Harnessing the Wisdom of Crowds*

Zhi Da[†] and Xing Huang[‡]

First Draft: May 2015
This Draft: December 2015

Abstract

We show that extensive herding is harmful on a crowd-based earnings forecast platform (Estimize.com). By tracking user viewing activities, we monitor the amount of information a user viewed before she makes an earnings forecast. We find the more public information viewed, the more she will underweigh 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, consistent with herding. To address the endogeneity concerning the information acquisition choice, we collaborate with Estimize.com to run experiments where we restrict the information set on randomly selected stocks and users. The experiments confirm that "independent" forecasts lead to more accurate consensus and convince Estimize.com to switch to a "blind" platform from November 2015. Overall, our findings suggest that wisdom of crowd can be better harnessed by encouraging independent voice from the participants.

*We thank Qi Chen, Cary Frydman, Stefano DellaVigna, Umit Gurun, David Hirshleifer, Harrison Hong, Byoung-Hyoun Hwang, Yin Luo, Adam Szeidl 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 the information from individuals. A great deal of evidence suggests that, under certain conditions, a large group's average 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 independent since individuals learn from observing other people's actions and belief. Banerjee (1992) and Bikhchandani et al. (1992) show that it is rational for individuals to "imitate" or "herd with" other people. At the same time, excessive imitation is irrational and harmful. Eyster and Rabin (2014) show that in a broad class of settings, abundant imitation will lead to a positive probability of people converging to wrong long-run beliefs.

By directly measuring and randomizing on individuals' information set, we are able to better isolate the impact of herding on economic outcomes. We focus on a specific setting where individuals make corporate earnings forecasts. Both earnings forecasts and realizations are easily observable and the forecast error can be clearly defined. Accurate earnings forecasts are of crucial importance to investors, the firms and the well-functioning of the financial market in general. Not surprisingly, a wide range of market participants provide earnings forecasts. They include equity analysts from

[1]Examples include the war on Iraq, jury verdicts, the setting of interest rate by the Federal Open Market Committee (FOMC), and the appointment of 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.

both the sell-side and buy-side, and more recently independent analysts.

Unfortunately, a long strand of literature on sell-side analyst forecasts (IBES) provides ample evidence that the analyst consensus is problematic. This is because the two conditions underlying the wisdom of crowds are often violated. First, analyst forecasts are often biased, driven by investment banking relation (Lin and McNichols (1998); Michaely and Womack (1999)) or career concerns (Hong and Kubik (2003)) among other things. Second, since earnings forecasts are also made sequentially, they are correlated as a result of either informational herding, reputational herding, or naive herding (Scharfstein and Stein (1990); Banerjee (1992); Bikhchandani et al. (1992); Trueman (1994); Hong et al. (2000); Welch (2000); Clement and Tse (2005); Eyster and Rabin (2010); Eyster and Rabin (2014)).[3] In the extreme case of an information cascade, private information of the subsequent forecasters are completely discarded so the crowd consensus is no more accurate than the first forecast in the sequence.[4]

Isolating the impact of herding behavioral in consensus earnings forecast accuracy is challenging, since it requires the researchers to observe the counter-factual of what will happen if analysts make their forecasts independently. In this paper, we tackle this challenge by taking advantage of an unique dataset on user activities and by running randomized experiments on a crowed-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 1773 users studied in our sample, one third are financial analysts coming from buy-side, sell-side, and 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. (2015) document that Estimize consensus is a better proxy for the market

[3] See Hirshleifer and Teoh (2003) for an excellent survey of herding behaviour in capital markets.

[4] Information cascade rarely happens with earnings forecasts though, since earnings are drawn from a continuous distribution.

expectation than the IBES consensus. In addition, they find the consensus computed using both Estimize and IBES forecasts to be even more accurate. A contemporaneous study by Bliss and Nikolic (2015) also finds that Estimize consensuses are more accurate than traditional I/B/E/S 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 and forecasts from other users. As a result, similar 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 the release view are more likely to only reflect the private information of the user.

During our sample period from March 2012 to July 2014, we examine 1317 quarterly firm earnings (releases) with at least 10 forecasts prior to the announcement. These releases come from 471 distinct firms in various sectors. We find the release viewing activity to have significant impact on the forecasts. First, release view is associated with underweighing of private information and positive autocorrelation in forecast revisions, suggesting the herding behavior. Second, while the release view improves the accuracy of individual forecast, it makes the consensus less accurate. This is because some useful private information maybe discarded when user herds with the prior forecasts. In particular, biases in earlier forecasts are more likely to persist subsequently and appear in the final consensus forecast. These findings are consistent with herding behavior.

We also find such a herding behavior to become more severe if the public information set contains estimates of the "influential" users. We identify "influential" users based on either the amount of their forecasts, the accuracy of their forecasts, whether their forecasts are viewed by other uses, or whether their forecasts "lead" subsequent forecasts.

We find very similar results regardless which definition of "influential" user is used. First, users are more likely to underweigh their private information when the releases they viewed contain prior forecasts of "influential" users. When this happens, the accuracy of the consensus forecasts goes down. Second, when "influential" users issued 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. Third, such a herding behavior predicts the accuracy of the final consensus forecasts. When contemporaneous stock return is negative and "influential" users issue forecasts that are lower than the current consensus early on, final consensus is more accurate, consistent with the notion that "influential" users facilitate the incorporation of negative information. On the other hand, when contemporaneous stock return is negative and "influential" users issue forecasts that are higher than the current consensus nevertheless, final consensus becomes less accurate. In this case, "influential" users' forecasts likely reflect positive sentiment that are propagated to subsequent users and drag the consensus to the wrong direction. The market does not completely undo this positive bias and we observe a significant negative return during the earnings announcement window.

