The Value of Diversity: Analyst Heterogeneity and Individual Analysts' Forecast Accuracy *

LING CEN

Chinese University of Hong Kong

YUK YING CHANG

Massey University

SUDIPTO DASGUPTA

Chinese University of Hong Kong, CEPR and ABFER

Preliminary: Please do not circulate

Previous version: November 2018

This version: February 2019

* We acknowledge the Massey Business School for providing financial support.

The Value of Diversity: Analyst Heterogeneity and Individual Analysts' Forecast Accuracy

Based on US data over the 1982 to 2014 period, we examine how the analyst heterogeneity affects the accuracy of the consensus (mean and median) and individual forecasts of the analysts following a firm. We document that the mean and median forecasts have smaller errors when these analysts are more diverse in terms of experience and location. However, we find no association between analyst diversity and consensus forecast bias. Our results are consistent with the notion that greater diversity is associated with more complete "averaging out" of idiosyncratic components of individual forecasts, similar to the portfolio diversification effect, as more diverse individuals bring different perspectives and interpretations to the public domain via their forecasts. Moreover, we find that the individual forecasts of these more dissimilar analysts have smaller errors as well. This suggests that analysts incorporate other analysts' information into their individual forecasts – a phenomenon that previous literature has not been able to document clearly because it is difficult to disentangle the influence of other analysts' forecasts from the effect of common shocks. By contrast, our results suggest that individual forecast errors decrease with diversity (due to the "averaging" of other analysts' forecast errors) only when analysts assign some weight to forecasts of other analysts in making their own forecasts.

Over the last forty years, research on the effects of diversity on group decision making has become increasingly important in a variety of disciplines (e.g., Erhardt et al. (2003) in multi-disciplinary corporate governance; Ashraf & Galor (2013) in economics; Pelled et al. (1999) in organisational science; Van Knippenberg and Schippers (2007) in psychology; Wise and Tschirhart (2000) for a review). There is considerable debate about whether group diversity enhances group performance (e.g., Van Dijk et al., 2012). Available evidence suggests that the type of diversity matters. *Visible demographic characteristics* (observable attributes which individuals are born into, such as race/ethnicity, gender, nationality) are generally found to affect group performance adversely; whereas *informational demographic characteristics* (such as work experience and education, that are typically related to skillsets individuals acquire and employ when undertaking a task) are deemed to have a positive impact (Thatcher 1999). Studies have shown that diversity along visible demographic characteristics increases relationship conflict and reduces communication and commitment within teams (Byrne 1971; Pelled 1996; Williams and O'Reilly 1998). In contrast, diversity on informational demographic characteristics brings a wider array of opinions, experiences and information that can improve group performance (e.g., Gruenfeld et al. 1996).

Along with the impact on group performance, the effect of diversity on individual performance is also of interest and has been explored, albeit to a more limited extent (Hansen et al. 2006). In many workplace situations, group output or performance is more readily observed than individual-specific performance. The two are often distinct – for example, a typical experimental setting is one in which students with diverse backgrounds are randomly assigned into groups, and are evaluated on a group task and individual tasks. However, outside of such experimental environments, it may not be easy to observe the two types of outcomes separately. For example, a team of innovators working together succeed or fail together. A corporate board's performance (measured, for example, in terms of the company's stock price performance) is only measurable at the group level. In these situations, one can examine the effect of diversity on group performance, but it is difficult to assess its impact on individual performance, *which could extend beyond the specific group task*. For example, individuals exposed to diverse approaches or ideas to a problem could benefit and perform better in another task. Put differently, a specific group task performed under a diverse environment could suffer when diversity is based on *visible demographic characteristics*, but individuals could benefit when it is based on *informational demographic characteristics*.

For these reasons, the effect of a particular type/dimension of diversity on individual performance is generally harder to assess than on group performance, aside from classroom-based studies. There is limited evidence on how diversity affects individual performance in professional or workplace settings (e.g., Richter et al. 2012). In this paper, we examine a unique setting – earnings forecast performance by analysts following a firm simultaneously. This setting is unique for several reasons. First, analysts do have common tasks. However, there is no collective "task" that the analysts have to do, even though the market pays close attention to a collective output – namely, the "consensus" (mean or median) forecast. Clearly, individual analysts are not evaluated on the basis of the consensus, but rather, their own forecasts. Second, the interaction is seemingly "arms length" – while we generally do not know whether analysts directly communicate with each other, obviously they are not required to do so. Third, since analyst forecasts are in the public domain, each analyst is able to observe other analysts' prior forecasts for the same firm-year and revise his forecast. Overall, this is a unique setting in which interaction is "arms length" so the interpersonal disruptive influences of visible demographic characteristics are likely to be absent. However, analysts are exposed to the "views" (estimates) of other analysts and the *informational demographic characteristics* may then become relevant.

