Harnessing the Wisdom of Crowds*

Zhi Da[†] and Xing Huang[‡]

This Draft: December 2016

Abstract

We examine the impact of herding on the accuracy of consensus earnings forecasts from a crowd-based forecast platform (Estimize.com). By tracking user viewing activities, we monitor the amount of information a user views before she makes an earnings forecast. We find that the more public information a user views, the less weight she will put on her private information. While this improves the accuracy of each individual forecast, it reduces the accuracy of the consensus forecast, since useful private information is prevented from entering the consensus. Predictable errors made by "influential users" early on persist in the consensus forecast and result in return predictability at earnings announcements. To address endogeneity concerns related to information acquisition choices, we collaborate with Estimize.com to run experiments where we restrict the information set for randomly selected stocks and users. The experiments confirm that "independent" forecasts lead to a more accurate consensus and convince Estimize.com to switch to a "blind" platform from November 2015. Overall, our findings suggest that the wisdom of crowds can be better harnessed by encouraging independent voices from the participants.

*We thank Renee Adams, Kenneth Ahern, Qi Chen, Erik Eyster (discussant), Cary Frydman, Stefano DellaVigna, Umit Gurun, David Hirshleifer, Harrison Hong, Byoung-Hyoun Hwang, Russell James (discussant), Petri Jylha (discussant), Peter Kelly, Tse-Chun Lin (discussant), Yin Luo, Davud Rostam-Afschar (discussant), Jacob Sagi (discussant), Adam Szeidl, Baolian Wang (discussant), Chishen Wei (discussant), and Holly Yang (discussant), as well as seminar participants at the Arizona State University, Emory University, HEC-Paris, Shanghai Advanced Institute of Finance, Tsinghua University, the University of Arizona, the University of Edinburgh, the University of Georgia, the University of Notre Dame, Villanova University, the 2016 FSU SunTrust Beach Conference, the 2016 SFS Cavalcade, the 2016 ABFER 4th Annual Conference in Singapore, the 2016 CEIBS finance conference, the 2016 WFA, the 2016 CICF, the 2016 Early Career Behavioral Economics conference, the NBER PRIT workshop, the 2016 Helsinki finance submit, the 2016 European Finance Association meeting, the 27th Annual Conference on Financial Economics and Accounting, and the 2016 CalTech behavioral finance junior faculty conference for their helpful comments and suggestions. We thank Leigh Drogen and Josh Dulberger from Estimize for their generous support.

[†]University of Notre Dame, Mendoza College of Business, Notre Dame, IN, 46556, USA. Email: zda@nd.edu

[‡]Michigan State University, Eli Broad College of Business, East Lansing, MI, 48824, USA. Email: huangx@bus.msu.edu

"The more influence we exert on each other, the more likely it is that we will believe the same things and make the same mistakes. That means it's possible that we could become individually smarter but collectively dumber." James Surowiecki, The Wisdom of Crowds.

1 Introduction

Many important decisions in life are made in a group setting.[1] Consequently, a crucial topic in social science is how to best elicit and aggregate information from individuals. A great deal of evidence suggests that, under certain conditions, the simple average of a large group's answers to a question involving quantity estimation is generally as good as, and often better than, the answer provided by any individual in that group.[2] This phenomenon is commonly referred to as the "wisdom of crowds." As long as individual estimates are unbiased and independent, the law of large numbers implies that the crowd average will be very accurate.

In most social and economic settings, however, individual estimates are unlikely to be independent since they are often issued sequentially, and individuals learn from observing other people's actions and beliefs. Will such a herding behavior result in a better or worse crowd average? We study this question in a specific setting where individuals make corporate earnings forecasts. Forecasting earnings is a suitable setting for studying the wisdom of crowds since it involves quantity estimations rather than predictions.[3] In addition, both earnings forecasts and realizations are easily observable and the forecast error can be clearly defined. Finally, accurate earnings forecasts are of crucial importance to investors, firms and the functioning of the financial market in general. Not surprisingly, a wide range of market participants provide earnings forecasts. They include equity analysts from both the sell-side and buy-side, and, more recently, independent analysts.

[1] Examples include the war on Iraq, jury verdicts, the setting of the interest rate by the Federal Open Market Committee (FOMC), and the appointment of a CEO by a firm's board of directors, just to name a few.

[2] See Sunstein (2005) for a general survey of this topic in the context of group judgments.

[3] Almost all forecasts we examined were issued after the earnings have already taken place. This is very different from other applications where the crowds are asked to predict future events (for example, the future stock price or the election outcome) and their answers could change their behaviors and therefore the future outcome.

There are several reasons why herding in a sequential forecast setting may generate a more accurate crowd average (the consensus earnings forecast). The sequential setting may encourage discussion and additional information production that improves the precision of subsequent analysts' private signals. In addition, if an analyst with a very precise private signal always issues her forecast early, then the subsequent herding by other analysts could improve the accuracy of the consensus forecast.

On the other hand, a long strand of literature on sell-side Wall Street analyst forecasts (from Institutional Brokers' Estimates System or IBES) find that herding results in correlated rather than independent forecasts. In the extreme case of an information cascade, subsequent forecasters' private information is completely discarded so the crowd consensus is no more accurate than the forecast that starts the cascade.[4] In general, as detailed in the Appendix, when a later analyst herds with the early analysts, the final consensus forecast will overweight the private signals of the early analysts. While still unbiased, the consensus forecast under a sequential setting maybe less efficient than the consensus under a simultaneous setting, which is simply the average of all analysts' private signals.

To clarify, throughout the paper, we follow Hirshleifer and Teoh (2003) and use the term "herding" in a broad sense to refer to situations where individuals place positive weights on other people's estimates when forming their own estimates. This behavior can be completely rational when the individual computes the weights using the Bayes' rule (see Banerjee (1992) and Bikhchandani et al. (1992) among others). The individual can also herd by underweighing her private signal (see Scharfstein and Stein (1990), Banerjee (1992), Bikhchandani et al. (1992), Trueman (1994), Hong et al. (2000), Welch (2000), and Clement and Tse (2005) among others). Alternatively, she can anti-herd by overweighing her private signals (see Ehrbeck and Waldmann (1996), Ottaviani and Sorensen (2006) and Bernhardt et al. (2006) among others). Finally, she can apply other naive weights (see

[4] As forcefully put by Bikhchandani et al. (1992), "the social cost of cascades is that the benefit of diverse information sources is lost. Thus a cascade regime may be inferior to a regime in which the actions of the first n individuals are observed only after stage n + 1." Information cascade rarely happens with earnings forecasts though, as earnings are drawn from a continuous distribution.

Eyster and Rabin (2010) and Eyster and Rabin (2014) among others). While we provide initial evidence in the Appendix regarding the specific herding behavior among individuals in our sample, differentiating among these various forms of herding behaviors is not the main goal of our paper.

Instead, we focus on empirically examining the net impact of herding on the accuracy of the consensus forecast. Quantifying such a net impact is usually challenging, as researchers are generally unable to observe the counter-factual, in which analysts make their forecasts independently. In this paper, we tackle this challenge by taking advantage of a unique dataset on user activities and by running randomized experiments on a crowd-based earnings forecast platform (Estimize.com).

Estimize.com, founded in 2011, is an open web-based platform where users can make earnings forecasts. The resulting consensus forecasts are available on both the company's website and Bloomberg terminals. A diverse group of users make forecasts. Among the 2516 users studied in our sample, one third are financial analysts coming from buy-side, sell-side, or independent research firms. The remaining users are working professionals from different industries and students. Both academic and practitioner studies have documented the value of the Estimize consensus forecasts. For example, Jame et al. (2016) document that the Estimize consensus is a better proxy for market expectations than the Wall Street consensus computed using IBES data. In addition, they find the consensus computed using both Estimize and Wall Street forecasts to be even more accurate. A contemporaneous study by Adebambo and Bliss (2015) also finds that Estimize consensuses are more accurate than traditional Wall Street earnings consensus 58%-64% of the time.

Users on Estimize.com make their forecasts sequentially as well. Indeed, before making her own forecast, a user can view a default webpage (the "release page") that contains information on past earnings, the current Estimize consensus forecast, and forecasts from other Estimize users. As a result, herding behavior is expected among Estimize users. The unique feature of our data is that we can observe the users' entire web activities on Estimize.com, which allows us to differentiate forecasts made with and without viewing the release page. Forecasts made without a release page view are less likely to be influenced by existing Estimize forecasts.

For our sample period from March 2012 to March 2015, we examine 2147 quarterly firm earnings (releases) with at least 10 forecasts prior to the announcement. These releases come from 730 distinct firms in various sectors. We find the release viewing activity to have significant impact on the forecasts. First, release viewing is associated with less weighting on private information, consistent with herding behavior. Second, while release viewing improves the accuracy of an individual forecast, it makes the consensus less accurate. This is because some useful private information may be discarded when a user place weights on the prior forecasts. In particular, errors in earlier forecasts are more likely to persist and appear in the final consensus forecast, making it less efficient.

However, our empirical tests are affected by the endogeneity associated with viewing choice. One could argue that a user may choose to view the release page only when he has little private information.[5] In order to address this endogeneity concern, we collaborate with Estimize.com to run experiments during the second and third quarter of 2015, in which we restrict the public information set on randomly selected stocks and users. Specifically, for randomly selected stock, we randomly select users, hide information on their release page, and ask them to make a blind forecast. Each blind forecast is then matched to a default forecast issued at about the same time by a user who could view the entire release page. Compared to the blind forecast, the default forecast uses significantly less private information and is more accurate on average. Nevertheless, the consensus computed from blind forecasts is significantly more accurate than that computed using matched default forecasts.

Immediately after the blind forecast is made, the release view is restored and the user can choose to update the forecast. During the pilot experiment in the second quarter of 2015, users are often genuinely surprised when they are selected to participate in the blind experiment and, as a result, they often revise their forecasts immediately when the release view is restored. We then compare the accuracy of two consensus forecasts: (1) the blind consensus computed using all blind forecasts; and (2) the revised consensus, computed using all revised forecasts made when the release view is

[5]see Trueman (1994) and Graham (1999) among others.

re-enabled. Out of the 13 stocks randomly selected in the pilot experiment, the blind consensus significantly outperforms the revised consensus 10 times, while the revised consensus outperforms the blind consensus only 2 times. They tie in the remaining case. In other words, our findings suggest that the wisdom of crowds can be better harnessed by encouraging independent voices from participants. These findings are so compelling that in November 2015, Estimize.com decided to switch to a blind platform, where users make forecasts without seeing the current consensus.[6] Our analysis of forecasts issued during the year after November 2015 confirms that Estimize consensus indeed becomes more accurate following the switch.

Having confirmed that herding reduces the accuracy of the consensus, we then examine when the herding behavior is predictably stronger. We find that the herding behavior becomes more severe when the public information set contains the estimates of influential users.

We first define a novel measure of user influence on the Estimize network using users' viewing activities and the PageRank algorithm invented by Google to rank webpages. We keep track of how many times Estimize users view each other on the website. Intuitively, users with high PageRank measures are viewed more by other users (either directly or indirectly), so their forecasts have more influence on subsequent forecasts. We also attempt to identify influential users using three other criteria: the total number of their forecasts, the total number of times when their forecasts are viewed by other uses, and whether their forecasts lead subsequent forecasts. Interestingly, we do not find influential users to provide more accurate forecasts. Hence, herding with influential users do not automatically improve the accuracy of subsequent forecasts.

We find very similar results regardless of which definition of influential user is used. First, users are more likely to underweight their private information when the releases they view contain the forecasts of influential users. Second, when influential users issue forecasts that are higher (lower) than the current consensus, the final consensus will move up (down), consistent with the notion that subsequent users are herding with the influential users.

[6] see http://blog.estimize.com/post/133094378977/why-the-estimize-platform-is-blind.

Third, this herding behavior predicts the accuracy of the final consensus forecasts. When the contemporaneous stock return is negative and influential users issue forecasts that are lower than the current consensus early on, the final consensus is more accurate, consistent with the notion that influential users facilitate the incorporation of negative information. On the other hand, when the contemporaneous stock return is negative and influential users nevertheless issue forecasts that are higher than the current consensus, the final consensus becomes less accurate. In this case, influential users' forecasts likely reflect positive sentiments that are propagated among subsequent users and drag the consensus in the wrong direction. In other words, because of herding, predictable errors made early by influential users are not offset by subsequent forecasts, and persist in the consensus forecast.