So far, the evidence using the unique release view information suggests that the influence we exert on each other can make the crowd collectively dumber. Our empirical tests are affected by the endogeneity associated with the viewing choice. One could argue that a user may choose to view the release page only when he has little private information.[5] However, as long as these forecasts are not biased, their average should still converge to the true earnings even though each forecast contains weak signal. The fact that the consensus forecasts aggregated from forecasts with release views is less accurate must in part reflect the influence of prior forecasts.

In order to address the endogeneity concerning the information acquisition choice nevertheless, we collaborate with Estimize.com to run experiments during the second and third quarter of 2015 where we restrict the public information set on randomly selected stocks and users. Specifically, for each randomly selected stock, we randomly select users and disable the release view function and ask them to make a "blind" forecast. Each "blind" forecast is then matched to a "default" forecast

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

issued at about the same time by an user who can view the release page. Compared to the "blind" forecast, the "default" forecast uses significantly less private information. More importantly, 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. In this case, the blind forecast can be viewed as the private signal and the revised forecast the combination of the private and public signals. We then compare the accuracy of two consensus forecasts: (1) "blind" consensus computed using all "blind" forecasts; and (2) "revised" consensus computed using all revised forecasts when release view is enabled. Out of the 13 stocks randomly selected in the pilot experiment, the "blind" consensus significantly outperform the "revised" consensus 10 times and the "revised" consensus wins only 2 times. They tie in the remaining case. In other words, our findings suggest that wisdom of crowd can be better harnessed by encouraging independent voice from the participants. These findings are so compelling that Estimize.com decided to switch to a "blind" platform on November 2015 where users make forecasts without seeing the current consensus.[6]

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 documented to be pervasive.[8] By measuring and randomizing on individual's information set on a large crowd-based earnings forecast

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

[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 security trading, security analysis, firm investment, financing, and reporting decisions.

platform, we are able to better isolate the impact of herding behavior on outcomes with direct real-life implications. Our findings also have broader implications regarding group judgment. Our results confirm that independent views are crucial for reaching efficient outcome in such a setting.

The remaining paper is organized as follows. Section 2 analyzes the impact of herding on consensus forecast accuracy in a simple theoretical framework and derives testable predictions. Section 3 briefly describes the Estimize data used in the paper. Section 4 empirically studies the impact of herding behavior on forecast accuracy using the unique Estimize viewing activity data. Section 5 presents further evidence from experiments and Section 6 concludes.

2 Herding and the Wisdom of Crowds

It is intuitive why herding can make individual forecast more accurate, yet at the same time the crowd consensus less accurate. 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.

When forecasts are made simultaneously, the crowd consensus will simply be the average of these private signals ($\bar{y}$). 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 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 with 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}):

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

When all individuals compute their forecasts rationally in a Bayesian manner, w_n will be $(n-1)/n$

and the individual forecast f_n, equal to the arithmetic average of all private signals up to n, will converge to the truth as n increases. Alternatively, the weight w_n can be a constant, similar in spirit to the naive herding behavior described in Eyster and Rabin (2010) 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. As a result, individual forecast f_n will never converge to the truth. But as long as the weight w_n is positive, early forecasts will exert some influence on later forecasts.

In our simple framework, the consensus after the nth individual submits her forecast is:

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

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

Lemma 1 *The final consensus of all forecasts could 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. In Appendix A. ■

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, private signals of earlier forecasters will receive much heavier weights. 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 $w \in (0, 1]$;*

Proof. In Appendix A. ■

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 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. In Appendix A. ■

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. Importantly, herding behavior, while improving the individual forecast accuracy, makes the forecast consensus less efficient.

The rest of our paper quantifies the impact of herding empirically using the earnings forecast data on a crowd-based forecasting platform.

3 Data and Sample Description

3.1 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, increasingly more contributors have joined the platform and the coverage of firms has also significantly expanded. As of May 2015, more than 7000 analysts contribute on the platform, resulting in coverage on over 1500 stocks each quarter.

Different from IBES, Estimize solicits contribution from a wide range of individuals, for example financial analysts, such as sell-side, buy-side or independent analysts, as well as non-professionals, such as students, private investors, industry experts, etc. With the contribution from individuals with diverse background and viewpoints, Estimize better represents the market's true expectation than the IBES consensus and could serve as a supplementary source of information to IBES, as documented by Jame et al. (2015) and Bliss and Nikolic (2015).

There are a few reasons why contributors have incentives to provide information and increasingly contribute to Estimize. First, contributors (e.g., independent analysts and students) could create a verifiable track record of their accuracy and foresight for the fundamental metrics. Second, Estimize implements a point system which rewards forecasts 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. Points winners get recognized on their website and and featured in podcasts, and awarded with a prize package, such as apple watch. Third, the goodwill factor may motivate some users to participate in the platform just for the sake of its success — the more the contribution, the more valuable the dataset is to everyone.

3.2 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 July 2014. The forecasted EPS value and the corresponding time at which the forecast was created are both provided.

The second dataset contains background information on users in the Estimize community. Based on a brief personal profile voluntarily provided by users themselves, Estimize classifies users in several career-biographical categories, such as buy-side and sell-side professionals, industry experts, students, etc.[9]

The third dataset records the users' entire web activities on Estimize.com, including the pages that users view, the actions that users take (e.g., creating forecasts), and 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 of Facebook, Inc. (FB). The release page contains two charts as shown in the figure. The left chart presents the actual EPS, the range and consensus of Wall Street forecasts, the range and consensus of Estimize forecasts for the current quarter and past few 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 could also click any individual listed in the right chart to access an estimate page which presents all forecasts created by that individual. We also exploit the number of one's estimate page viewed by other users to construct a measure for influential users.

