delta BNH}_{i,t-4}$	-0.365^{**}
	(0.153)
$D_{i,CG} \times Z^{\Delta edgar}_{i,t-4}$	0.005^{*}
	(0.003)
$D_{t,2012} \times Z^{\Delta edgar}_{i,t-4}$	-0.003
	(0.004)
Obs.	25391
Sargan	3.2372
m-stat. (1st)	-11.0775^{***}
m-stat. (2nd)	0.38574