One metric in terms of which both group and individual performance are measured for analysts is the accuracy of the forecast. This is usually defined as the forecast error, or the deviation of the forecast from the actual earnings, scaled by the last price of the previous year. The mean absolute deviation (MAD) could be larger if either (a) forecasts are more biased (for example, analysts may generate more optimistic forecasts to be in the good books of management (Hong and Kubik 2003)), or (b) unbiased forecast errors are larger (e.g., because the information environment is more opaque).

Through either or both these channels, a more diverse group of analysts can affect both group performance (the MAD of the consensus forecast) as well as individual performance (MAD of the individual analyst forecasts). In a recent paper, Merkley et al. (2017), compiling detailed data on analysts' cultural backgrounds, find that greater cultural diversity reduces the absolute forecast error of the consensus forecast. They attribute this improvement to greater information sharing among analysts with diverse cultural backgrounds, and suggest earnings conference calls as a possible mechanism. They also find that the average bias of analyst forecasts decreases as diversity increases. Note that even though individuals are "born into" their cultures and normally one would consider culture to be a *visible demographic variable*, the usual arguments as to why greater diversity in visible demographic variables diminishes performance could be absent here since there is no interpersonal interaction. The results in Merkley et al. (2017) are based on the premise that people from different cultural backgrounds have independent ways of looking at issues, and also suggest that the interpersonal elements possibly explain why previous studies in group settings do not find culture diversity to enhance performance.

In contrast to MerkIey et al. (2017), in this paper, we focus on individual analyst forecasts. We directly examine whether greater diversity affects accuracy of individual analyst forecasts because an analyst pays attention to other analysts' forecasts in coming up with his own forecast, and find supportive evidence. Studies show that analyst forecasts are influenced by their peers' (Hong et al. 2000; Clement and Tse 2005; Jegadeesh and Kim 2010). This can happen when analysts react one after another to new common or public information as Clement et al. (2010) document. Alternatively, the peer influence can be real if the analyst takes into consideration of the forecasts of the other analysts because these forecasts are information relevant. However, these two channels are difficult to disentangle (Welch 2000).[1] Our paper fills this gap in the literature.

Our measures of diversity are the *Herfindahl-Hirschman indices* (HHI) (Herfindahl 1950 or Hirschman 1964) based on two variables – analyst location and analyst experience. The HHI is widely used in diversity research ((e.g., Lang & Stulz 1994; Denis et al. 1997)) and inversely related to diversity. We also conduct similar tests and add a new test on the analysis of the consensus forecast accuracy as Merkley et al. (2017) in order to contrast our results from theirs and to the distinguish the channel through which diversity affects forecast error in our setting.

We first examine the effect of diversity on the MAD of the consensus forecast (as in Merkely et al. (2017)) and find a negative effect. However, unlike their findings, we find no evidence that the average forecast bias decreases as diversity increases for our diversity variables. Rather, the reason why greater diversity increases group performance (i.e., lowers the MAD of the consensus forecast) is essentially similar to a portfolio diversification effect. If each analyst's forecast represents the fundamental (true earnings), plus an analyst-specific bias, plus a random signal, then the consensus forecast simply averages out the signals. If greater diversity implies that the signals are less correlated, diversity would mechanically

[1] See Cohn and Juergens (2014) for a new approach to determine whether analysts influence each other.

improve the MAD of the consensus forecast, even when the average forecast bias did not change.

In contrast, there is no reason why analysts' *individual* forecast errors should mechanically improve with diversity, unless the analysts pay attention to other analysts' forecasts. Suppose, for example, that an individual analyst's last forecast during the forecasting period is a linear function of his own estimate (fundamental plus bias plus noise) and the average of all other analysts' last forecasts. Then greater diversity increases the average analyst's forecast accuracy if (a) the analyst assigns a positive weight on the average of other analysts' forecasts, and (b) the standard deviation of the average forecast of all other analysts decreases in diversity. The first condition is precisely the idea that an informational demographic measure of diversity presents an array of different views that could potentially get incorporated into individual decision making, while the latter should be the case if greater diversity in analysts' location or experience corresponds to individual analyst errors being less correlated with each other.

We then examine the forecast accuracy of the last forecast of any analyst following the firm in a given year and the last individual forecasts of all analysts. Our tests are done in levels including year, firm and analyst fixed effects, in first differences with year fixed effects, and in levels controlling for the fixed effects plus the accuracy of the analyst's first forecast or the accuracy of the consensus forecast based on each analyst's first forecast. For these individual-level tests, we have two additional diversity measures. First, we exclude the group to which the analyst in question belongs and construct HHIs based on the remaining groups. Second, for each analyst, we construct a new variable inversely related to diversity – the fraction of analysts covering the firm that share the same demographics (e.g., location, experience) as that analyst. We find that for all the tests, all our diversity measures are positively related to forecast accuracy.