3.3 Sample Construction

We match the information on the forecasts and on the web activities to form a comprehensive dataset with forecast-level observations, covering the period from March 2012 through July 2014.[10] For each forecast created by a user, we count how many times she views the related release page longer than 5 seconds.[11]

The initial sample includes 46084 forecasts with 10084 releases that have announced quarterly

[9] 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.

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

[11] We set a cutoff for the length staying on one page, because we want to exclude the cases where a user just passes a page to access the next one. Nonetheless, same results will hold if we count the release views without a minimum viewing time.

earnings before July 31st, 2014. We drop forecasts where the users can not be successfully linked with an identifier in the activity dataset. We also exclude forecasts that are flagged manually or algorithmically unreliable by Estimize.[12] Finally, in order to ensure a decent size of crowd for each release, we only consider in our analysis the releases with at least 10 forecasts.

3.4 Descriptive Statistics

Our final sample consists of 21778 forecasts with 1317 releases. Figure 1 presents the coverage of our sample over time and demonstrates a trend of increasing number 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 on the release level. On average, about 17 users contribute 20 forecasts to a single release. There is a fair amount of web activities of viewing the release page — the average release has around 20 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 at making forecasts for some releases (e.g., only one release view), while they may check the release pages frequently for other releases (e.g., more than 100 release views). The wide range of release viewing activities provides considerable variation across releases.

The "runs test p-value" is the p value of a runs test of the hypothesis that the EPS forecasts occur in a random order by counting how many runs there are above and below the consensus. Small p-value indicates a highly correlated forecast sequence. The average (median) of the p-values is 0.41 (0.38), modestly smaller than 0.5, which indicates that the forecast sequences in the sample generally have higher correlation than a random sequence would suggest. The average consensus on Estimize is slightly pessimistic with an average consensus error -0.02. The average absolute value of the consensus error is 0.07, which is slightly more accurate than the average absolute value of

[12] 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.

the wall street consensus error.

We also obtain financial characteristics data from Compustat. Panel B presents the size and book-to-market (B/M) statistics for release-level observations.[13] To compare the financial characteristics with NYSE stocks, we also report the statistics of the size and B/M NYSE quintile group for firms in our sample.[14] The average firm size is $29.7 billion, while the median firm size is considerably smaller (about $10 billion). The average B/M ratio is 0.41 and the median B/M is 0.31. Our sample covers significantly larger firms compared to 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, consumer staples, and etc. 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 1824 users (Panel C). The average user covers 9 firms and contributes 14 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 background. As Panel E shows, 34.51% of the contributors studied in our sample are financial professionals, including sell-side (6.43%), buy-side (11.84%) and independent analysts (16.24%). The rest of contributors are non-professionals. Two major groups of non-professionals are industry experts and students (16.92%).

4 Empirical Analysis

Taking advantage of the the unique release view information, we examine the impact of herding on the behavior and accuracy of consensus earnings forecasts.

[13] Only 1266 out of 1317 release-level observations are successfully matched with data from Compustat.

[14] 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.

4.1 Release view and weighing of information

We first examine how release view affects the relative weighting between private and public information when a user makes a forecast. We follow the empirical framework in 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' best forecast according to the 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 efficient weight h in reality. Instead, the actual forecast f could be $f = ky + (1 - k)c$. Chen and Jiang (2006) shows that when regressing forecast error ($FE = f - z$) on 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 ($k > h$).

Panel A of 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 overweighting their private signals on average. Comparing to sell-side equity analysts studied in Chen and Jiang (2006), Estimize users seem to overweight their private signals a little more. Most importantly, we find a significant negative coefficient on the interaction term between *Dev* and the release view dummy. The result suggests that viewing of the current consensus, not surprisingly, is associated with more weight on the consensus and less weight on the private signal, consistent with the "herding" behavior.

Panel B of Table 2 then links the release view to herding behavior at the release level. We again use the "runs test p-value" as a measure of herding behavior. A smaller p-value implies stronger autocorrelation in the forecast revisions, which in turn reflects a more severer herding tendency. In the regressions, we find significant negative coefficients on the release view dummy, confirming the fact that more viewing of public information is associated with forecast revisions that are more autocorrelated.

4.2 Release view and forecast accuracy

How does viewing of public information affect the forecast accuracy? We first examine this question at individual forecast level. In Panel A of Table 3, we regress absolute forecast error on the release view dummy. We find a significant negative coefficient in column (1). Release view reduces the forecast error by more than 0.30 cents. Of course, forecast accuracy is also driven by the uncertainty associated with the earnings. In column (2), we control for such an uncertainty using the dispersion of forecast errors and result actually becomes stronger. Another concern is about the endogeneity associated with the viewing choice. One could argue that a more sophisticated user is more likely to take advantage the release page but she gives better forecasts regardless. In column (3), we include user profession fixed effect and again the result does not change much. Overall, it is clear that viewing public information including the current 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 view as the logarithm of one plus the ratio of the number of forecasts with release views longer than 5s to the number of total forecasts (LnNumView). In other words, if most users viewed the release page before making their forecasts in that release, LnNumView for that release will be higher. Interestingly, when regressing absolute consensus forecast error on LnNumView, we find a significant positive coefficient on LnNumView, suggesting that viewing of public information actually makes the consensus forecast less accurate.