Finally, building on early research that finds the Estimize consensus to better proxy market expectations for corporate earnings, we then examine the important question of whether predictable error in consensus Estimize earnings forecasts induced by herding affects the stock prices. Put differently, is the financial market smart enough to correct for these errors? When we examine the stock returns during earnings announcements, we find no return predictability in general, except when influential users made early forecasts that are too optimistic. This optimism bias persists in the release and the market does not completely undo it, so we observe a significant negative return during the subsequent earnings announcement window.

Our paper contributes directly to the literature on herding. Much progress has been made in understanding various mechanisms underlying herding behavior.[7] Herding behavior has been documented in various lab settings (see Anderson and Holt (1997) and Kubler and Weizsacker (2004) among others). Empirically, herding behavior has been found to be pervasive.[8] By measuring and randomizing an individual's information set in a large crowd-based earnings forecast platform, we are better able to isolate the net impact of herding behavior on outcomes with direct real-life

[7] Hirshleifer and Teoh (2003) review several possible sources including (1) payoff externalities, (2) sanctions upon deviants, (3) preference interactions, (4) direct communication, and (5) observational influence.

[8] Hirshleifer and Teoh (2003) review evidence for herding behavior in securities trading, security analysis, firm investment, financing, and reporting decisions.

implications.

Our findings also have broader implications regarding group judgment.[9] Our results confirm that independent views are crucial for reaching an efficient outcome in such a setting. We focus on the simple arithmetic average in computing the group consensus estimate in this paper and find that this simple consensus can be significantly improved in a blind forecasting environment where herding is difficult. While the simple arithmetic average seems most natural approach in an egalitarian society, there are of course other ways of averaging individual estimates to reach a more accurate consensus. For example, one could use the median to alleviate the impact of outliers. One could also overweight the estimates from users with better track records or more experience. In our view, the simple average of independent estimates still offers a robust and efficient group consensus, especially when the exact nature of herding behavior and the precision of individual signals are unknown.

The blind forecasting environment may also improve group judgment by eliminating other inefficient strategic behavior. For example, it has been shown that with a convex payoff, individuals may even anti-herd. In other words, they may exaggerate their private signals in order to stand out from the crowd (see Ehrbeck and Waldmann (1996), Ottaviani and Sorensen (2006) and Bernhardt et al. (2006) among others). Since Estimize's scoring method also penalizes a bold forecast exponentially if it turns out to be deviating in the wrong direction, anti-herding behavior is not prevalent on Estimize.com. Nevertheless, the blind forecasting environment also prevents anti-herding, by hiding information about the crowd.

[9] Recent field studies by Barber et al. (2003), Charness et al. (2011), Adams and Ferreira (2010), and Charness and Sutter (2012), among others, all demonstrate that group decisions are moderate and reason-based.

2 Data and Sample Description

2.1 Brief introduction to Estimize

Estimize.com is an open web-based platform that facilitates the aggregation of financial estimates from a diverse community of individuals. Since the firm was founded in 2011, increasing numbers of contributors have joined the platform and the coverage of firms has also significantly expanded. As of December 2015, more than 10,000 regular users contribute on the platform, resulting in coverage of more than 1500 stocks each quarter.

Unlikely the IBES, Estimize solicits contribution from a wide range of individuals, including both professionals, such as sell-side, buy-side, or independent analysts, and non-professionals, such as students, private investors, and industry experts. Because of the contributions of these individuals, who have diverse background and viewpoints, Estimize better represents the market's true expectation than the Wall Street consensus and can serve as a supplementary source of information to IBES, as documented by Jame et al. (2016) and Adebambo and Bliss (2015).

There are several reasons why Estimize consensus can be more accurate than the Wall Street consensus in IBES. First, Wall Street forecasts often contain predictable biases, driven by investment banking relations (Lin and McNichols (1998); Michaely and Womack (1999)) or career concerns (Hong and Kubik (2003)) among other things. In contrast, Estimize users do not suffer from these biases. Nevertheless, the fact that market reacts more to the earnings surprise computed using Estimize consensus than that computed using Wall Street consensus suggests that the Estimize consensus goes beyond simply de-biasing the Wall Street consensus. Indeed, the fact that Estimize consensus is complementary to the Wall Street consensus suggests that it contains incremental value. For example, Estimize users maybe the firm's employees or consumers and they could have private information about the earnings of the firm that is not available to the Wall Street analysts.

Estimize users have several incentives to provide information and contribute to Estimize. First, many users (e.g., independent analysts and students) can create a verifiable track record of their accuracy and ability to predict the fundamental metrics.

Second, Estimize assigns points to its contributors' forecast. Points winners get recognized on their website, featured in podcasts, and awarded with a prize package, such as an Apple watch. Recently, Estimize organized All-America student analyst competitions; winners received awards at Institutional Investor Magazine's annual awards dinner. The point system rewards forecasts that are more accurate than the Wall Street consensus and punishes forecasts less accurate than the Wall Street consensus. The system also incentivizes aggressive estimation by awarding points on an exponential scale in order to elicit more private information.[10] Since the point system also penalizes bold forecasts exponentially if they turn out to be incorrect, deviating from the crowd systematically without private information should not be the optimal strategy in most cases.

Consistent with the incentive structure underlying the point system, our empirical analysis confirms that Estimize contributors on average overweight their private signals relative to a Bayesian benchmark, even though they still put positive weights on the current consensus. Importantly, since the exact formula for computing points is never made public, it is not easy to strategically game the scoring system or to compute the exact optimal forecasting strategy.

Third, the goodwill factor may motivate some users to participate in the platform, especially during the site's early days, just for the sake of its success — the more contributions, the more valuable the dataset is to everyone.

2.2 Dataset

We collect three sets of data from Estimize. The first dataset contains information on the forecasts created by users in the Estimize community. The sample period is March 2012 through March 2015. The forecasted earnings per share (EPS) value and the time at which the forecast was created are

[10] Specifically, according to Estimize.com, "the number of points received is determined by the distance of your estimate to the reported results of the company and the distribution of all other estimates for that earnings release. The system incentivizes aggressive estimation by awarding points on an exponential scale. While being a little more accurate than Wall Street may score you a few points, an aggressive estimate well outside the mean will have both a higher risk, and a far higher reward."

both provided.

The second dataset contains background information on users in the Estimize community. Based on a brief personal profile voluntarily provided by the users themselves, Estimize classifies users in several career-biographical categories, such as buy-side and sell-side professionals, industry experts, students, etc.[11]

The third dataset records users' entire activities on Estimize.com, including the pages that they view and the actions that they take (e.g., creating forecasts); the data includes the time stamps of all activities. The detailed web activities are made available through Mixpanel, an advanced analytics platform for mobile and web. We mainly focus on how many times a user views the release page of a specific firm that she covers. Figure 2 gives an example of a typical release page. The figure presents a screenshot of the release page corresponding to the 2015 Q2 earnings of Facebook, Inc. (FB). The release page contains two charts as shown in the figure. The left chart presents the actual EPS of the past 8 quarters, the range and consensus of Wall Street forecasts, and the range and consensus of Estimize forecasts for the current quarter and past 8 quarters. The right chart contains information on all individual forecasts created for the current quarter. The count of views on the release page could proxy for whether the user's information set contains information from other users on the platform. Users can also click any individual listed in the right chart to access an estimate page that presents all forecasts created by that individual. We also exploit the number of views of a user's estimates page to construct a measure of influence.

2.3 Sample construction

We match the information on forecasts and web activities to form a comprehensive dataset with forecast-level observations, covering the period from March 2012 through March 2015.[12] For each

[11]The profile information, though voluntarily provided, should be reasonably reliable. When a new analyst contributes to Estimize, they are put through a manual review process which considers the depth of their biographical information and the reliability of their first 5 estimates.

[12]These two datasets exploit different identifiers for users. We first use the time stamp of forecast creation activities in both datasets to construct a table to link the two identifiers.

forecast created by a user, we track whether she views the related release page for longer than 5 seconds.[13]

The initial sample includes 91,411 forecasts with 14,209 releases. We drop forecasts if the users cannot be successfully linked with an identifier in the activity dataset. We also exclude forecasts that are flagged manually or algorithmically unreliable by Estimize.[14] Finally, in order to ensure a reasonably sized crowd for each release, we only consider in our analysis releases with at least 10 forecasts. The consensus forecast is always computed using the most recent forecast from a user.

2.4 Descriptive statistics

Our final sample consists of 38,115 forecasts with 2,147 releases. Figure 1 presents the coverage of our sample over time and demonstrates a trend of increasing numbers of contributors and expanding coverage of firms, which is similar to the trend in the full sample. In Table 1, we provide descriptive statistics for our final sample. Panel A presents descriptive statistics for the release level. On average, about 16 users contribute 20 forecasts to a single release. The average release has around 19 views of the release page, though the median count of release views is smaller (12 views). It is worth noting that we observe a wide range in the number of release views. Users may be very independent when making forecasts for some releases (e.g., only 1 release view), but check the release pages frequently for other releases (e.g., more than 114 release views). The wide range of release viewing activities provides considerable variation across releases.

The average consensus on Estimize is slightly pessimistic, with an average consensus error of -0.02. The average absolute value of the consensus error is 0.08, which is one cent more accurate than the average Wall Street consensus. When we examine a typical release in our sample, on average, 35.6% of all forecasts in that release are issued after viewing the release page. Across

[13]We set a cutoff for the length of time spent on one page, because we want to exclude cases where a user just passes a page to access the next one. We obtain similar results when we use other cutoff points and when we do not use a cutoff.

[14]According to Estimize.com, forecasts will be flagged and not included in the Estimize consensus if they have been manually or algorithmically unreliable, or if they have not been revised within the past 60 days and fall well outside of the current consensus. About 2.5% of all estimates made on the platform are determined unreliable.

different releases, there is a lot of variations in the average release viewing activity, which allows us to examine the impact of release page viewing on forecast accuracy.

We also obtain financial characteristics data from Compustat. Panel B presents the size and book-to-market (B/M) statistics for release-level observations.[15] To compare the financial characteristics with NYSE stocks, we also report statistics on the size and B/M NYSE quintile group for firms in our sample.[16] The average firm size is $24.5 billion, while the median firm size is considerably smaller, about $7.6 billion. The average B/M ratio is 0.40 and the median B/M is 0.31. Our sample covers significantly larger firms compared than NYSE stocks, with a strong growth-tilt. These firms cover a wide range of sectors (Panel D), such as information technology, consumer discretionary, industrials, health care, and consumer staples. Information technology and consumer discretionary are the two major sectors and account for more than 50% of our sample.

The forecasts covered in our sample are contributed by 2,516 users (Panel C). The average user covers 10 firms and contributes 17 forecasts, and the distribution is strongly skewed to the right — there are many users contributing a moderate number of forecasts, while a few users frequently contribute on the platform. Estimize obtains contribution from individuals with remarkably diverse backgrounds. As Panel E shows, 33.31% of the contributors studied in our sample are financial professionals, including sell-side (6.42%), buy-side (11.41%) and independent analysts (15.48%). The rest of the contributors are not professional analysts. Two major groups of non-professionals are information technology (21.08%) and students (20.02%).

3 Herding and Forecast Accuracy

In this section, we examine the impact of herding on the behavior and accuracy of individual and consensus earnings forecasts.

[15]Only 1,953 out of 2,147 release-level observations are successfully matched with data from Compustat.

[16]The size group and B/M group are obtained by matching each release with one of 25 size and B/M portfolios at the end of June based on the market capitalization at the end of June and B/M, the book equity of the last fiscal year end in the prior calendar year divided by the market value of equity at the end of December of the prior year.

In our empirical analysis, we focus on the raw and unscaled earnings forecasts. Cheong and Thomas (2011) document that analysts' earnings forecast errors and dispersions do not actually vary with scale in the cross-section. We find similar scale-invariance with the Estimize earnings forecasts. Robustness checks confirm that the results are qualitatively similar when we scale the earnings forecasts by the (split-adjusted) stock price at the end of previous quarter. To save space, these results are not reported.

In addition, we control for various fixed effects in our regressions. In our forecast-level regressions, release fixed effects subsume the need to control for stock characteristics and seasonality. Professional and individual fixed effects subsume the need to control for user characteristics. In our release-level regressions, we incorporate sector and quarter fixed effects.

Standard errors in our main regressions are double-clustered by sector and quarter. They are clustered by stock in regressions using our experimental data. In both cases, however, herding-induced correlations among different forecasts in the same release are accounted for, since a release is nested in either the sector, or the stock cluster. We confirm that the clustered standard errors are more conservative than those estimated from a random effect model, which represents an alternative way to deal with forecast error autocorrelation.