Endogeneity is a major challenge for our tests. Our tests control for year and firm fixed effects, so "within-variation" drives our results. However, change in diversity need not be exogenous – in particular, it is possible that factors that drive changes in the diversity of the group of analysts following a firm also drive the accuracy of analyst forecasts. Most of the endogenous changes in the diversity, however, are likely to happen via a change in the number of analysts following the firm. For example, if the information environment of the firm improves for exogenous reasons, this could encourage new analysts to cover the firm, resulting in an increase in the diversity of the analyst pool at the same time. Indeed, in support of such a channel affecting our results, we find that there is a strong negative contemporaneous correlation between changes in our HHI measures and the change in coverage (i.e. the number of analysts following the firm). While we control for lagged coverage in our fixed effects regressions, to control for normal or mechanical change in diversity related to a change in coverage, we run our regressions in first-differences. After controlling for the change in contemporaneous coverage, the change in diversity thus picks up random variation in diversity as coverage changes.[2] We find that in all these difference regressions, our diversity results still hold.

In addition, we provide another approach to addressing the reverse causality concern that changes in firm environment simultaneously improve forecast accuracy and increase analyst coverage (and hence diversity). First, as noted above, for both of our diversity measures, we find no evidence that greater diversity reduces forecast bias. This suggests that diversity reduces the mean absolute deviation mainly through the "averaging" effect, and not because forecasts are based on better information. To further reduce the likelihood that both diversity and forecast accuracy reflect an improvement in the firm's information environment,

[2]Controlling for the change in contemporaneous coverage captures "mechanical" variation in diversity as coverage changes due to factors such as an improvement in the information environment. Any residual change in diversity thus can be regarded as unrelated to such factors.

we control for the absolute error of the consensus forecast based on the *first* forecast of every analyst who covers the firm during the forecast period. The dependent variable (absolute error of the consensus based on the *last* forecast of every analyst) and this control variable are based on the forecasts of the same set of analysts. Therefore, the incremental effect of the HHI on the former cannot be due to factors that cause firm environment to change, and simultaneously change analyst coverage and diversity. We find that after controlling for the first forecast errors, our diversity results remain.

The rest of the paper is organized as follows. Section 1 describes our data and introduces variables. Section 2 explains the relation between diversity and mean absolute deviation. Section 3 reports and discusses our empirical results. Section 4 concludes.

1 Data and variables

We obtain realized earnings and analyst earnings forecasts from the I/B/E/S Detail History files. The sample covers the 1982 to 2014 period. We primarily consider the last annual earnings per share (EPS) forecast of each analyst for each firm year. We use the I/B/E/S actual earnings instead of the Compustat earnings because the I/B/E/S has a policy of reporting actual earnings numbers that are consistent with forecasts, i.e., it excludes the same items from actual EPS numbers that the majority of analysts exclude from their forecasts (Christensen 2007). The sources of accounting and financial data are Compustat and CRSP, respectively.

The main dependent variable is forecast error. For each firm year, we first calculate the difference between the mean or median of the earnings forecasts of all analysts and the corresponding actual earnings being forecasted, scaled by the share price as of the end of the previous fiscal year. To obtain our primary dependent variable (*ABS ERROR*), we take the absolute value of the difference calculated. For the analyses of individual forecasts of the analysts, we similarly take the absolute difference between the individual forecast value of each analyst and the actual value being forecasted, scaled by the prior share price. The other dependent variable is forecast bias. It is calculated in the same manner as *ABS ERROR*, except that we do not take the absolute value of the difference.

The key explanatory variables are location- and experience- diversity of analysts, which are constructed for each firm year. Our diversity measure is the Hirschman-Herfindahl index (Herfindahl 1950 or Hirschman 1964) (HHI), mathematically equivalent to the Simpson's (1949) index that has long been considered the primary measure of diversity (McDonald and Dimmick 2003). The HHI is also akin to the effective number parties index (Laakso and Taagepera 1979) in politics and the inverse participation ratio (e.g., Kramer and MacKinnon 1993) in physics. Palan (2010) compares nine common specialisation and heterogeneity indices and mentions that the HHI meets all criteria of a sound index. The HHI is not only widely used in academic research (e.g., Lang & Stulz 1994; Denis et al. 1997), but also employed by the Department of Justice and the Federal Reserve (Rhoades 1993). The constructions of our HHIs are as follows:

1.1 Construction of location diversity:

For each firm year, we pool all analysts following the same firm in the same year together and then divide them into 9 location groups based on the 8 Bureau of Economic Analysis (BEA) regions and non-US countries in which they are located. All non-US countries are combined into one non-US group. Our analyst location data are based on annual volumes of Nelson's Directory of Investment Research.[3] We then calculate the BEA-Herfindahl-Hirschman Index for each firm year ($BEA\_HHI_{j,t}$) as follows:

[3] We acknowledge Kee-Hong Bae and Hongping Tan for providing us with location data for the period 1995 – 2010. The procedure used to identify analysts' locations is the same as that used in Bae, Stulz and Tan (2008) and Bae, Tan and Welker (2008). Using Nelson's Directories, we manually check cases with the same analyst, the

$$BEA\_HHI_{j,t} = \sum_{q=1}^{9}\left(\frac{Number\ of\ analysts\ in\ location\ q_{j,t}}{Total\ number\ of\ analysts_{j,t}}\right)^2 \quad (1.1)$$

where the subscripts *j* and *t* index firm and year, respectively. $Total\ number\ of\ analysts_{j,t}$ is the total number of analysts producing forecasts for firm *j* in year *t*. *Number of analysts in location* $q_{j,t}$ is the number of analysts, producing forecasts for firm *j* in year *t*, in the q^{th} location. The largest possible value of the *BEA_HHI* is 1.0, when all analysts are based in the same region. The smallest possible value of the *BEA_HHI* is 1/9 or 0.11, when each of the 9 regions has the same number of analysts. The lower the *BEA_HHI*, the more diverse the geographical location of analysts. Hence, the HHI is an inverse measure of diversity. Our location sample covers the period between 1994 and 2010, for which the location data has good coverage. We also consider two alternative geographical classifications: the 4 Census Bureau-designated regions and the 52 states. The respective HHIs are constructed in the same manner as *BEA_HHI* and are referred to as *CENSUS_HHI* and *STATES_HHI*. All non-US countries are combined into one group for *CENSUS_HHI* whereas each non-US country is treated as a separate group for *STATES_HHI*.

1.2 Construction of experience diversity:

For each year, we pool all analysts of all firms together and then divide all analysts into 5 groups ($k = \{1, 2, \ldots, 5\}$) based on the quintiles of the length of their experience, which we estimate by how long the analyst has been filed in the I/B/E/S database. To mitigate the problem of tenure truncation in the early years for which the original data is available, our experience sample begins in year 1994. We then calculate the experience-Herfindahl-Hirschman Index for each firm year ($EXP\_HHI_{j,t}$) as follows:

same research firm and multiple locations. We exclude observations for which there is insufficient information to clearly identify the location of the analyst.

$$EXP\_HHI_{j,t} = \sum_{k=1}^{5} \left(\frac{Number\ of\ analysts\ in\ group\ k_{j,t}}{Total\ number\ of\ analysts_{j,t}} \right)^2 \tag{1.2}$$

where the subscripts *j* and *t* index firm and year, respectively. $Total\ number\ of\ analysts_{j,t}$ is the total number of analysts producing forecasts for firm *j* in year *t*. *Number of analysts in group* $k_{j,t}$ is the number of analysts, producing forecasts for firm *j* in year *t*, in the k^{th} experience group. The *EXP_HHI* can range from 0.2, when the number of analysts in each of the five groups is the same, to 1.0, when all the analysts fall into one group. The lower the *EXP_HHI*, the more diverse the experience of analysts. As robustness checks, we also estimate experience diversity based on 4 and 3 experience groups.

There are four other variables of interest. They are proportions of analyst's own location and experience groups and the location- and experience-diversity measures of the remaining groups to which the analyst does not belong. They are constructed as follows.

1.3 Construction of OWN REGION'S PROPORTION:

$$OWN\ REGION'S\ PROPORTION_{i,j,t} = \frac{Number\ of\ analysts\ in\ analyst\ i's\ region_{j,t}}{Total\ number\ of\ analysts_{j,t}} \tag{1.3}$$

$OWN\ REGION'S\ PROPORTION_{i,j,t}$ is the proportion of the analysts following firm *j* in year *t* located in analyst i's region. $Total\ number\ of\ analysts_{j,t}$ is the total number of analysts producing forecasts for firm *j* in year *t*. *Number of analysts in analyst i's region*$_{j,t}$ is the number of analysts, producing forecasts for firm *j* in year *t*, in analyst *i*'s region, based on the Bureau of Economic Analysis classification.