Another way of seeing this result is through a simple horserace as we conduct in Panel C. In each release, we separate all forecasts into two groups. The view group contains all forecasts that are made after viewing the release page. The no-view group contains the remaining forecasts that are made without release view. We then compute two consensus forecasts using forecasts from the two groups respectively and compare which consensus is more accurate. Out of 1331 releases we studied, the no-view consensus wins 62.39% of the time, which is significantly than 50%. Again, viewing of public information makes the consensus forecast less accurate.

How can release view improve the accuracy of individual forecast 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 extreme forecast error, thus 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 extreme case, if all subsequent users completely herd on the first user, then private information of the subsequent users is completely discarded so the crowd consensus is no more accurate than the first forecast in that sequence. In particular, biases in earlier forecasts are more likely to persist subsequently and show up in the final consensus forecast.

Table 4 examines such a persistent bias at the release level. 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 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 upward (downward) 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. The control variable include the same measure of forecast uncertainty, and sector and quarter fixed effects. The results confirm a strong link between the persistence of bias and the release view. When more forecasts are made after release view, the initial bias is more likely to persist and show up in the final consensus.

4.3 The role of "influential" users

So far, the evidence using the unique release view information suggests that the influence we exert on each other can make the crowd collectively dumber, consistent with the prediction of herding. Of course, not all users are created equal. Some users potentially can exert stronger influence on the others. We would expect the herding behavior to be more severe when more "influential" users are present in the crowd.

We identify "influential" users as those (1) who made more forecasts, (2) whose forecasts are more accurate, (3) whose forecasts are more often viewed by other uses, and (4) whose forecasts tend to "lead" subsequent forecasts. To measure 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 this 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 for each user, we count the number of times his/her estimates are identified as leading (among smallest three ratios in 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 less than 20 forecasts are assigned to the lowest value. The users who rank above 80th percentile on the measure are identified as influential users. While neither of the four criteria gives a complete description of a "influential" user. By finding consistent results across all four criteria, we are confident that we are capturing many 'influential" users indeed.

Table 5 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 influenced dummy variable that equals 1 when a large number of influential users have made forecasts. As in Table 2, we find a significant 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 also negative, significant and twice as large in absolute term. In other words, when the current release page contains the forecasts of influential users, viewing this page is associated with even more weight on the consensus and less weight on the private signal. Simply put, users are herding more with influential users.

Table 6 examines how influential users affect the consensus accuracy at the release level. We find a significant positive coefficient on the interaction term between LnNumViews and the influenced dummy, which means that more viewing of the influential users' forecasts is associated with less accurate consensus forecast. As subsequent users are herding more with influential users, even more private information is discarded and the accuracy of the consensus forecast is further reduced.

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 7, we then regress the consensus forecast revisions in the later period (close-to-announcement periods) on influential users' forecast revisions in the earlier period. Across all four definitions of influential user, 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, consistent with the notion that subsequent users are "herding" with the "influential" users.

Knowing that influential users' forecasts strongly swing subsequent forecasts, we conjecture that if influential users' early forecasts are biased, such a bias is likely to drag the consensus to the wrong direction. To identify such a bias ex-ante, we compare the direction of influential users' forecast revisions against the sign of contemporaneous stock return. In Panel B, we find that when contemporaneous stock return is negative and influential users issue forecasts that are lower than the current consensus, final consensus is more accurate, consistent with the notion that influential users facilitate the incorporation of negative information. On the other hand, when contemporaneous stock return is negative and influential users issue forecasts that are higher than the current consensus nevertheless , final consensus becomes less accurate. In this case, influential users' forecasts likely reflect positive sentiment that are propagated to subsequent users and drag the consensus to the wrong direction.

Does the market fully understand such a predictable bias? The answer seems to be No as we examine the earnings-announcement window return in Table 8. When the positive sentiment in influential users drags the the final consensus too high, the market is negatively surprised on the earnings announcement, as evident in a significant lower cumulative abnormal return.

5 Blind Experiment

Our empirical tests so far are affected by the endogeneity associated with the viewing choice. One could argue that a user may choose to view the release page only when he has little private information. However, as long as these forecasts are not biased, their average should still converge to the true earnings even though each forecast contains weak signal. The fact that the consensus forecasts aggregated from forecasts with release views is less accurate must in part reflect the influence of prior forecasts.

Of course, when the sample size is small, the average of less informative signals may converge to the truth at a slower pace than the average of more informative signals. Therefore, it is still possible that forecasts with release views may result in a less accurate consensus due to a small sample. In order to address the endogeneity concerning the information acquisition choice nevertheless, we collaborate with Estimize.com to run randomized experiments during the second and third quarters of 2015.

The stocks in our experiments are randomly selected to come from a wide range of industries. For each selected stock, we then randomly pick a set of users to participate in the experiment. When the user is selected, she will be asked to make a earnings forecast when the release page is disabled. The resulting forecast is labelled as the blind forecast (f_b). Each blind estimate is matched with the closest estimate in the sequence made by a different user who can view the release page. The matched estimate is labelled as 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. There are 103 releases in the final sample, 13 from the first round pilot experiment and the remaining 90 from the second round experiment.

We first compare the blind forecasts to their matching default forecasts in terms of information weighting. Similar to panel A of Table 2, we regress forecast errors (FE) on Dev and its interaction with the default forecast dummy ($Default$). The results are reported in Table 9. The regression in Column (1) does not include sector fixed effects. First, the large, positive, and significant coefficient on Dev (0.735) confirms that blind forecasts are made almost exclusively with private information. The coefficient is higher than that (0.645) 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 -0.276 on $Dev \times Default$ indicates the ability to view public information results in less overweighing 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 sector fixed effect in Column (2) is very similar.