3.1 Release view and weighing of information

We first examine how release viewing affects the relative weighting between private and public information when a user makes a forecast. We follow the empirical framework of Chen and Jiang (2006).

Let z denote the true earnings and c denote the current market consensus about z. The user has a private signal y about z. Assume

$$
\begin{aligned}
c &= z + \varepsilon_c, \\
y &= z + \varepsilon_y,
\end{aligned}
$$

where ε_c and ε_y are independent and normally distributed with zero means and precisions of p_c and p_y, respectively. The user's best forecast according to Bayes' rule is:

$$
\begin{aligned}
E[z|y,c] &= hy + (1-h)c, \\
h &= \frac{p_y}{p_c + p_y}.
\end{aligned}
$$

The user may not apply the most efficient weight h in reality. Instead, the actual forecast f could be $f = ky + (1-k)c$. Chen and Jiang (2006) show that when regressing forecast error ($FE = f - z$) on a forecast's deviation from the consensus ($Dev = f - c$), the slope coefficient converges to $1 - \frac{h}{k}$. In other words, in the regression of:

$$FE = \alpha + \beta_0 \cdot Dev + \varepsilon,$$

β_0 measures the actual weighting of private and public information relative to the optimal weighting. For example, a positive β_0 implies overweighting of private information ($k > h$).

Table 2 reports the regression results at the forecast level. In addition to Dev, we also include a release view dummy and its interaction with Dev as independent variables in the regressions. We find a significantly positive β_0, suggesting that Estimize users are, on average, overweighting their private signals.[17] Most importantly, we find a significant negative coefficient on the interaction

[17] Without the interaction terms, β_0 is 0.18, similar to that reported by Chen and Jiang (2006) who examine sell-side

term between *Dev* and the release view dummy. For example, the coefficients reported in Column (1) suggests that release viewing reduces the excessive weight on private information by 0.274 (from 0.424 to 0.150). In other words, viewing of the current consensus, not surprisingly, is associated with placing more weight on the consensus and less weight on the private signal, consistent with herding behavior. To rule out the possibility that our results are driven by a particular user type or by a particular release, we include firm-quarter (or release), profession, and individual fixed effects in Columns (2) to (5). The results are very similar.

3.2 Release view and forecast accuracy

How does the viewing of public information affect the forecast accuracy? We first examine this question at the individual forecast level by regressing the absolute forecast error on the release view dummy. We include release fixed effects. Effectively, we are comparing forecasts for the same release, with and without release views. In addition, we include a Close-to-Announcement dummy variable that is equal to 1 if the forecast was issued during the last three days before the earnings announcement. This dummy variable controls for the fact that forecasts closer to the announcement should be more accurate.

In Panel A of Table 3, we find a significant negative coefficient in Column (1). Release viewing reduces the forecast error by more than 0.73 cents. In Column (2), we further include user profession fixed effects and again the result does not change much. In Column (3), we replace user profession fixed effects with individual fixed effects. We still find that viewing the release page reduces individual forecast error. Overall, it is clear that viewing public information, including the current Estimize consensus, improves the accuracy of each individual forecast.

But what about the accuracy of the consensus forecast, or the wisdom of the crowd? We examine this question at the release level in Panel B. For each release, we measure the frequency of release viewing as the logarithm of one plus the ratio of the number of forecasts made by users who viewed

equity analysts.

the release for longer than 5 seconds to the number of total forecasts (LnNumView). In other words, if most users viewed the release page before making their forecasts for that release, LnNumView for that release will be higher. Interestingly, when we regress absolute consensus forecast error on LnNumView, we find a significant positive coefficient on LnNumView, suggesting that the viewing of public information actually makes the consensus forecast less accurate. Compared to a release where all forecasts are made without viewing the release page (LnNumView = 0), a release where all forecasts are made after viewing the release page (LnNumView = ln(2) = 0.69) is 3.82 (= 0.0551×0.69 using the coefficient reported in Column (3)) cents less accurate. This represents a significant decrease in accuracy as the median forecast error is only 3 cents in our sample (see Table 1, Panel A).

Another way of seeing this result is through a simple horse race, which we conduct in Panel C. In each release, we separate all forecasts into two groups. The view group contains all forecasts made after viewing the release page. The no-view group contains the remaining forecasts, made without first viewing the release page. We then compute two consensus forecasts using the forecasts from the two groups and compare which consensus is more accurate. Out of the 2,127 releases we studied, the no-view consensus wins 59.24% of the time, which is significantly more than 50%. Again, the viewing of public information makes the consensus forecast less accurate.

How can viewing a release page improve the accuracy of individual forecasts but at the same time make the consensus less accurate? The intuition is simple: when a user herds with the prior forecasts, he is less likely to make an extreme forecast error, the individual forecast error is reduced on average. At the same time, herding prevents useful private information from entering the final consensus, making the consensus less accurate. In the most extreme case, if all subsequent users completely herd on the first user, then the private information of the subsequent users is entirely discarded, so the crowd consensus is no more accurate than the first forecast in that sequence. In particular, errors in earlier forecasts are more likely to persist and show up in the final consensus forecast.

Table 4 examines one such persistent error at the release level. The dependent variable is a dummy variable that is equal to one if earlier and close-to-announcement estimates are biased in the same direction. The close-to-announcement window is defined as extending from five days before the announcement date through the announcement date ([-5,0]). The early window is defined as any of the days prior to day -5. The consensus within the window is upwardly (downwardly) biased if the difference between the consensus and the actual EPS is above H-th percentile (below L-th percentile). The main independent variable is again LnNumView, but measured using only forecasts in the close-to-announcement window. The control variables include the same measure of forecast uncertainty, and sector and quarter fixed effects. The results confirm a strong link between the persistence of error and release views. When more forecasts are made after viewing the release page, the initial error is more likely to persist and show up in the final consensus, making it a less efficient forecast.

3.3 Blind experiments

Our empirical tests so far are affected by the endogeneity associated with the choice of whether or not to view. One could argue that a user may choose to view the release page only when he has little private information. In order to address the endogeneity concerning the information acquisition choice, we collaborate with Estimize.com to run randomized experiments during the second and third quarters of 2015. Note that the experiments take place after the sample period of our main analysis.

The stocks in our experiments are randomly selected to come from a wide range of industries. We then randomly pick a set of users to participate in the experiment. When a user is selected, she will be asked to make an earnings forecast while the release page is disabled. Figure 3 gives an example of a disabled (blind) release page. The figure presents a screenshot of the blind release page for Lululemon Athletica Inc. (LULU) for the fourth quarter of 2015. The left chart plots the historical data of the actual EPS, the range and consensus of Wall Street forecasts, and the range and consensus of Estimize forecasts. Note that no information on the consensus is provided for the fourth quarter. The right chart shows that all Estimize estimates of LULU's EPS, including the current Estimize consensus, are hidden. Importantly, the current Wall Street consensus is still available on the blind release page. Even if a selected user has no private information about the earnings, she can always use the current Wall Street consensus as her default forecast and revise it later when the release page is restored. In addition, making the current Wall Street consensus always available also limits the downside associated with the blind forecasting environment by eliminating completely uninformed forecasts.

The resulting forecast is labeled the blind forecast (f_b). Each blind estimate is matched with the closest estimate in the sequence made by a different user who could view the release page. The matched estimate is labeled the default forecast. The pair is removed if the time difference between the blind estimate and the default estimate exceeds 24 hours.[18] The final sample contains releases with at least 15 matched pairs. There are 103 releases in the final sample, 13 from the first round pilot experiment and the remaining 90 from the second round experiment.

In Panel A of Table 5, we first confirm that the users who issued blind forecasts (the blind users) are similar to those who issued default forecasts (the default users). Blind and default users are equally likely to be professional analysts (27.2% vs. 27.1%) and highly viewed by other users. Prior to the experiment, they have covered similar numbers of stocks (55 vs. 53) and have participated in similar numbers of releases (102 vs. 94) on average. Their past forecast accuracy is also similar.

While blind users are randomly selected and behaved similarly to the default users in the past, they may choose to issue blind forecasts during the experiment only if they have precise private signals. This endogenous choice may drive the accuracy of the blind consensus. Empirically, we observe the opposite pattern. The ratio between the number of forecasts and the number of release page views is actually higher among the blind users than the ratio among the default users. In other

[18] We also examined a more conservative matching procedure where the default estimate is always chosen during the 24 hours *after* the blind estimate. To the extent that a more recent estimate is usually more accurate, this match procedure biases against the blind estimate. We find similar results under this alternative approach.

words, conditioning on arriving at the release page, blind users are more likely to issue forecasts than default users. More directly, Panel B of Table 5 repeats the analysis from Panel A of Table 3 on the experimental sample. We again find the blind forecast to be individually less accurate than the matching default forecast. Blind users do not seem to issue forecasts only when they have precise private signals.

We also examine the dispersion in the blind forecasts versus that in the default forecasts. We find that, on average, the standard deviation of the default forecasts is 11.09% lower than that of the blind forecasts (t-value = 1.95). In other words, the ability to view the current Estimize consensus and other individual users' forecasts reduces the forecast dispersion. This finding is more consistent with herding behavior than with an anti-herding strategy where a winner-takes-all payoff scheme induces the user to strategically deviate from the crowd. Nevertheless, for a small set of releases where the number of contributing users is less than 15, the default forecasts can have larger dispersions, suggesting that strategic behavior can be relevant when the number of players is small. Eliminating such strategic behavior offers another channel for blind forecasts to improve the accuracy of the consensus forecast.

We then compare the blind forecasts to their matching default forecasts in terms of information weighting. As in Panel A of Table 2, we regress forecast errors (FE) on Dev and its interaction with the default forecast dummy ($Default$) with release fixed effects. The results are reported in Table 5. The regression in Column (1) does not include profession fixed effects. First, the large, positive, and significant coefficient on Dev (0.670) confirms that blind forecasts are made almost exclusively with private information. The coefficient is higher than the corresponding number (0.489) in Panel A of Table 2, suggesting that the blind forecasts in the experiment rely more on private information than forecasts from full sample made without viewing the release page. Second, the significant negative coefficient of -0.113 on $Dev \times Default$ indicates that the ability to view public information results in less overweighting of private information, and more reliance on public information. Importantly, since both experiment participants and stocks are randomly selected, the difference between the blind forecast and the default forecast cannot be driven by the endogenous decision to view the release page. The results with profession fixed effects in Column (2) are very similar.

Since users can always see the Wall Street consensus, we consider a placebo test where we replace the Estimize consensus (c) with the Wall Street consensus in the above regression (c_{ws}). We find a small and insignificant coefficient of less than 0.1 on $Dev \times Default$. This is not surprising as both blind and default forecasts are made with c_{ws} included in the information set.

The more interesting question is whether blind forecasts result in a more accurate consensus than the default forecasts. We examine this question with a simple horse race. For each release, we compute two consensus forecasts. The blind consensus is computed as the average of all blind forecasts and the default consensus is computed as the average of all default forecasts. By construction, the two consensuses are computed using the same number of forecasts. Out of the 103 releases examined, we find the blind consensus to be more accurate 62 times. The associated one-tail p-value is smaller than 0.0001 in rejecting the hypothesis that blind and default consensus are equally accurate.

To gauge the statistical significance in each pairwise comparison, we also conduct jackknife resampling. Take the Q1 earnings for Facebook (F) as an example, 24 distinct users are randomly selected to participate in the experiment. They issue 24 blind forecasts, which are in turn matched to 24 default forecasts. In each resample, we remove one user and compute the blind and default consensus using the remaining 23 forecasts, and check which is more accurate. We find the blind consensus to beat the revised consensus in all 24 resamples, resulting in a p-value of 0. Out of the 103 releases examined, blind consensus significantly beats the default consensus 58 times, with a p-value of less than 10%, while default consensus wins significantly only 38 times.

The experimental evidence so far confirms that limiting information access may encourage the user to express more independent opinions and therefore improve the accuracy of the consensus forecast. So far, we have compared the forecasts from two different groups of users (blind and default). Next, we compare two different forecasts from the same user from the pilot experiment.