1.4 Construction of OWN EXP_GROUP'S PROPORTION:

$$OWN\ EXP\_GROUP'S\ PROPORTION_{i,j,t} = \frac{Number\ of\ analysts\ in\ analyst\ i's\ group_{j,t}}{Total\ number\ of\ analysts_{j,t}} \tag{1.4}$$

$OWN\ EXP\_GROUP'S\ PROPORTION_{i,j,t}$ is the proportion of the analysts following firm *j* in year *t* in analyst *i*'s experience group. $Total\ number\ of\ analysts_{j,t}$ is the total number of analysts producing forecasts for firm *j* in year *t*. *Number of analysts in analyst i's group*$_{j,t}$ is the number of analysts, producing forecasts for firm *j* in year *t*, in analyst *i*'s experience group where the experience group is one of the 5 groups based on the quintiles of the length of experience of all analysts in the same year.

1.5 Construction of OTHER REGIONS' BEA_HHI:

$$OTHER\ REGIONS'\ BEA\_HHI_{i,j,t} = \sum_{q=1}^{8}\left(\frac{Number\ of\ analysts\ in\ region\ q_{j,t}}{Total\ number\ of\ remaining\ analysts_{j,t}}\right)^2 \qquad (1.5)$$

where the q^{th} region is not analyst *i*'s region. $OTHER\ REGIONS'\ BEA\_HHI_{i,j,t}$ is the BEA_HHI of the analysts, covering firm *j* in year *t*, in the regions outside analyst *i*'s region. $Total\ number\ of\ remaining\ analysts_{j,t}$ is the total number of the remaining analysts producing forecasts for firm *j* in year *t*, excluding the number of analysts in analyst *i*'s region. *Number of analysts in location* $q_{j,t}$ is the number of analysts, producing forecasts for firm *j* in year *t*, in the q^{th} region where *q* is one of the 8 remaining regions outside analyst *i*'s region.

1.6 Construction of OTHER GROUPS' EXP_HHI:

$$OTHER\ GROUPS'\ EXP\_HHI_{i,j,t} = \sum_{k=1}^{4}\left(\frac{Number\ of\ analysts\ in\ group\ k_{j,t}}{Total\ number\ of\ remaining\ analysts_{j,t}}\right)^2 \qquad (1.6)$$

where the k^{th} group is not analyst *i*'s experience group. $OTHER\ GROUPS'\ EXP\_HHI_{i,j,t}$ is the EXP_HHI of the remaining 4 experience groups that are not analyst *i*'s group. $Total\ number\ of\ remaining\ analysts_{j,t}$ is the total number of the remaining analysts producing forecasts for firm *j* in year *t*, excluding the number of analysts in analyst *i*'s experience group. *Number of analysts in group* $k_{j,t}$ is the number of analysts, producing forecasts for firm *j* in year *t*, in the k^{th} group based on the quintiles of the length of experience of all analysts in the same year and to which analyst *i* does not belong.

We define the other variables in Appendix A. In all regressions, all variables except count, time and dummy variables are winsorized at 1% and 99% to minimize the influence of outliers and errors in the data, such as recording errors.

2 Diversity and Mean Absolute Deviation

Let us fix a firm and for simplicity, assume that the ratio of earnings to the previous period's price remains unchanged over time. Let x_i denote an individual analyst *i*'s forecast of the earnings-to-price ratio. We assume that

$$x_{it} = f + b_i + \epsilon_{it} \tag{2.1}$$

where *f* denotes the actual earnings-to-price ratio, b_i is an analyst-specific bias, and $\in_{it}$ denotes all other sources of the analyst-specific forecast error, i.e., noise. We assume that $\in_{it}$ is normally distributed with mean 0 and variance σ^2.

The individual forecast error is given by $|x_{it} - f|$, that is, the individual absolute deviation. Einhorm et al. (1977) derives the expression for the mean absolute deviation (*MAD*) for the case of normally distributed errors. This is given by:

$$MAD = \sigma\left[\frac{b_i}{\sigma}\left(2\Phi\left(\frac{b_i}{\sigma}\right) - 1\right) + 2\phi\left(\frac{b_i}{\sigma}\right)\right] \tag{2.2}$$

where Φ and ϕ denote, respectively, the cumulative distribution and density of the standard normal variable.

It can be checked that the *MAD* is increasing in the size of bias (whether it is positive or negative). Figure 1 illustrates for σ^2=1.