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 consensus 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 turned 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 actually 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). We then 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 enabled so the user can view the public information and choose to revise her forecast and the new forecast is labelled as the revised forecast (f_r). Users can of course choose not to change her forecast, in which case, 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 and the release page is enabled.[15] In this case, we could interpret f_b as the private signal and f_r as the forecast combining both public and private signals: $f_r = kf_b + (1-k)c$, or

$$f_b - f_r = (1-k)(f_b - c).$$

In other words, if we regress $f_b - f_r$ on $f_b - c$, the slope coefficient identifies the weight the users place on public information. When we run the regression in Panel A of Table 10, we find a coefficient of about 0.534 (column 2), suggesting that estimize user on average place more weight on public information then on her private information.

In Panel B, we compare the accuracy of two consensus forecasts: (1) “blind” consensus computed using all “blind” forecasts; and (2) “revised” consensus computed using all revised forecasts when release view is enabled. Out of the 13 randomly selected releases in the pilot experiment, the “blind” consensus significantly outperform 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 crowd can be better harnessed by encouraging independent voice from the participants. Motivated by our findings, Estimize.com decided to switch to the “blind” forecast platform since November 2015 where forecasts from all 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.”

[15] 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 often lags f_b significantly. Since new information may have arrived during that gap, f_r became less comparable to f_b in the second experiment.

6 Conclusion

The wisdom of crowd hinges on having 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, herding behavior arises, resulting in the loss of useful private information.

Take 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 the more public information viewed, the more a user will underweigh 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, consistent with herding. We also find such a herding behavior to become more severe if the public information set contains estimates of the "influential" users.

Finally, a randomized experiment offers clean evidence that wisdom of crowd can be better harnessed by encouraging independent voice from the participants. Ironically, by limiting the crowd's information access, we can actually improve the accuracy of their average forecast. We are confident that by adopting such a "blind" forecast platform, Estimize.com will generate more accurate corporate earnings forecasts that are crucial for the efficiency and well-functioning of the financial market.

References

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.

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.

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

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

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

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 Wolfe. The value of crowdsourced earnings forecasts. *Working Paper*, 2015.

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.

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

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 A: Proofs

Proof of Lemma 1 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:

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

Proof of Proposition 2 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 $w \in (0, 1]$.

Proof of Proposition 3 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,

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

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 sample covers releases from March 2012 to July 2014 with at least 10 estimates. The left axis represents the number of users and releases, and the right axis represents the number of estimates.