In our experiment, immediately after the blind forecast (f_b) is issued, the release page is re-enabled so the user can view the Estimize forecasts and consensus and choose to revise her forecast. The new forecast is labeled the revised forecast (f_r). Users can of course choose not to change their forecasts, in which case, the revised forecast is the same as the blind forecast. In the pilot experiment, many users are genuinely surprised when they are first selected to participate in the blind experiment. Consequently, many of them choose to immediately revise their forecasts after issuing the blind forecast, when the release page is enabled.[19] In this case, we could interpret both f_b and f_r as the combination of the same private signal y and the Estimize consensus: $f_b = w_b y + (1 - w_b)c$ and $f_r = w_r y + (1 - w_r)c$. It can then be shown that:

$$f_b - f_r = \frac{w_b - w_r}{w_b}(f_b - c).$$

In other words, if we regress $f_b - f_r$ on $f_b - c$ and obtain a positive slope coefficient, it means that the blind forecast places more weight on the private signal than the revised forecast does ($w_b > w_r$). When we run the regression in Panel A of Table 6, we indeed find a positive and significant coefficient of about 0.534 (Column 2). We again consider a placebo test where we replace the Estimize consensus (c) with the Wall Street consensus in the above regression (c_{ws}). We find a small and insignificant coefficient on $f_b - c_{ws}$.

In Panel B, we compare the accuracy of two consensus forecasts: (1) the blind consensus computed using all blind forecasts; and (2) the revised consensus computed using all revised forecasts. Out of the 13 randomly selected releases in the pilot experiment, the blind consensus significantly outperforms the revised consensus 10 times and the revised consensus wins only 2 times. They tie in the remaining 1 case. The statistical inference is again conducted using jackknife resampling.

To summarize, our experiment results suggest that wisdom of crowds can be better harnessed

[19] As the users became more familiar with the experiment, they realized that they do not have to immediately revise their blind forecasts. Indeed, in the second experiment, f_r lags f_b by 2 days on average. Since new information may have arrived during that gap, f_r became less comparable to f_b in the second experiment.

by encouraging independent voices from the participants. Motivated by our findings, Estimize.com decided to switch to the blind forecast platform; since November 2015, forecasts from other users are always blocked initially. As stated in their announcement of the switch, "(consensus) only gets better with a greater number of independent opinions, ... , while your estimate for a given stock may be less accurate than the average of your peers, it is an important part of building a better consensus."

A natural question is whether the blind platform indeed improves the accuracy of Estimize consensus. Our analysis of the Estimize forecasts during the four quarters immediately after the switch shows that the answer is yes. Specifically, we compare the Estimize consensuses during the before-experiment-period (before Mar 2015) to those during the after-experiment-period (Nov 2015 - October 2016). We limit the comparison to the same set of stocks and the same quarter of the year. For example, we compare the consensus of Facebook's 2016 Q3 earnings to other Facebook Q3 consensuses before the experiment. There are 1,641 stocks that are covered by Estimize in both the before-experiment-period and the after-experiment-period. To control for market wide variations during our sample period, we always compare the Estimize consensus against the corresponding Wall Street consensus.

Column 1 in Table 7 shows that the accuracy of Estimize consensus improves after the blind platform was adopted. In the before-experiment-period, the Estimize consensus beat the Wall Street consensus 56.67% of the time. In the after-experiment-period, the Estimize consensus is more accurate 64.11% of the time. The increase in the winning percentage of 7.44% is highly significant (t-value of 6.28). We also confirm that the improvement is not driven by increasing participation of Estimize users during the after-experiment-period so the Estimize consensus is computed using more individual forecasts. Columns 2 to 4 confirm the improvement even when we restrict the comparisons to consensuses computed using similar numbers of individual forecasts. Not surprisingly, when the number of individual forecasts exceeds 30, the improvement in Estimize consensus becomes smaller.

4 Influential Users and Return Predictability

So far, we have confirmed that herding, while it improves the accuracy of individual forecasts, reduces the accuracy of the consensus forecast; interestingly, withholding certain information from individual users actually improves the average forecast. Two questions follow. First, when is herding behavior more severe and resulting in predictable errors in the consensus forecast? Second, does predictable forecasting error lead to return predictability? Put differently, is the market smart enough to correct these errors?

4.1 The role of "influential" users on herding

The evidence using the unique release view information suggests that the influence we exert on each other can make the crowd's average estimate less accurate. Of course, not all users are created equal. Some users can potentially exert stronger influence on the others. We would therefore expect herding behavior to be more severe when more influential users are present in the crowd.

To measure the influence of Estimize users, we make use of the user viewing activity data and the PageRank algorithm developed by Google for ranking webpages. Figure 4 contains an illustrative example. Different Estimize users are represented by different circles and they are linked by arrows that capture viewing activities. For example, when user D views user A, it results in an arrow going from user D to user A. Such view activities provide direct observations of person-to-person influence, just like insider trading legal documents provide direct observations of person-to-person communication as in Ahern (2016). An influential user (represented by a bigger circle) either receives more incoming arrows (as in the case of user B) or receives an arrow from another influential user (as in the case of user C). The user influence is measured by the PageRank measure, which is reported inside the circle. Intuitively, users with high PageRank measures are viewed more by other users (either directly or indirectly), so their forecasts are more influential; they have more impact on subsequent forecasts.

In computing the PageRank measure for each Estimize user, we also account for the number of times user A viewed user B. When we regress PageRank scores on user characteristics across different Estimize users, we find a user to be more influential if he makes more forecasts and if his forecasts are more often viewed by other users. Interestingly, the user's average forecast accuracy and average forecast bias do not affect the PageRank measure. As two simple alternative measures of user influence, we therefore also consider the total number of forecasts made by the user and the total number of times the user has been viewed by others.

Our fourth measure of user influence attempts to capture the extent to which a user's forecasts lead subsequent forecasts. For each estimate in a release, we measure the ratio of (the distance of subsequent estimates from the current estimate) over (the distance of subsequent estimates from the consensus of previous estimates). A smaller ratio means subsequent estimates are dragged towards the current estimate. In other words, a smaller ratio indicates a leading estimate. Then we count the number of times each user's estimates are identified as leading (among the smallest three ratios for that release), and normalize the count by the total number of submitted estimates by the user as the probability of being a leader.

The measures for users who submit fewer than 20 forecasts are assigned to the lowest value. The users who rank above the 80th percentile on the measure are identified as influential users. None of the four criteria gives a complete description of an influential user; however, when we find consistent results across all four criteria, we are confident that we are indeed capturing many influential users.

Table 8 examines how influential users affect subsequent users in their relative weighting of public and private information at forecast level. The key independent variable of interest is a triple interaction term among Dev, the release view dummy, and an influence dummy variable that equals 1 when a large number of influential users have made forecasts. As in Table 2, we find a negative coefficient on the interaction term between Dev and the release view dummy, so that viewing of the release page is associated with more weight on the consensus and less weight on the private signal. More importantly, the coefficient on the triple interaction term is negative and significant. In other words, when the current release page contains the forecasts of influential users, viewing this page is associated with placing even more weight on the consensus and less weight on the private signal. Simply put, users herd more with influential users.

4.2 Predictable forecast error

Since influential users issue more accurate earnings estimates on average, herding with influential users may not always result in a less accurate consensus forecast. Given that influential users' forecasts strongly swing subsequent forecasts, we conjecture that if influential users' early forecasts are inaccurate, this is likely to drag the consensus in the wrong direction. To identify such a forecasting error ex-ante, we use the contemporaneous stock return as a proxy for the information content and compare the direction of influential users' forecast revisions against the sign of the contemporaneous stock return. If their signs are consistent, then the revision is likely to be informative; if they are opposite to each other, then the revision is likely to contain an error.

To directly examine how influential users' forecasts affect subsequent forecasts, we again separate the forecasting period into earlier and close-to-announcement periods, as in Table 4. In Panel A of Table 9, we then regress the consensus forecast revisions in the later period (the close-to-announcement period) on influential users' forecast revisions in the earlier period. Across all four definitions of influential users, we find very consistent results: if influential users issued forecasts that are higher (lower) than the current consensus in the earlier period, the consensus will move up (down) in the later period, confirming that influential users' forecasts strongly swing subsequent forecasts.

In Panel B, we find that when the contemporaneous stock return is negative and influential users issue forecasts that are lower than the current consensus, the final consensus becomes more accurate, consistent with the notion that influential users facilitate the incorporation of negative information. On the other hand, when the contemporaneous stock return is negative and influential users nevertheless issue forecasts that are higher than the current consensus, the final consensus becomes less accurate. In this case, influential users' forecasts likely reflect positive sentiments that are propagated to subsequent users and drag the consensus in the wrong direction.

4.3 Return predictability

Both Jame et al. (2016) and Adebambo and Bliss (2015) provide evidence suggesting that the Estimize consensus is a better proxy for the market's expectations of future firm earnings. Our analysis of influential users so far shows that such a consensus may contain predictable errors. Does the market fully understand these predictable errors? If it does, then it should not be surprised by the actual earnings.

In Table 10, we examine the earnings-announcement window returns and find strong return predictability in only one scenario: when the initial positive sentiment expressed by influential users persists in the final Estimize consensus, the market is negatively surprised at the earnings announcement, as evident in a significantly lower cumulative abnormal return. Specifically, a one standard deviation increase (=0.66) in ln(1+Num of UD), which captures an influential user's positive sentiment, lowers the earnings announcement window returns by 46 basis points ($= 0.66 \times -0.007$). This return predictability is not too surprising. Using IBES forecasts, So (2013) also documents that stock prices do not fully reflect the predictable components of analyst errors. Our analysis provides at least one channel where these predictable errors may arise.

5 Conclusion

The wisdom of crowds hinges on each crowd member making independent estimates. In many real life applications, however, estimates and opinions from a crowd are elicited in a sequential basis. Since participants learn from observing each other, they also exert influence on each other, and herding behavior arises, resulting in the loss of useful private information.

By taking advantage of a unique dataset from a web-based corporate earnings forecast platform, we can better isolate the impact of user influence on the ultimate accuracy of the consensus forecasts.

We find that the more public information a user views, the more she will underweight her private information. While this improves the accuracy of the individual's forecast, it reduces the accuracy of the consensus forecast, since useful private information is prevented from entering the consensus, consistent with herding. We also find that herding behavior becomes more severe if the public information set contains the estimates of more influential users. Interestingly, the resulting errors in the earnings consensus, while predictable, do affect stock returns. In other words, our preliminary evidence suggests that the market does not always undo errors-in-expectations that arise from herding behavior.

A randomized experiment offers clean evidence that the wisdom of crowds can be better harnessed by encouraging independent voices from the participants. Ironically, by limiting the crowd's access to information, we can actually improve the accuracy of their consensus forecast. We are confident that by adopting such a blind forecast platform, Estimize.com will continue to generate more accurate corporate earnings forecasts, which are crucial for the efficiency and function of the financial market.

References

Renee Adams and Daniel Ferreira. Moderation in groups: Evidence from betting on ice break-ups in alaska. *Review of Economic Studies*, 77:882–913, 2010.

Biljana N. Adebambo and Barbara Bliss. The value of crowdsourcing: Evidence from earnings forecasts. *Working Paper*, 2015.

Kenneth Ahern. Information networks: Evidence from illegal insider trading tips. *Journal of Financial Economics*, forthcoming, 2016.

Lisa R. Anderson and Charles A. Holt. Information cascades in the laboratory. *American Economic Review*, 87(5):847–862, 1997.

Abhijit V. Banerjee. A simple model of herd behavior. *Quarter Journal of Economics*, 107(3): 797–818, 1992.

Brad Barber, Chip Heath, and Terrance Odean. Good reasons sell: Reason-based choice among group and individual investors in the stock market. *Mangement Science*, 49(12):1636–1652, 2003.

Dan Bernhardt, Murillo Campello, and Edward Kutsoati. Who herds? *Journal of Financial Economics*, 80:657–675, 2006.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, 1992.

Gary Charness and Matthias Sutter. Groups make better self-interested decisions. *Journal of Economic Perspectives*, 26(2):157–176, 2012.

Gary Charness, Edi Karni, and Dan Levin. Individual and group decision making under risk: An experimental study of bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and Uncertainty*, 35(2):129–148, 2011.

Qi Chen and Wei Jiang. Analysts' weighting of private and public information. *Review of Financial Studies*, 19(1):319–355, 2006.

Foong Soon Cheong and Jacob Thomas. Why do eps forecast error and dispersion not vary with scale? implications for analyst and managerial behavior. *Journal of Accounting Research*, 49(2): 359–401, 2011.

Michael B. Clement and Senyo Y. Tse. Financial analyst characteristics and herding behavior in forecasting. *Journal of Finance*, 60(1):307–341, 2005.

Tilman Ehrbeck and Robert Waldmann. Why are professional forecasters biased? agency versus behavioral explanations. *Quarterly Journal of Economics*, 111(1):21–40, 1996.

Erik Eyster and Matthew Rabin. Naive herding in rich-information settings. *American Economic Journal: Microeconomics*, 2:221–243, 2010.