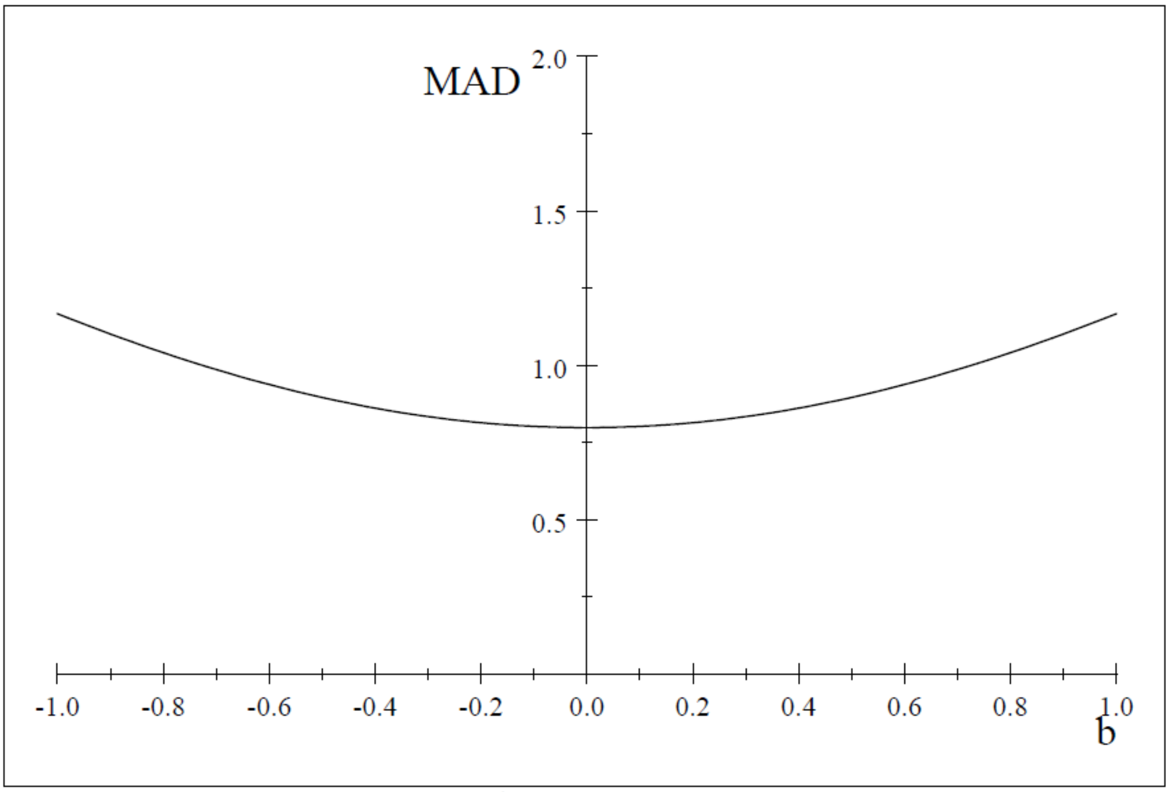

Figure 1. MAD as a function of bias

It is also straightforward to show that

$$\frac{\partial\left(\sigma\left[\frac{b_i}{\sigma}\left(2\Phi\left(\frac{b_i}{\sigma}\right)-1\right)+2\phi\left(\frac{b_i}{\sigma}\right)\right]\right)}{\partial\sigma}=\frac{\sqrt{2}}{\sqrt{\pi}}e^{-\frac{1}{2}\left(\frac{b_i^2}{\sigma^2}\right)}>0.$$

Thus, the *MAD* is increasing in the bias and increasing in the standard deviation of the forecast noise.

Extension to the *MAD* to the consensus is straightforward. The consensus is the mean of *N* forecasts x_{it}, given by $X_t = f + \frac{\sum_{i=1}^{N} b_i}{N} + \frac{\sum_{i=1}^{N} \epsilon_{it}}{N}$. The standard deviation is given by $\sigma_N = \sqrt{\frac{1}{N}[\sigma^2 + (N-1)\sigma_{nm}]}$, where σ_{nm} denotes the covariance between the errors. The expression for the *MAD* of the consensus forecast follows from (2.2) by replacing b_i in (2.2) with the average bias, $\bar{b} = \frac{\sum_{i=1}^{N} b_i}{N}$ and replacing σ with σ_N. If analyst diversity reduces either the average bias or the covariance between the errors, then the effect is to reduce the *MAD* for the consensus forecast.

Finally, turning to the last forecast of each analyst, assume the following setting: each analyst *i* observes the remaining analyst *w*'s forecast from the previous round,

$$x_{wt} = f + b_w + \bar{\in}_{wt}. \quad (2.3)$$

Analyst *i* then observes his new signal $f + \epsilon_{it}$, and all analysts "simultaneously" make their last forecasts for the forecasting period. The variances of ϵ_{it} are assumed to be σ^2, and the covariances $Cov(\bar{\in}_{wt}, \bar{\in}_{it})$ are denoted by σ'_{iw}. For simplicity, we also assume that $Cov(\bar{\in}_{wt}, \in_{it}) = \sigma'_{iw}$. WLOG, assume that analyst *i*, in making the last forecast, assigns a weight (1-α) to the mean forecast of the previous round of all other analysts, and a weight α to his own signal in the last round.[4] Then his forecast in the last round will be

$$x_{it} = f + \alpha b_i + (1-\alpha)\left(\frac{\sum_{w\neq i} b_w}{N-1}\right) + \alpha\epsilon_{it} + (1-\alpha)\frac{\sum_{w\neq i}\bar{\in}_{wt}}{N-1} \quad (2.4)$$