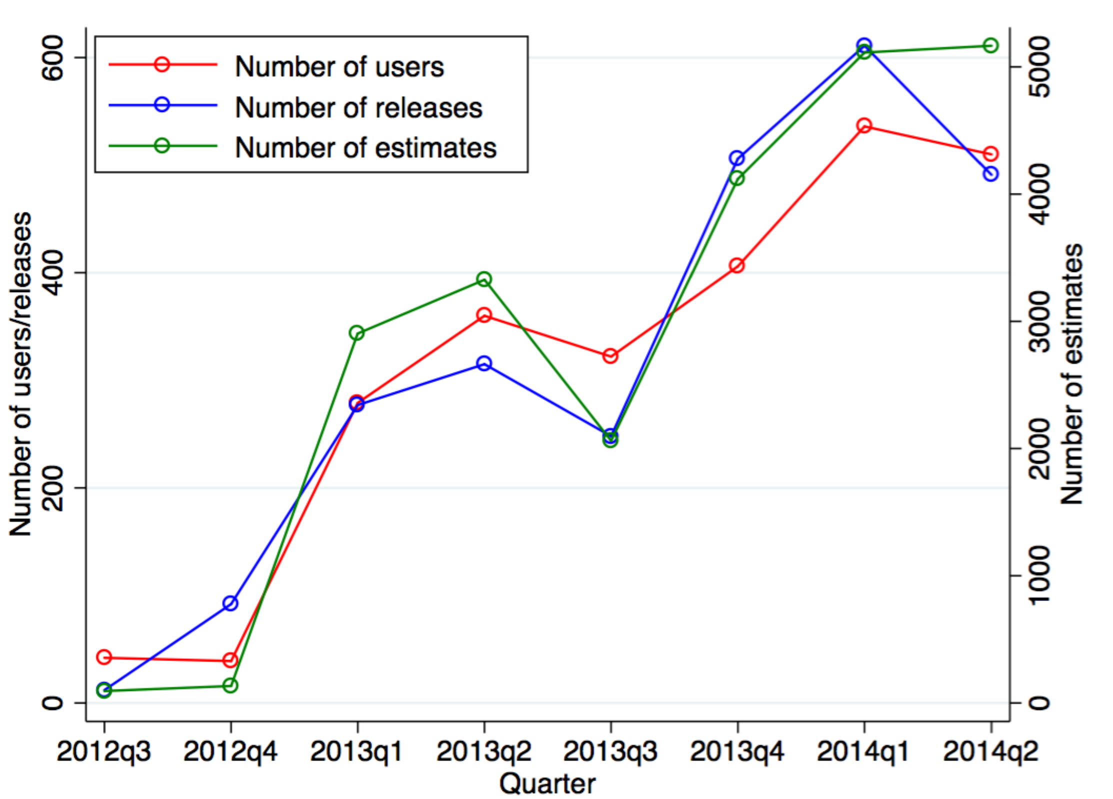

Figure 2: **Example of a Release Page**

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

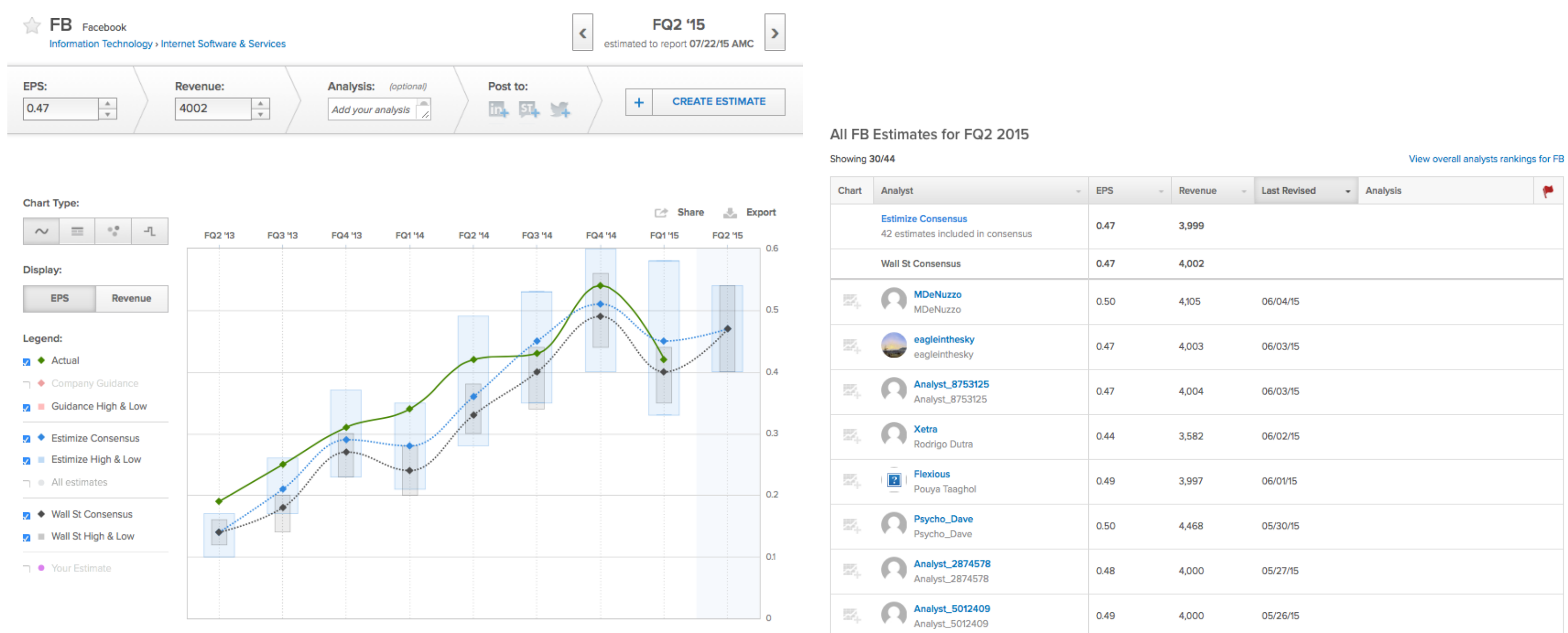

Table 1: **: Descriptive statistics for Estimize sample**

The table presents descriptive statistics for forecasts submitted on Estimize from March 2012 to July 2014. The sample covers 1317 releases with at least 10 estimates. Panel A reports release-level forecast characteristics. The “Runs test p-value” is the p value of a runs test of the hypothesis that the EPS forecasts occur in a random order by counting how many runs there are above and below the consensus. Panel B reports release-level financial characteristics. The sample contains 1266 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 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. Panel C reports user-level characteristics. Panel D reports the sector distribution of the 471 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 = 1317)							
Number of forecasts	19.56	13.17	10.00	11.00	15.00	22.00	70.00
Number of distinct users	16.53	10.67	4.00	10.00	13.00	19.00	56.00
Number of release views	20.03	25.29	1.00	8.00	12.00	22.00	135.00
Runs test p-value	0.41	0.31	0.00	0.12	0.38	0.67	1.00
Consensus error (=consensus-actual)	-0.02	0.15	-0.48	-0.05	-0.01	0.02	0.30
Abs (consensus error)	0.07	0.13	0.00	0.01	0.03	0.07	0.60
Estimize abserr - WS abserr	-0.01	0.05	-0.22	-0.02	-0.01	0.01	0.15
Panel B: Release-level Financial Characteristics (#Obs = 1266)							
Size (in million)	29715.31	52798.17	461.73	3295.56	9597.14	31156.45	241996.03
B/M	0.41	0.41	0.03	0.19	0.31	0.49	1.88
Size group (=1: bottom 20%; =5, top 20%)	4.11	1.08	1.00	3.00	4.00	5.00	5.00
B/M group (=1: bottom 20%; =5: top 20%)	1.90	1.20	1.00	1.00	1.00	2.00	5.00
Panel C: User-level Characeristics (#Obs = 1824)							
Number of tickers covered	9.01	28.92	1.00	1.00	2.00	4.00	170.00
Number of forecasts submited	14.21	55.76	1.00	1.00	2.00	6.00	307.00

Panel D: Distribution of Stocks by Sector

Sector	Freq	Pct
Information Technology	144	30.57
Consumer Discretionary	126	26.75
Industrials	47	9.98
Health Care	46	9.77
Consumer Staples	39	8.28
Others	69	14.65
Total	471	100

Panel E: Distribution of Users by Profession

	Freq	Pct
Financial Professionals:		
Buy Side	210	11.84
Sell Side	114	6.43
Independent	288	16.24
Non Professionals:		
Information Technology	399	22.50
Student	300	16.92
Financials	103	5.81
Consumer Discretionary	90	5.08
Health Care	74	4.17
Others	195	11.00
Total	1773	100.00

Table 2: **: Release view and weighting of information**

Panel A presents results of forecast-level weighting regression. The dependent variable is forecast error, which is defined as the difference between user's forecasted EPS and the actual EPS. The main independent variables include: (1) Dev: the forecast distance from the consensus prior to the submitted forecast; (2) Nonzero views: a dummy variable for viewing the release page for longer than 5s at least once; (3) the interaction term between Dev and Nonzero views. Panel B presents results of release-level regression. The dependent variable is the p-value of a runs test of the hypothesis that the EPS forecasts occur in a random order by counting how many runs there are above and below the consensus. The main independent variable is the logrithm of one plus the ratio of the number of forecasts with release views longer than 5s to the number of total forecasts. Standard errors are in parentheses and double-clustered by sector and quarter. ***, **, * - significant at the 1, 5, and 10% level.