Erik Eyster and Matthew Rabin. Extensive imitation is irrational and harmful. *Quarterly Journal of Economics*, 129(4):1861–1898, 2014.

John Graham. Herding among investment newsletters: Theory and evidence. *Journal of Finance*, 54:237–268, 1999.

David Hirshleifer and Siew Hong Teoh. Herd behaviour and cascading in capital markets: a review and synthesis. *European Financial Management*, 9(1):25–66, 2003.

Harrison Hong and Jeffrey D. Kubik. Analyzing the analysts: Career concerns and biased earnings forecasts. *Journal of Finance*, 58(1):313–351, 2003.

Harrison Hong, Jeffrey D. Kubik, and Amit Solomon. Securityanalysts'careerconcernsand herding of earnings forecasts. *RAND Journal of Economics*, 31:121–144, 2000.

Russell Jame, Rick Johnston, Stanimir Markov, and Michael C. Wolfe. The value of crowdsourced earnings forecasts. *Journal of Accounting Research*, forthcoming, 2016.

Dorothea Kubler and Georg Weizsacker. Limited depth of reasoning and failure of cascade formation in the laboratory. *The Review of Economic Studies*, 71(2):425–441, 2004.

Hsiou-Wei Lin and Maureen F. McNichols. Underwriting relationships, analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics*, 25(1):101–127, 1998.

Roni Michaely and Kent L. Womack. Conflict of interest and the credibility of underwriter analyst recommendations. *Review of Financial Studies*, 12(4):653–686, 1999.

Marco Ottaviani and Peter Norman Sorensen. The strategy of professional forecasting. *Journal of Financial Economics*, 81(2):441–466, 2006.

David S. Scharfstein and Jeremy C. Stein. Herd behavior and investment. *The American Economic Review*, 80(3):465–479, 1990.

Eric C. So. A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts? *Journal of Financial Economics*, 108(3):615–640, 2013.

Cass R. Sunstein. Group judgments: Statistical means, deliberation, and information markets. *New York University Law Review*, 80:962–1049, 2005.

Brett Trueman. Analyst forecasts and herding behavior. *Review of Financial Studies*, 7:97–124, 1994.

Ivo Welch. Herding among security analysts. *Journal of Financial Economics*, 58:369–396, 2000.

Appendix: Herding and the Wisdom of Crowds

Herding can make individual forecasts more accurate, yet at the same time make the crowd consensus less accurate; the reason why are intuitive. Consider a crowd of N individuals. Each has an independent private signal about the earnings: $y_1, y_2, ..., y_N$. For simplicity of illustration, we assume these signals are drawn from identical distributions with zero mean and variance σ^2. The true earnings is zero. The consensus after the nth individual submits her forecast (f_n) is:

$$c_n = \frac{1}{n}\sum_{i=1}^{n} f_i, \text{ for all n = 1,..., N}$$

.

When forecasts are made simultaneously, the crowd consensus will simply be the average of these private signals ($\bar{y}$), as each individual will just issue her private signal ($f_i = y_i$). By the law of large numbers, when N is large, the crowd consensus will be very close to the true mean (zero), and is likely to be more accurate than any individual forecast (y_n in this case). This phenomenon is known as the "wisdom of crowds."

When the forecasts are made sequentially, however, each individual may herd by placing a positive weight on the current consensus, with the exception of the first individual, whose forecast will still be her private signal ($f_1 = y_1$). In other words, individual n's forecast (f_n) is a weighted average between her private signal (y_n) and the consensus of all prior forecasts (c_{n-1}):

$$f_n = (1 - w_n)y_n + w_n c_{n-1}, 1 > w_n > 0.$$

For example, an individual may display the naive herding behavior described in Eyster and Rabin (2010) the or extensive imitation discussed in Eyster and Rabin (2014) when the current individual ignores prior individual forecasts and/or fails to account for the fact that prior forecasts are issued sequentially. In this case, the individual will simply equally weight all previous forecasts with her own private signal, or $w_n = (n-1)/n$. Alternatively, the individual may always place a constant weight on the current consensus, or $w_n = w$.

The benchmark case is when all individuals compute their forecasts rationally, in a Bayesian manner. The individual forecast f_n, in this case, will equal the arithmetic average of all private signals up to n, and it will converge to the truth as n increases. In addition, as n increases, the individual forecast (f_n) will be a more efficient estimator than the consensus forecast (c_n), since the consensus will overweight earlier private signals.

Figure A1, Panel A differentiates the above herding behaviors by their corresponding impulse response functions: how the first forecast impacts subsequent forecasts. In the rational case, the impact of the first forecast decays at the rate of $1/n$. In contrast, with naive herding, the impact of the first forecast is always 0.5 and does not decay over time. Finally, in the case of a constant weight on the prior consensus, when the constant weight is high ($w_n = 0.9$), the first forecast can have large impact on subsequent forecasts. While the impact decays over time, the decaying speed can be slow.

When we estimate the impulse response function empirically, we find evidence supporting the idea that Estimize users herd by placing a constant weight on the current consensus. Figure A1, Panel B sheds some light on the herding behavior of these Estimize users. We estimate the impulse response function: how the first forecast in a release impacts subsequent forecasts in the same release. Specifically, we regress the forecast errors associated with the second, third, ..., twentieth forecasts on the forecast errors associated with the first forecast, across different releases in our sample. The regression coefficients therefore measure the impact of the first forecast on the second, third, ..., twentieth forecasts and are plotted in Panel B together with their confidence bands. Comparing our results to the different theoretical cases shown in Panel A, we find the herding behavior of Estimize users most closely resembles the case where individuals put a large and constant weight on the current consensus when forming their own forecasts. Of course, the plot in Panel B needs to be interpreted with caution given the implicit assumption that Estimize users have uniform private signals. Nevertheless, a comparison of Panels A and B suggests that

Estimize users are unlikely to behave completely rationally in forming their forecasts. In addition, we also find the last forecast in a release to be significantly less accurate than the final consensus forecast in that release on Estimize, again rejecting the notion that users are rationally forming their forecasts.

For emphasis, the main objective of our paper is not to formally distinguish the different herding behaviors of Estimize users. Instead, as long as individuals place a positive weight on prior consensus ($w_n > 0$), early forecasts will exert some influence on later forecasts. The goal of our paper is to isolate and quantify this phenomenon's net impact on the accuracy of both individual forecasts (f_n) and the consensus forecast (c_n).

The following lemma shows that the final consensus c_N can be expressed as a weighted-average of private signals, with more weight on earlier signals in the sequence.

Lemma 1 *The final consensus of all forecasts can be described as a weighted sum of all private signals:*

$$c_N = \sum_{i=1}^{N} l^N(i) y_i,$$

where weights ($l^N(i)$) sum up to one, $\sum_{i=1}^{N} l^N(i) = 1$.

Proof According to the definition, the general form of the consensus of the first n forecasts could be written as:

$$\begin{aligned} c_n &= \frac{1}{n}(f_n + (n-1)c_{n-1}), \text{ for } n \geq 2 \\ &= \frac{1-w_n}{n} y_n + \frac{n-1+w_n}{n} c_{n-1} \end{aligned}$$

We will prove by induction.

Base case: when n=2,

$$\begin{aligned} c_2 &= \frac{1}{2}(f_2 + f_1) \\ &= \frac{1-w_2}{2}y_2 + \frac{1+w_2}{2}y_1 \end{aligned}$$

So c_2 is a weighted average of the first two private signals, and the weights sum up to 1.

Induction step: Assume c_{n-1} is a weighted average of the first $(n-1)$ private signals, $c_{n-1} = \sum_i^{n-1} l^{n-1}(i)y_i$, where $\sum_i^{n-1} l^{n-1}(i) = 1$.

Hence,

$$\begin{aligned} c_n &= \frac{1}{n}(f_n + (n-1)c_{n-1}) \\ &= \frac{1-w_n}{n}y_n + \frac{n-1+w_n}{n}\sum_{i=1}^{n-1} l^{n-1}(i)y_i \\ &= \sum_i^n l^n(i)y_n \end{aligned}$$

Therefore, c_n could be written as a weighted sum of all private signals with the weights satisfy:

$$\begin{aligned} l^n(n) &= \frac{1-w_n}{n} \\ l^n(i) &= \frac{n-1+w_n}{n}l^{n-1}(i) \text{ for } i < n \end{aligned}$$

We can easily prove that the weights also sum up to 1:

$$\begin{aligned} \sum_{i=1}^n l^n(n) &= \frac{1-w_n}{n} + \sum_{i=1}^{n-1}\frac{n-1+w_n}{n}l^{n-1}(i) \\ &= \frac{1-w_n}{n} + \frac{n-1+w_n}{n}\sum_{i=1}^{n-1} l^{n-1}(i) \\ &= 1 \end{aligned}$$

Lemma 1 shows that since forecasts are made sequentially, private signals will not be equally weighted in the final consensus. In fact, as long as w_n is non-decreasing over time, the private signals of earlier forecasters will be much more heavily weighted. Consequently, if earlier forecasts contain large errors, they will "drag" the final consensus away from the true mean.

We then examine the impact of herding on forecast accuracy in the next two propositions.

Proposition 2 *The mean squared error of the consensus of all private signals ($\bar{y}_N \equiv \frac{1}{N}\sum_{n=1}^{N} y_n$ is smaller than the consensus of all forecasts (c_N) for any $w_n \in (0, 1]$;*

Proof According to Lemma 1, the final consensus of all forecasts (c_N) is a weighted average of all private signals. Since the mean of all private signals is the actual earnings, c_N is an unbiased estimator, and the mean squared error is the variance of c_N.

$$\begin{aligned} Var(c_N) &= Var(\sum_{n=1}^{N} l^N(n) y_n) \\ &= \sum_{n=1}^{N} l^N(n)^2 \sigma^2 \end{aligned}$$

According to Jensen's inequality,

$$\frac{\sum_{n=1}^{N} l^N(n)^2}{N} \geq (\frac{\sum_{n=1}^{N} l^N(n)}{N})^2$$

Therefore,

$$Var(c_N) \geq N \cdot (\frac{\sum_{n=1}^{N} l^N(n)}{N})^2 \sigma^2 = \frac{1}{N}\sigma^2$$

The equality holds if and only if $w_1 = w_2 = \cdots = w_N$ (or $w = 0$), which are the weights for the consensus of all private signals. In other words, the mean squared error of the consensus of all private signals is smaller than the consensus of all forecasts ($\bar{y}_N \equiv \frac{1}{N}\sum_{n=1}^{N} y_n$) for any $w \in (0, 1]$.

Proposition 2 is a simple result of Jensen's inequality. Herding places unequal weights on different private signals, making the resulting weighted average a less efficient estimator of the mean. Of course if the weight (w_n) is known, one can always back out the private signals (y)

from forecasts (f) and consensus (c) and reproduce the efficient mean estimate. In the more likely case where the weight (w_n) is unknown, directly observing the private signals and computing their average still produces the most efficient estimator.

Proposition 3 *The mean squared error of the forecast (f_n) is smaller than that of the private signal (y_n).*

Proof Since the forecast is also an unbiased estimator, the mean squared error of f_n is the variance of f_n. According to the definition,

$$Var(f_n) = (1 - w_n)^2 Var(y_n) + w_n^2 Var(c_{n-1})$$

We can easily prove that $Var(c_{n-1}) \leq \sigma^2$, because $Var(c_{n-1}) - \sigma^2 = \sum_{i=1}^{n-1} l^{n-1}(i)^2\sigma^2 - \sum_{i=1}^{n-1} l^{n-1}(i)\sigma^2 = \sum_{i=1}^{n-1} l^{n-1}(i)(l^{n-1}(i) - 1)\sigma^2 \leq 0$.
Therefore,

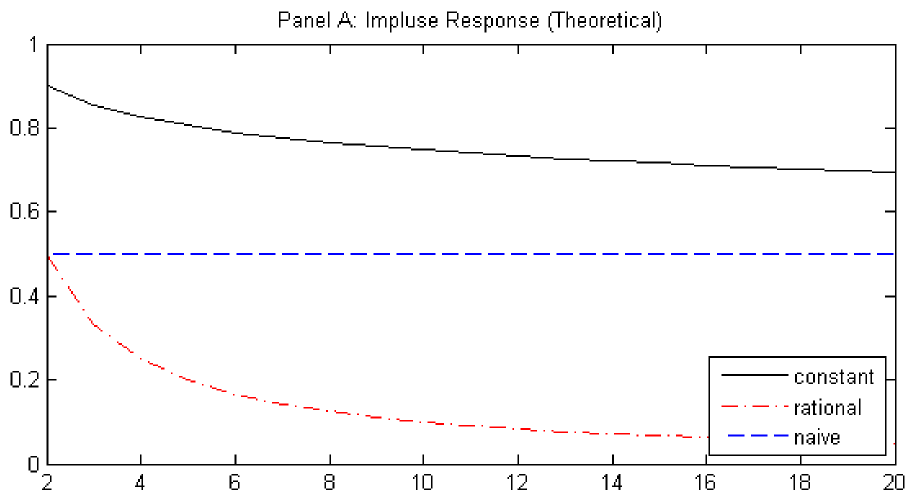
$$\begin{aligned} Var(f_n) &\leq (1 - w_n)^2\sigma^2 + w_n^2\sigma^2 \\ &= (1 + 2w_n \underbrace{(w_n - 1)}_{<0})\sigma^2 \\ &< \sigma^2 = Var(y_n) \end{aligned}$$

According to Proposition 3, herding makes each individual forecast more accurate on average. This is because each forecast puts a positive weight on the current consensus and the current consensus, being the average of multiple private signals, has a lower variance than each private signal does. Importantly, herding behavior, while it improves individual forecast accuracy, could make the forecast consensus less efficient.