It is straightforward to show that the standard deviation of x_{it} is given by σ_i , where

$$\sigma_i^2 = \left(\alpha^2 + \frac{(1-\alpha)^2}{N-1}\right)\sigma^2 + (1-\alpha)(\alpha + (1-\alpha)\frac{N-2}{N-1})\sigma'_{iw} \quad (2.5)$$

The *MAD* of analyst *i*'s last forecast of the round is given by equation (2.2), with $\sigma = \sigma_i$, and $b_i = \alpha b_i + (1-\alpha)\overline{b_o}$, where $\overline{b_o}$ represents the mean bias of all other analysts. It is increasing in the covariance $\sigma'_{iw}$and the average bias size. If diversity lowers the covariance and/or the average bias, each analyst's forecast accuracy will improve with greater diversity.

3 Empirical results

[4] Without any loss of generality, we could include analyst *i*'s own forecast in the previous round in the mean from the previous round.

3.1 Summary statistics and correlation

Table 1 reports the summary statistics. Concerning analyst diversity, the analysts covering a given firm have generally moderate geographical diversity. To see this, we first estimate the geographical HHI lower bound, the minimum possible value. The lower bound of *BEA_HHI* is 1/the maximum number of regions, i.e., 1/9 or 0.11. The lower bound of *CENSUS_HHI* is 1/5 or 0.2. However, the lower bound of the *STATES_HHI* should be 1/$\textit{Total number of analysts}_{j,t}$ because the maximum number of the possible locations (52+) is far more than the typical number of analysts covering a firm year (*COVERAGE*). Using the mean and median of *COVERAGE* from Table 1, the *STATES_HHI* lower bound is estimated as 0.09 and 0.125. The upper bound of all geographical HHIs is 1.0. As shown in Table 1, means of *BEA_HHI*, *CENSUS_HHI* and *STATES_HHI* are 0.526, 0.558 and 0.462, and their medians are 0.5, 0.5 and 0.401. Therefore, the analysts are generally moderately geographically diverse.

By contrast, in an average year, for a typical firm, the analysts have generally relatively more diverse forecasting experience. *EXP_HHI* has mean and median of 0.344 and 0.288 whereas the lower and upper bounds of the HHI are 0.2 and 1.0, respectively.

Concerning forecast errors, the median and mean of the absolute error of the mean consensus forecasts are 0.291% and 1.163%, respectively. As expected, the median and mean of the *median* consensus absolute errors are smaller of 0.228% and 1.042%, respectively. Meanwhile, the statistics of the absolute errors of the last forecasts of all analysts have a median of 0.210% and mean of 0.889%. As for the corresponding statistics for the very last individual forecast, the median and mean are 0.238% and 1.066%, respectively. The reason why the statistics for individual forecasts are smaller than those for consensus forecasts is that there are larger number of individual forecasts for firm years with larger coverage and forecasts of firms with larger coverage tend to be more accurate.

Our forecast error statistics are comparable to summary statistics reported in the literature. For example, Clement et al. (2011) report a median of the mean consensus absolute errors of 0.2%. Walther et al. (2013) show a mean of the individual forecast absolute errors of 0.95%.

Importantly, the statistics of the difference between the mean of individual forecast absolute errors and the mean consensus absolute error of the same firm year (*MEAN DIFF* or *MEDIAN* DIFF) confirm that the former is not smaller than the latter, i.e., Jensen's Inequality holds. The difference statistics between the median of the individual forecasts and the median consensus forecast are also positive in our sample.

Table 2 reports variance inflation factors (VIFs) and correlation coefficients of the explanatory variables. All VIFs are smaller than 10. Hence, multicollinearity is not an issue. The correlation coefficients between location HHIs are large, with 0.9 between *STATES_HHI* and *BEA_HHI*, 0.85 between *STATES_HHI* and *CENSUS_HHI*, and 0.9 between *BEA_HHI* and *CENSUS_HHI*. By contrast, the correlation coefficients between HHIs of different dimensions are generally low. As for the correlation coefficients between the HHIs and the other explanatory variables, they are also generally small. Two exceptions are -0.49 between *COVERAGE* and *EXP_HHI* and -0.48 between *LNSIZE* and *EXP_HHI*.