	(1)	(2)
Panel A: Forecast-level analysis		
Dependent variable:	Forecast error (= Forecast-Actual)	
Dev	0.645***	0.686***
(= Forecast - Pre consensus)	(0.067)	(0.047)
Dev X Nonzero Views	-0.123**	-0.162**
	(0.054)	(0.065)
Nonzero Views	-0.000103	-0.000547
	(0.001)	(0.004)
Profession effect	No	Yes
Observations	20741	20741
R-squared	0.063	0.102

Panel B: Release-level analysis

	(1)	(2)
Dependent variable:	Runs test p-value (small p-value indicates forecasts are more correlated)	
LnNumView	-0.0471*	-0.0818***
(= ln(1+Num of Release View/Num of Forecasts)	(0.028)	(0.019)
Sector effect	No	Yes
Quarter effect	No	Yes
Observations	1331	1331
R-squared	0.012	0.032

Table 3: **: Release view and forecast accuracy**

Panel A presents results of forecast-level regression. The dependent variable is the absolute value of forecast error. Forecast error is defined as the difference between user's forecasted EPS and the actual EPS. The main independent variable is "Nonzero Views", a dummy variable for viewing the release page for longer than 5s at least once. The control variable include the standard deviation of forecast error normalized by the absolute value of median forecast error, and users' profession fixed effects. Panel B presents results of release-level regression. The dependent variable is the absolute value of forecast error. Forecast error is defined as the difference between user's forecasted EPS and the actual EPS. The main independent variable is the logarithm of one plus the ratio of the number of forecasts with release views longer than 5s to the number of total forecasts. The control variable 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 results of a comparison of the consensus of forecasts with release views versus the consensus of forecasts without release views within each release.

	(1)	(2)	(3)
Panel A: Forecast-level analysis			
Dependent variable:	Abs(FE = Forecast - Actual)		
Nonzero Views	-0.00318**	-0.00347***	-0.00331***
	(0.001)	(0.001)	(0.001)
Std Dev (FE) / Abs (Median(FE))		0.0400**	0.0398**
		(0.016)	(0.016)
Profession Effect	No	No	Yes
Observations	22022	21925	21844
R-squared	0.000	0.007	0.007
Panel B: Release-level analysis			
Dependent variable:	Abs(FE = Forecast - Actual)		
LnNumView	0.0295**	0.0262*	0.0546***
(= ln(1+Num of Release View/Num of Forecasts)	(0.013)	(0.013)	(0.007)
Std Dev (FE) / Abs (Median(FE))		0.0168**	0.0103**
		(0.008)	(0.004)
Sector effect	No	No	Yes
Quarter effect	No	No	Yes
Cluster	Sec and Quar	Sec and Quar	Sec and Quar
Observations	1318	1312	1312
R-squared	0.005	0.011	0.100

Panel C: Within-release horse race

	Freq. of Releases	Percentage
Average of forecasts with release views wins	501	37.61%
Average of forecasts without release views wins	831	62.39%
Binomial test p-value:	<0.0001	

Table 4: **: Release view and lead-lag biases**

The table presents results of forecast-level regression. 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 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 upward (downward) 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 the logrithm of one plus the ratio of the number of forecasts with release views longer than 5s to the number of total forecasts within the close-to-announcement window. The control variable 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:	Early and close-to-announcement estimates are biased in the same direction		
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	0.429***	0.519***	0.482***
(= ln(1+Num of Release View/Num of Forecasts)	(0.153)	(0.141)	(0.144)
Sector effect	Yes	Yes	Yes
Quarter effect	Yes	Yes	Yes
Observations	1178	1178	1180
Pseudo R2	0.0288	0.0352	0.0460

Table 5: **: The impact of influential users on the weighting of information**

The table presents results of forecast-level weighting regression. The dependent variable is forecast error, which is defined as the difference between user's forecasted EPS and the actual EPS. The main independent variables include: (1) Dev: the forecast distance from the consensus prior to the submitted forecast; (2) Nonzero views: a dummy variable for viewing the release page for longer than 5s at least once; (3) Influenced: a dummy variable which is equal to one if the number of influential users ahead of the user to which the observation corresponds is above the 80th percentile across all observations; and the interaction terms among these three variables. To identify the influential users, we consider four measures: (1) number of releases; (2) accuracy; (3) number of releases being viewed; (4) prob of being a leader. The measures for users who submit less 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	Number of releases	Accuracy	Number of releases being viewed	Prob of being leader
Dev	0.592***	0.603***	0.589***	0.608***
(= Forecast - pre consensus)	(0.050)	(0.030)	(0.043)	(0.036)
Dev X Nonzero Views	-0.0879***	-0.0986***	-0.0920***	-0.100***
	(0.026)	(0.019)	(0.011)	(0.020)
Dev X Nonzero Views X Influenced	-0.271**	-0.216***	-0.279**	-0.206***
	(0.108)	(0.055)	(0.113)	(0.045)
Dev X Influenced	0.368**	0.312**	0.397**	0.292*
	(0.174)	(0.142)	(0.166)	(0.160)
Nonzero Views X Influenced	-0.00882*	-0.00893	-0.00622	-0.00944
	(0.005)	(0.005)	(0.005)	(0.006)
Influenced	-0.0187*	-0.0176**	-0.0202**	-0.0176**
	(0.010)	(0.008)	(0.010)	(0.008)
Nonzero Views	0.00282	0.00259	0.00226	0.00263
(Nonzero release views longer than 5s)	(0.003)	(0.003)	(0.003)	(0.003)
Profession effect	Yes	Yes	Yes	Yes
Observations	20741	20741	20741	20741
R-squared	0.070	0.069	0.070	0.069