So far, we consider a simple case where the precision of the private signal is fixed and identical across individuals. This assumption may not hold in reality. For example, observing the difference between her private signal (y_n) and the current consensus (c_{n-1}) may encourage the individual to increase her information production effort and therefore improve the precision of her private signal. In addition, the individual with the most precise private signal may be incentivized to always issue her forecast first. In both cases, herding may actually improve the precision of the final consensus forecast (c_N). In other words, the impact of herding on the accuracy of consensus forecast is ultimately an empirical question we examine on Estimize.com.

Figure A1: **Differentiating Herding Behaviors Using Impulse Responses**

Panel A plots the theoretical impulse response functions corresponding to different herding behaviors. The impulse response function measures the impact of the first forecast on subsequent forecasts. The dot-dashed line represents the rational case where individuals always equally weight all prior private signals. The dashed line represents the case of naive herding where individuals always equally weight all prior forecasts. The solid line represents the case where individuals always place a constant weight (0.9) on the current consensus. Panel B plots the empirical impulse response function estimated from the Estimize data.



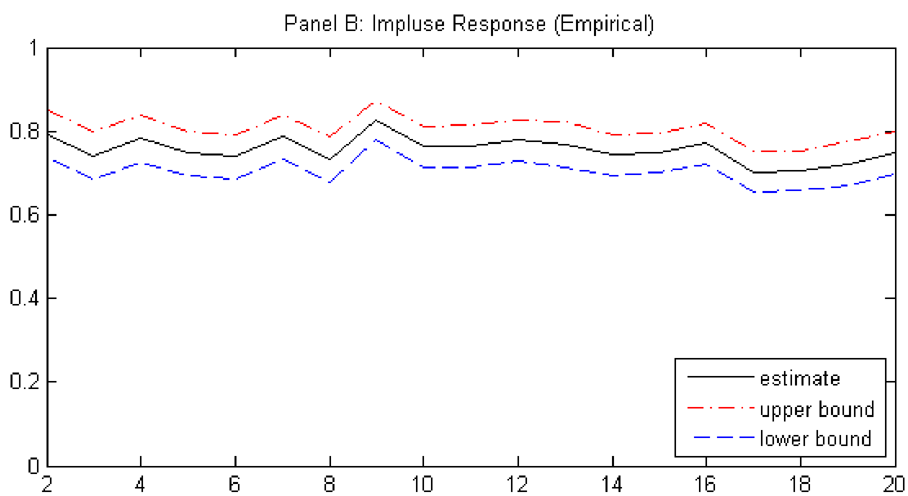


Figure 1: **Coverage of Estimize Sample over Time**

The figure plots the number of users, releases, and estimates in each quarter covered by our sample. Our main sample covers releases with at least 10 estimates, from March 2012 to March 2015. The left axis represents the number of users and releases, and the right axis represents the number of estimates.

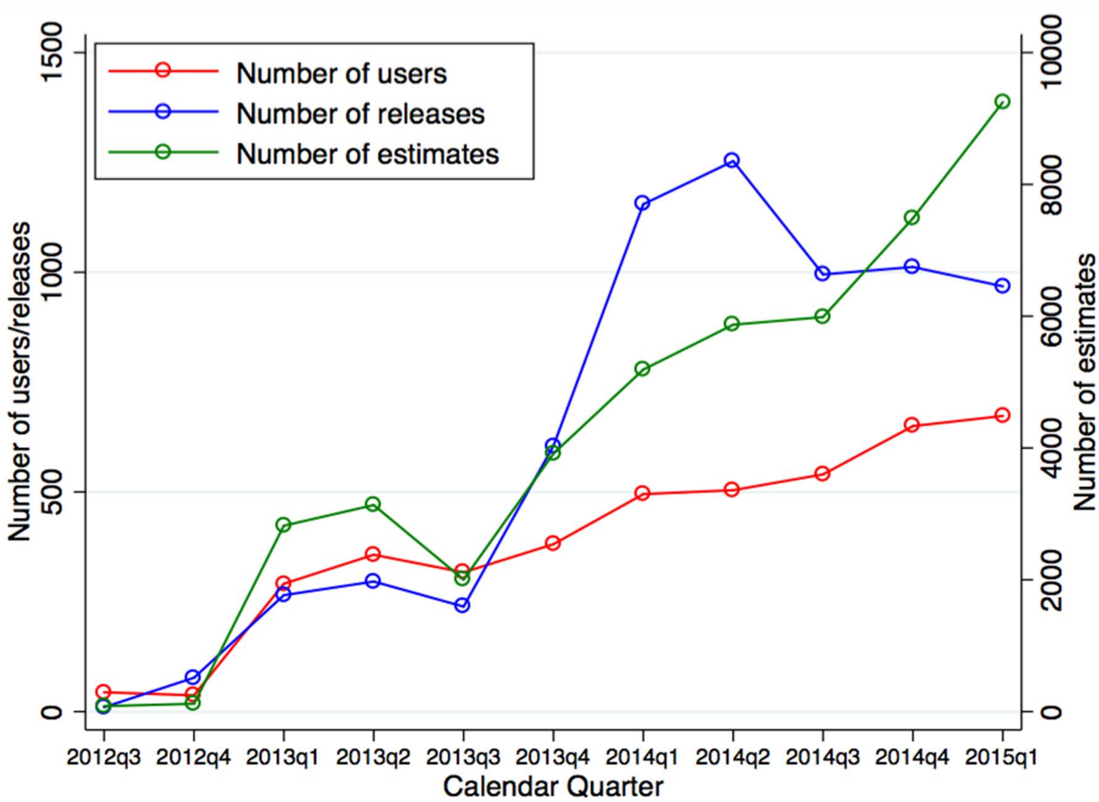


Figure 2: **Example of a Release Page**

The figure presents a screenshot of the release page for Facebook, Inc. (FB) for the second fiscal quarter of 2015. The left chart plots the historical data of actual EPS, the range and consensus of Wall Street forecasts, and the range and consensus of Estimize forecasts. It also includes the current Wall Street and Estimize consensus. The right chart lists all Estimize estimates of FB's EPS underlying the current Estimize consensus.

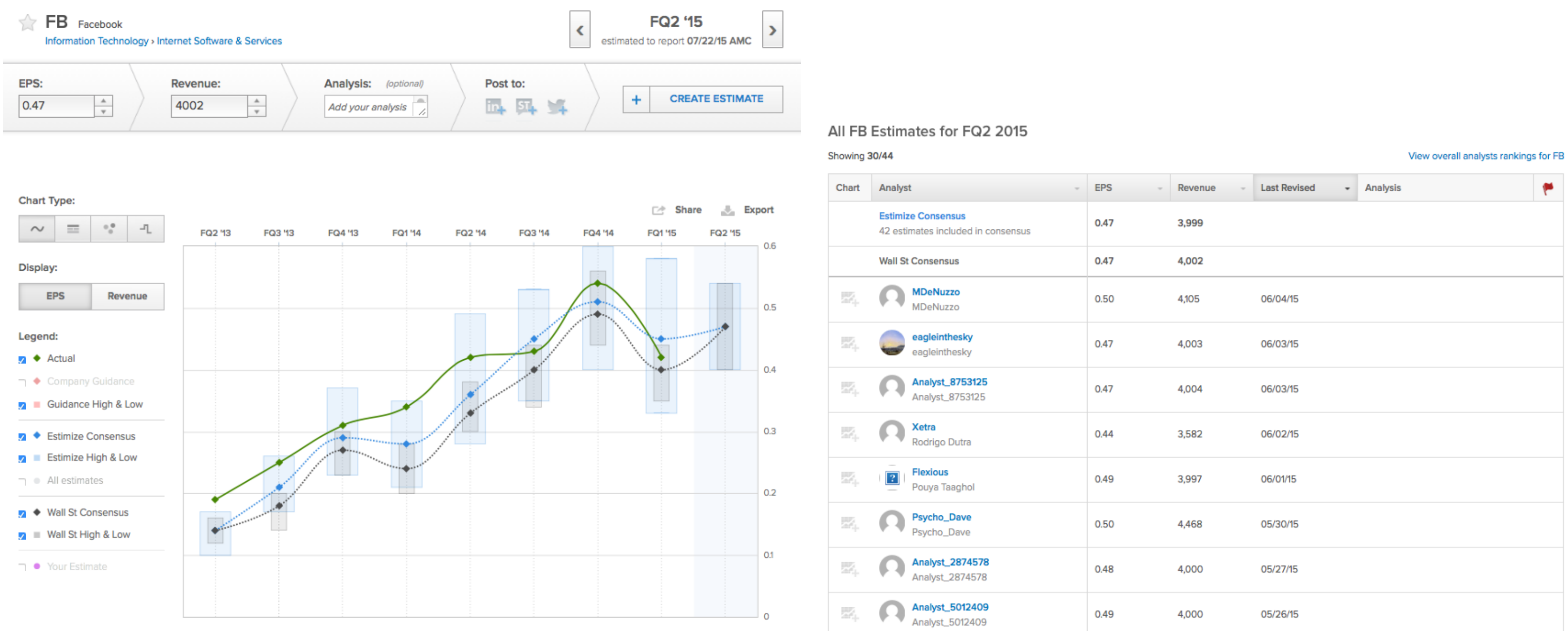


Figure 3: **Example of a Blind View**

The figure presents a screenshot of a blind release page for Lululemon Athletica Inc. (LULU) for the fourth fiscal quarter of 2015. The left chart plots the historical data of actual EPS, the range and consensus of Wall Street forecasts, and the range and consensus of Estimize forecasts. Note no information on the Estimize consensus is provided for the fourth quarter. The right chart shows that all individual Estimize estimates of LULU's EPS are also hidden.