Table 6: **: The impact of influential users on the consensus during the close-to-announcement period**

The table presents results of release-level regression. The dependent variable is the absolute consensus error during the close-to-announcement period. The consensus error is defined as the difference between the mean of user's forecasted EPS and the actual EPS. The main independent variables include: (1) LnNumView: the logarithm of one plus the ratio of the number of forecasts with release views longer than 5s to the number of total forecasts within the close-to-announcement window; (2) Influenced: a dummy variable which is equal to one if the number of influential users in the early period is above the 80th percentile across all releases; and (3) the interaction term between these two variables. 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)
Dependent variable:	Abs consensus error (= abs(Consensus - Actual))			
Measure of Influential Users	Number of releases	Accuracy	Number of releases being viewed	Prob of being leader
LnNumView X Influenced	0.695**	0.614*	0.606*	0.612*
	(0.338)	(0.333)	(0.338)	(0.334)
Influenced	-0.349	-0.407	-0.357	-0.416*
	(0.257)	(0.250)	(0.248)	(0.251)
Sector effect	Yes	Yes	Yes	Yes
Quarter effect	Yes	Yes	Yes	Yes
Observations	1179	1179	1179	1179
R-squared	0.0401	0.0382	0.0383	0.0382

Table 7: **: Predicting the change in consensus error and change in consensus accuracy from the early to the close-to-announcement period**

The The table presents results of release-level regression. 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; (2) whether the cumulative abnormal returns (CAR) on the corresponding day of the forecast is 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 variable, 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	Number of releases	Accuracy	Number of releases being viewed	Prob of being leader
ln(1+Num of Upward revision, Neg CAR)	0.014***	0.015***	0.014***	0.014***
	(0.002)	(0.002)	(0.002)	(0.002)
ln(1+Num of Upward revision, Pos CAR)	0.017***	0.017***	0.017***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)
ln(1+Num of Downward revision, Pos CAR)	-0.009***	-0.010***	-0.009***	-0.010***
	(0.002)	(0.002)	(0.003)	(0.002)
ln(1+Num of Downward revision, Neg CAR)	-0.014***	-0.014***	-0.015***	-0.014***
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	-0.015***	-0.014***	-0.013***	-0.013***
	(0.003)	(0.003)	(0.003)	(0.003)
Observations	1199	1199	1199	1199
R-squared	0.140	0.139	0.131	0.137

Panel B: Predicting the change in accuracy				
Dependent variable:	Change in Abs(consensus error)			
ln(1+Num of Upward revision, Neg CAR)	0.005***	0.005**	0.005**	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)
ln(1+Num of Upward revision, Pos CAR)	0.002	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
ln(1+Num of Downward revision, Pos CAR)	-0.002	-0.002	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
ln(1+Num of Downward revision, Neg CAR)	-0.005**	-0.005***	-0.005**	-0.005***
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	-0.014***	-0.013***	-0.013***	-0.013***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	1199	1199	1199	1199
R-squared	0.014	0.014	0.013	0.014

Table 8: **: Predicting the cumulative abnormal return around announcement date**

The table presents the results of regressing the cumulative abnormal return 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; (2) whether the cumulative abnormal returns (CAR) on the corresponding day of the forecast is 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	Number of releases	Accuracy	Number of releases being viewed	Prob of being leader
ln(1+Num of UD)	-0.008**	-0.009**	-0.008*	-0.009**
	(0.004)	(0.004)	(0.004)	(0.004)
ln(1+Num of UU)	0.005	0.004	0.002	0.005
	(0.004)	(0.004)	(0.004)	(0.004)
ln(1+Num of DU)	-0.002	-0.002	0.001	-0.002
	(0.004)	(0.004)	(0.004)	(0.004)
ln(1+Num of DD)	-0.001	-0.001	-0.002	-0.000
	(0.004)	(0.004)	(0.004)	(0.004)
Time effect (Year-month)	Yes	Yes	Yes	Yes
Observations	1097	1097	1097	1097
R-squared	0.027	0.027	0.025	0.027

Table 9: **: Blind experiment: Blind vs. Default**

When a user is randomly selected to participate in the experiment, she will be asked to make a earnings forecast when the release page is disabled. The resulting forecast is labelled as the blind forecast (f_b). Each blind estimate is matched with the closest estimate in the sequence made by a different user who can view the release page. The matched estimate is labelled as 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. 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 equal to one if it is a default forecast; 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.

	(1)	(2)
Dependent variable:	Forecast error (= Forecast-Actual)	
Dev	0.735***	0.715***
(= Forecast - Pre consensus)	(0.133)	(0.105)
Dev X Default	-0.276**	-0.181*
	(0.134)	(0.105)
Default	-0.00388	-0.00351*
	(0.002)	(0.002)
Sector effect	No	Yes
Observations	20741	20741
R-squared	0.063	0.102

Table 10: **: 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 regression. 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 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 results of within-release horserace between the blind consensus and the revised consensus. When calculating 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