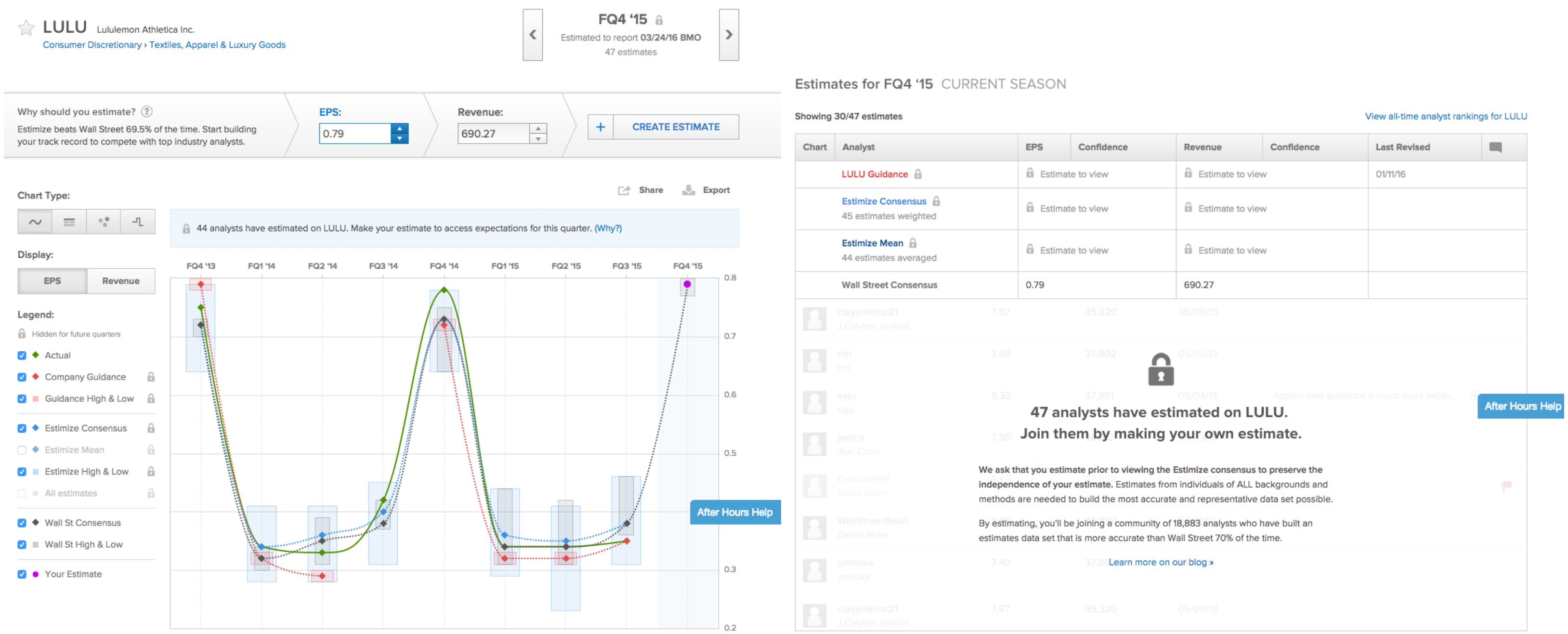


Figure 4: **Measuring Estimze User Influence Using PageRank**

The figure presents an example of how we measure user influence on the Estimize network using the PageRank method. Users are represented by circles and are linked by arrows which capture viewing activities. For example, when user D views user A, it results in an arrow going from user D to user A. An influential user (represented by a bigger circle) either receives more incoming arrows (as in the case of user B) or receives an arrow from another influential user (as in the case of user C). User influence is measured by the PageRank, reported inside the circle.

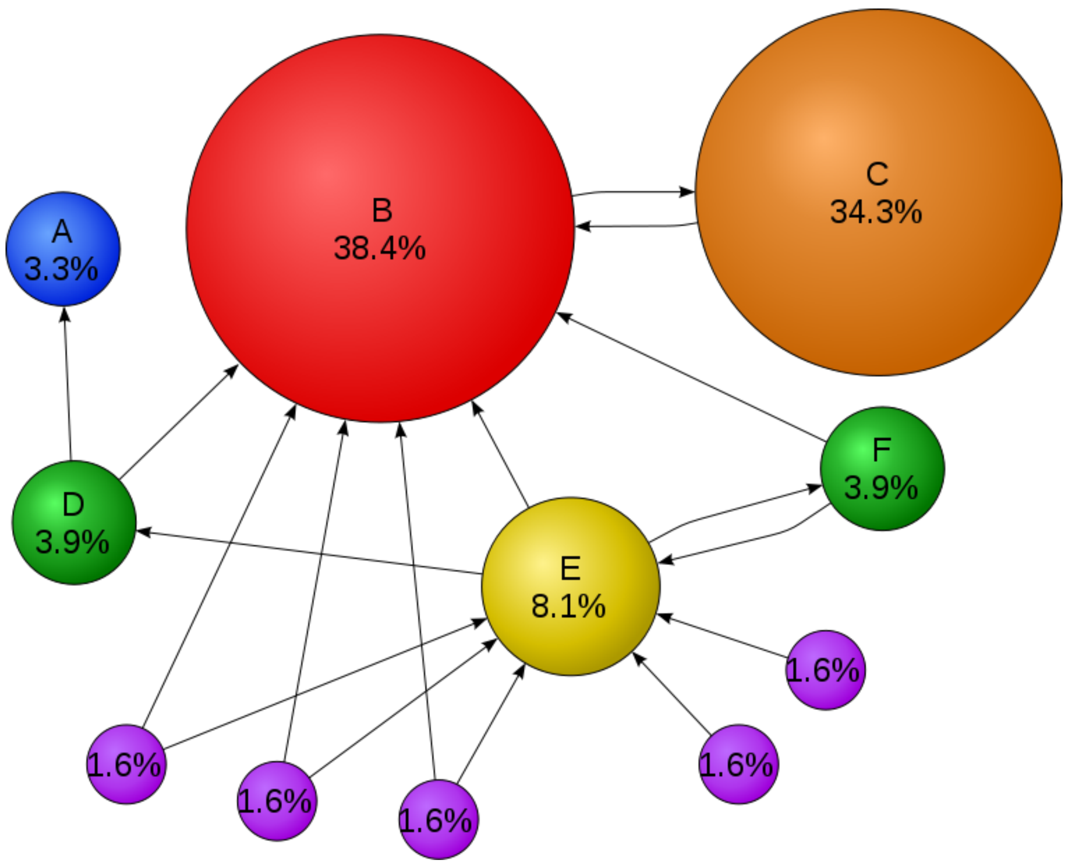


Table 1: **: Descriptive Statistics for the Estimize Sample**

The table presents descriptive statistics for the forecasts made on Estimize from March 2012 to March 2015. The sample covers 2,147 releases with at least 10 estimates. Panel A reports release-level forecast characteristics. Panel B reports release-level financial characteristics. The sample contains 1,953 releases with at least 10 estimates and matched financial data from Compustat. The size group and B/M group are obtained by matching each release with one of 25 size and B/M portfolios at the end of June based on market capitalization at the end of June and B/M, the book equity of the last fiscal year end in the prior calendar year divided by the market value of the equity at the end of December of the prior year. Panel C reports user-level characteristics. Panel D reports the sector distribution of the 730 distinct stocks in our sample. Panel E reports the distribution of users in our sample by their professions.

	mean	sd	p1	p25	p50	p75	p99
Panel A: Release-level Estimize Forecast Characteristics (#Obs = 2147)							
Number of forecasts	20.03	15.01	10.00	12.00	15.00	23.00	74.00
Number of distinct users	16.08	11.92	4.00	10.00	13.00	19.00	59.00
Number of release views	18.97	32.85	1.00	7.00	12.00	21.00	114.00
Consensus error (=consensus-actual)	-0.02	0.19	-0.51	-0.05	-0.01	0.02	0.31
Abs (consensus error)	0.08	0.18	0.00	0.01	0.03	0.08	0.63
Estimize abserr - WS abserr	-0.01	0.10	-0.17	-0.02	-0.01	0.01	0.14
% of release view	35.60%	17.20%	0.00%	23.08%	35.29%	46.94%	76.47%
Panel B: Release-level Financial Characteristics (#Obs = 1953)							
Size (in million)	24512.35	46548.06	430.66	2896.06	7635.54	21862.95	241171.05
B/M	0.40	0.40	0.03	0.18	0.31	0.49	2.08
Size group (=1: bottom 20%; =5, top 20%)	3.91	1.16	1.00	3.00	4.00	5.00	5.00
B/M group (=1: bottom 20%; =5: top 20%)	2.04	1.26	1.00	1.00	2.00	3.00	5.00
Panel C: User-level Characeristics (#Obs = 2516)							
Number of tickers covered	10.22	35.68	1.00	1.00	2.00	5.00	181.00
Number of forecasts submited	17.10	78.43	1.00	1.00	2.00	6.50	320.00

Panel D: Distribution of Stocks by Sector

Sector	Freq	Pct
Consumer Discretionary	146	20.00
Consumer Staples	47	6.44
Energy	40	5.48
Financials	40	5.48
Health Care	76	10.41
Industrials	95	13.01
Information Technology	224	30.68
Materials	41	5.62
Telecommunication Services	7	0.96
Utilities	14	1.92
Total	730	100.00

Panel E: Distribution of Users by Profession

	Freq	Pct
Financial Professionals:		
Buy Side	281	11.41
Sell Side	158	6.42
Independent	381	15.48
Non Professionals:		
Information Technology	519	21.08
Student	493	20.02
Financials	142	5.77
Consumer Discretionary	110	4.47
Health Care	94	3.82
Others	284	11.54
Total	2462	100.00

Table 2: **: Release Views and Weighting of Information**

The table presents the results of forecast-level weighting regressions. The dependent variable is forecast error, which is defined as the difference between a user's forecasted EPS and the actual EPS. The main independent variables include: (1) Dev: the forecast's distance from the consensus prior to the submitted forecast, (2) Nonzero views: a dummy variable for viewing the release page for longer than 5 seconds at least once, (3) the interaction term between Dev and Nonzero views. Standard errors are in parentheses and double-clustered by sector and quarter. ***, **, * - significant at the 1, 5, and 10% level.

Panel A: Forecast-level analysis

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Forecast error (= Forecast-Actual)				
Dev	0.424***	0.425***	0.489***	0.489***	0.470***
(= Forecast - Pre consensus)	(0.087)	(0.087)	(0.077)	(0.077)	(0.057)
Dev X Nonzero Views	-0.274***	-0.274***	-0.250**	-0.250**	-0.218***
	(0.090)	(0.090)	(0.110)	(0.110)	(0.062)
Nonzero Views	0.00177	-0.00129	0.00102	0.00160	0.000262
	(0.002)	(0.005)	(0.001)	(0.001)	(0.001)
Release effect	No	No	Yes	Yes	Yes
Profession effect	No	Yes	No	Yes	No
Individual effect	No	No	No	No	Yes
Observations	30429	30429	30429	30429	30429
R-squared	0.034	0.035	0.917	0.918	0.934

Table 3: **: Release Views and Forecast Accuracy**

Panel A presents the results of forecast-level regressions. The dependent variable is the absolute value of forecast error. Forecast error is defined as the difference between a user's forecasted EPS and the actual EPS. The main independent variable is "Nonzero Views," a dummy variable that is equal to 1 when the user views the release page for longer than 5 seconds at least once. The control variables include a Close-to-Announcement (CTA) dummy that is equal to 1 if the forecast was issued in the last three days before the announcement, and release, individual, and user profession fixed effects. Panel B presents the results of release-level regressions. The dependent variable is the absolute value of forecast error. The main independent variable is the logarithm of one plus the ratio of the number of forecasts made following release views longer than 5 seconds to the number of total forecasts. The control variables include the standard deviation of forecast error normalized by the absolute value of median forecast error, and sector and quarter fixed effects. Standard errors are in parentheses and double-clustered by sector and quarter. ***, **, * - significant at the 1, 5, and 10% level. Panel C presents the results of a comparison of the consensus of forecasts with release views versus the consensus of forecasts without release views within each release.

Panel A: Forecast-level analysis

	(1)	(2)	(3)
	Abs(FE = Forecast - Actual)		
Nonzero Views	-0.00737***	-0.00731***	-0.00267***
	(0.001)	(0.001)	(0.001)
CTA dummy	-0.0119***	-0.0116***	-0.00678***
	(0.002)	(0.002)	(0.001)
Release Effect	Yes	Yes	Yes
Profession Effect	No	Yes	No
Individual Effect	No	No	Yes
Observations	37674	37674	37674
R-squared	0.824	0.824	0.835

Panel B: Release-level analysis

	(1)	(2)	(3)
Dependent variable:	Abs(FE = Consensus Forecast - Actual)		
LnNumView	0.0450***	0.0401***	0.0551***
	(0.008)	(0.006)	(0.010)
Std Dev (FE) / Abs (Median(FE))		0.0517**	0.0508**
		(0.022)	(0.020)
Sector effect	No	No	Yes
Quarter effect	No	No	Yes
Observations	2122	2122	2122
R-squared	0.011	0.025	0.110

Panel C: Within-release horse race

	Freq. of Releases	Percentage
Average of forecasts with release views wins	867	40.76%
Average of forecasts without release views wins	1260	59.24%
Binomial test p-value:	<0.0001	

Table 4: **: Release Views and Lead-lag Biases**

The table presents the results of forecast-level regressions. The dependent variable is a dummy variable which is equal to one if earlier and close-to-announcement estimates are biased in the same direction. The close-to-announcement window is defined as extending from five days before the announcement date through the announcement date ([-5,0]). The early window is defined as days prior to day -5. The consensus within the window is upwardly (downwardly) biased if the difference between the consensus and the actual EPS is above the H-th percentile (below the L-th percentile). The main independent variable is the logarithm of one plus the ratio of the number of forecasts made following release views longer than 5 seconds to the number of total forecasts within the close-to-announcement window. The control variables include the standard deviation of forecast error normalized by the absolute value of median forecast error, and sector and quarter fixed effects. Standard errors are in parentheses and double-clustered by sector and quarter. ***, **, * - significant at the 1, 5, and 10% level.

	(1)	(2)	(3)
Dependent Variable:	Consistent bias indicator		
Bias is defined as (average forecasts-actual) above H-th percentile or below L-th percentile	H=60, L=40	H=70, L=30	H=80, L=20
LnNumView	0.221*	0.292***	0.466***
	(0.120)	(0.112)	(0.115)
Sector effect	Yes	Yes	Yes
Quarter effect	Yes	Yes	Yes
Observations	1770	1770	1770
Pseudo R2	0.0317	0.0359	0.0614

Table 5: **: Blind Experiment: Blind vs. Default**

When a user is randomly selected to participate in the experiment, she is asked to make an earnings forecast while the release page is disabled. The resulting forecast is labeled the blind forecast (f_b). Each blind forecast is matched with the closest estimate in the sequence made by a different user who could view the release page. The matched estimate is labeled the default forecast. The pair is removed if the time difference between the blind estimate and the default estimate exceeds 24 hours. The final sample contains releases with at least 15 matched pairs. Blind and default forecasts are pooled in the regression. In Panel A, we compare user characteristics between the blind group and the default group. Column (1) represents the percentage of professional users. Column (2) represents the percentage of users with the number of estimate view over the 80th percentile. Column (3) represents the average number of releases submitted by the users in each group. Column (4) represents the average number of firms covered by the users in each group. Column (5) represents the average absolute value of forecast errors of users in each group. Column (2) - (5) are based on users' forecasts before the experiment. T-stat and p-value of the difference between these two groups are also reported. In Panel B, the dependent variable is the absolute forecast error. The control variables include a Close-to-Announcement (CTA) dummy that is equal to 1 if the forecast was issued in the last three days before the announcement, and release, individual, and user profession fixed effects. In Panel C, the dependent variable is the forecast error defined as the difference between the blind forecast and the actual EPS. Independent variables include: (1) Dev: the forecast distance from the consensus prior to the submitted forecast, (2) Default: a dummy variable that is equal to one if it is a default forecast, and zero if it is a blind forecast, (3) the interaction term between Dev and Default. Standard errors are in parentheses and clustered by ticker. ***, **, * - significant at the 1, 5, and 10% level.

Panel A: Blind vs. default users

	(1)	(2)	(3)	(4)	(5)
	Professional	Highly viewed	#releases	#tickers	Abs(FE)
Blind	27.23%	19.96%	102	55	0.111
Default	27.10%	17.88%	94	53	0.115
Diff (Blind - Default)	0.13%	2.08%	8	2	-0.004
T-stat	0.07	0.69	0.29	0.16	0.32
p-value	0.947	0.49	0.770	0.871	0.750

Panel B: Release view and forecast accuracy

	(1)	(2)
Dependent variable:	Abs (Forecast error)	
Default	-0.00221**	-0.00253**
	(0.001)	(0.001)
CTA dummy	-0.00648***	-0.00647***
	(0.001)	(0.001)
Release effect	Yes	Yes
Profession effect	Yes	Yes
Individual effect	No	Yes
Observations	8630	8630
R-squared	0.825	0.825

Panel C: Release view and information weighting

	(1)	(2)
Dependent variable:	Forecast error (= Forecast-Actual)	
Dev	0.670***	0.670***
(= Forecast - Pre consensus)	(0.068)	(0.068)
Dev X Default	-0.113**	-0.113*
	(0.059)	(0.058)
Default	-0.00311**	-0.00299*
	(0.001)	(0.002)
Release effect	Yes	Yes
Profession effect	No	Yes
Observations	8198	8198
R-squared	0.956	0.956

Table 6: **: Blind Experiment: Blind vs. Revised**

We consider the blind and revised forecasts from the pilot experiment. Panel A presents the results of forecast-level weighting regressions. The dependent variable is the difference between the blind forecast and the revised forecast from the same user in the blind experiment. The main independent variable is the blind forecast's distance from the consensus prior to the submitted forecast. Standard errors are in parentheses and clustered by ticker. ***, **, * - significant at the 1, 5, and 10% level. Panel B presents the results of within-release horseraces between the blind consensus and the revised consensus. When calculating the revised consensus, we fill the forecast with the initial one for users who choose not to revise their forecasts. The forecast error (FE) is defined as the difference between the consensus and the actual EPS.

Panel A: Forecast-level weighting regression

	(1)	(2)
Dependent variable:	Forecast (Blind) - Forecast (Revised)	
Forecast (Blind) - Pre-Consensus	0.523***	0.534***
	(0.050)	(0.051)
Sector Effect	No	Yes
Observations	104	104
R-squared	0.466	0.481

Panel B: Within-release horserace: Blind consensus vs Revised consensus

Ticker	Total number of blind users	Blind FE	Revised FE	p-value
WFM	20	-0.0020	-0.0030	0.05
UA	40	0.0170	0.0173	0.05
F	24	0.0342	0.0392	0.00
CMG	35	-0.2114	-0.2086	1.00
AA	22	-0.0059	-0.0059	1.00
XOM	19	-0.3195	-0.3232	0.05
BAC	16	0.0263	0.0306	0.06
GS	17	-1.6682	-1.6812	0.00
GILD	58	-0.5341	-0.5047	1.00
JNJ	17	-0.0318	-0.0329	0.06
AAPL	133	-0.0744	-0.0745	0.06
FB	91	0.0227	0.0236	0.00
FEYE	16	0.0344	0.0369	0.00
				# Blind wins (p<0.1)
				10

Table 7: **: Estimize Consensus: Before March 2015 and After November 2015**

The table presents the comparison of the probability that Estimize consensus wins Wall Street consensus between the before-experiment period (before Mar 2015) and the after-experiment period (Nov 2015 - Feb 2016). The sample is limited to the same set of stocks before and after the experiment. Column (1) include all releases. Column (2), (3) and (4) include releases with the number of releases below 10, between 10 and 30, and greater than 30, respectively.

	(1)	(2)	(3)	(4)
Pr(Estimize consensus wins WS consensus)	All	#Estimates (0,10]	#Estimates (10,30]	#Estimates>30
N	1641	660	315	210
Before experiement	56.67%	54.85%	56.51%	62.38%
After experiement	64.11%	64.24%	64.44%	66.67%
Z-test (After-Before)	6.28	5.03	2.94	1.32
p-value	0.000	0.000	0.002	0.090

Table 8: **: The Impact of Influential Users on the Weighting of Information**

The table presents the results of forecast-level weighting regressions. The dependent variable is forecast error, which is defined as the difference between a user's forecasted EPS and the actual EPS. The main independent variables include: (1) Dev: the forecast's distance from the consensus prior to the submitted forecast, (2) Nonzero views: a dummy variable for forecasts made after viewing the release page for longer than 5 seconds at least once, (3) Influenced: a dummy variable that is equal to one if the number of influential users ahead of the observed user is above the 80th percentile across all observations, and the interaction terms among these three variables. To identify influential users, we consider four measures: (1) PageRank, (2) number of releases, (3) number of releases being viewed, (4) probability of being a leader. The measures for users who submit fewer than 20 forecasts are assigned to the lowest value. The users who rank above 80th percentile on the measure are identified as influential users. Standard errors are in parentheses and double-clustered by sector and quarter. ***, **, * - significant at the 1, 5, and 10% level.

	(1)	(2)	(3)	(4)
Dependent variable:	Forecast error (= Forecast - Actual)			
Measure of Influential Users	PageRank	Number of releases	Number of releases being viewed	Prob of being leader
Dev	0.496***	0.496***	0.485***	0.498***
(= Forecast - pre consensus)	(73.26)	(74.60)	(71.10)	(75.61)
Dev X Nonzero Views	-0.206***	-0.214***	-0.200***	-0.210***
	(-28.31)	(-29.92)	(-27.28)	(-29.53)
Dev X Nonzero Views X Influenced	-0.134***	-0.125***	-0.153***	-0.140***
	(-10.95)	(-9.93)	(-12.57)	(-10.97)
Dev X Influenced	0.0940***	0.108***	0.128***	0.107***
	(8.13)	(8.99)	(11.07)	(8.73)
Nonzero Views X Influenced	-0.00341***	-0.00249***	-0.00193**	-0.00192**
	(-3.74)	(-2.72)	(-2.11)	(-2.09)
Influenced	0.00202**	0.000808	0.00106	0.000873
	(2.26)	(0.89)	(1.21)	(0.97)
Nonzero Views	0.00262***	0.00214***	0.00184***	0.00185***
	(4.17)	(3.41)	(2.95)	(2.97)
Release effect	Yes	Yes	Yes	Yes
Profession effect	Yes	Yes	Yes	Yes
Observations	33264	33264	33264	33264
R-squared	0.920	0.920	0.920	0.920

Table 9: **: Predicting the Change in Consensus Error and Change in Consensus Accuracy from the Early to the Close-to-announcement Period**

The table presents the results of release-level regressions. Panel A regresses the change in the consensus error from the early to the close-to-announcement period on four variables constructed on forecasts made by influential users in the early period. All forecasts made by influential users in the early period are sorted into four groups by two dimensions: (1) whether the forecast leads to an upward or a downward revision of the consensus, and (2) whether the cumulative abnormal returns (CAR) on the corresponding day of the forecast are positive or negative. The main independent variables are the logarithm of one plus the number of forecasts in each group. Panel B uses the same set of independent variables, while the dependent variable is the change in the absolute value of the consensus error. The close-to-announcement window is defined as from five days before the announcement date through the announcement date ([-5,0]). The early window is defined as days prior to day -5. Standard errors are in parentheses and double-clustered by sector and quarter. ***, **, * - significant at the 1, 5, and 10% level.

	(1)	(2)	(3)	(4)
Panel A: Predicting the change in consensus error				
Dependent variable:	Change in concensus error			
Measure of Influential Users	PageRank	Number of releases	Number of releases being viewed	Prob of being leader
ln(1+Num of Upward revision, Neg CAR)	0.003**	0.003**	0.003**	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
ln(1+Num of Upward revision, Pos CAR)	0.006***	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
ln(1+Num of Downward revision, Pos CAR)	-0.005***	-0.005***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
ln(1+Num of Downward revision, Neg CAR)	-0.007***	-0.007***	-0.007***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-0.008***	-0.005	-0.008***	-0.005
	(0.003)	(0.003)	(0.003)	(0.003)
Observations	1988	2004	2004	2004
R-squared	0.098	0.102	0.100	0.102
Panel B: Predicting the change in accuracy				
Dependent variable:	Change in Abs(consensus error)			
ln(1+Num of Upward revision, Neg CAR)	0.002*	0.003**	0.003**	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
ln(1+Num of Upward revision, Pos CAR)	-0.001	-0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
ln(1+Num of Downward revision, Pos CAR)	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
ln(1+Num of Downward revision, Neg CAR)	-0.002*	-0.002*	-0.002*	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-0.006**	-0.005**	-0.006**	-0.005**
	(0.003)	(0.003)	(0.003)	(0.003)
Observations	1988	2004	2004	2004
R-squared	0.031	0.028	0.028	0.028

Table 10: **: Predicting the Cumulative Abnormal Return Around Announcement Date**

The table presents the results of regressing the cumulative abnormal returns during [-1,1] around the announcement date on four variables constructed on forecasts made by influential users during the period before day -1. All forecasts made by influential users in that period are sorted into four groups by two dimensions: (1) whether the forecast leads to an upward or a downward revision of the consensus, and (2) whether the cumulative abnormal returns (CAR) on the corresponding day of the forecast are positive or negative. The main independent variables are the logarithm of one plus the number of forecasts in each group. Standard errors are in parentheses and double-clustered by sector and quarter. ***, **, * - significant at the 1, 5, and 10% level.

	(1)	(2)	(3)	(4)
Dependent variable:	CAR[-1,1]			
Measure of Influential Users	PageRank	Number of releases	Number of releases being viewed	Prob of being leader
ln(1+Num of Upward revision, Neg CAR)	-0.006**	-0.007**	-0.007**	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
ln(1+Num of Upward revision, Pos CAR)	0.004	0.003	0.004	0.003
	(0.003)	(0.003)	(0.003)	(0.003)
ln(1+Num of Downward revision, Pos CAR)	-0.001	-0.002	-0.000	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)
ln(1+Num of Downward revision, Neg CAR)	-0.001	-0.004	-0.004	-0.005
	(0.003)	(0.003)	(0.003)	(0.003)
Time effect (Year-month)	Yes	Yes	Yes	Yes
Observations	1668	1687	1687	1687
R-squared	0.022	0.027	0.027	0.027