3.2 The baseline: location diversity and absolute errors of mean consensus forecasts

Table 3 reports our baseline results. Our dependent variable is the forecast error of the consensus forecast, measured by the absolute difference of the realized earnings and the mean of the last forecasts of all analysts in a given firm year (hereafter referred to as the mean consensus), scaled by the end-of-period stock price of the prior year. The independent variable of interest is location diversity of the analysts in that firm year, as measured by the HHI. Because of an effect akin to a portfolio diversification effect, we expect that more analyst diversity (smaller HHI in (3.1) below) is associated with lower mean consensus absolute error (smaller values of the dependent variable in (3.1) below), following the discussion in Section 2. Specifically, we estimate the following model:

$$Mean\ Consensus\ ABS\ ERROR_{j,t} = a + b \times HHI_{j,t} + c \times Controls + d \times Firm\ FE + e \times Year\ FE + \varepsilon_{j,t} \quad (3.1)$$

where subscript *j* and *t* index firms and years, respectively. $\varepsilon_{j,t}$ is the error term. The variables are described in Section 1 and Appendix A. We include year and firm fixed effects (*Year FE and Firm* FE) in our regressions. Our robust standard errors are based on clustering at the firm level. As discussed above, we expect the HHI coefficients to be positive, i.e., *b*>0.

Consistent with the expectation, the estimated coefficients of all location HHIs are positive and statistically significant at the 1% level. Hence, the more the analyst diversity (the lower the location HHI), the smaller the mean consensus forecast error.

To illustrate the economic significance of the location HHIs, consider a homogeneous group and a heterogeneous group. Both groups have two analysts. The homogeneous group has a HHI of 1.0 whereas the heterogeneous group has a HHI of 0.5. Based on the estimated coefficients in Table 3, compared with the homogeneous group, the heterogeneous group have, on average, 0.289%, 0.261%, and 0.240% (half the HHI estimated coefficients) lower absolute mean consensus errors when the HHI are based on *STATES_HHI*, *BEA_HHI* and *CENSUS_HHI*, respectively, compared with the unconditional median of the consensus absolute errors of 0.395%. These translate into 4.3¢, 3.9¢, and 3.6¢ (6.3¢, 5.7¢, and 5.2¢) per share based on the median (mean) share price. These are economically significant as Bhojraj et al. (2009) suggest that two-cent differences between the actual and forecast earnings can drive significantly different and material managerial actions.[5]

[5] Hotchkiss and Strickland (2003) report that the mean and median two-day cumulative abnormal returns (CARs) are 1.30% and 0.99% when the consensus earnings forecasts are smaller than the actual earnings by more than 2.0¢; the mean and median two-day CARs are -0.83% and -0.52% when the consensus earnings forecasts are larger than the actual earnings.

Concerning the average characteristics of analysts at the firm-year level, we have the following observations. First, the average number of industries covered (*NSIC3*) is associated with the forecasting complexity faced by the analysts on average. Consistent with Clement (1999), we find that when analysts typically follow a larger number of industries, their mean consensus forecasts generally have larger absolute errors. Second, the average number of firms covered by the analysts (*NCO*) can be a proxy for the forecasting expertise of these analysts. We find that consensus forecasts of analysts with more forecasting expertise generally have smaller absolute errors. Third, while the average length of time for which analysts have followed the firm (*LNEXP_WITH_FIRM*) can be a proxy for their experience with that firm, it is also likely to be positively associated with conflicts of interest. If the former is true, then this variable will be negatively related with *ABS ERROR*. If the latter is the case, then this variable will be associated with more optimism and thus may be positively related with *ABS ERROR*. Our results indicate a positive and marginally significant relation between them, possibly reflecting presence of closer relationships and more conflicts of interest when the analysts have, on average, followed the firm for a longer time. Finally, *LNGENERAL_EXP*, the average length of time for which the analysts have made forecasts, shows no significant relation with *ABS ERROR*, which suggests that the average maturity of the analysts may not help accuracy of the their consensus forecasts.

Endogeneity is clearly an important concern in our setting. Next, we discuss several approaches to deal with possible reverse causality and endogeneity concerns. Specifically, we first perform regression analysis in first differences and then run level regressions by controlling for the first forecast error.

3.3 First differences: location diversity and absolute errors of consensus forecasts

Even though we use firm fixed effects to control for unobserved heterogeneity, and control for the (lagged) number of analysts following the firm, it is possible that change in diversity is related to factors that change analyst coverage and also affect the information environment (and thus forecasting performance). In fact, we find a statistically highly significant and negative relation between the change in HHI and the contemporaneous change in coverage. We therefore conduct our tests in first-difference – the idea being that controlling for change in coverage captures the expected change in diversity due to a change in coverage, so that any residual change in diversity is essentially random and unrelated to factors that might have caused forecasting performance and coverage to change simultaneously. In all these first difference regressions, we find our location diversity measures as significant as before.

The first three columns of Table 4 report the results of the regressions in first differences for mean consensus forecasts, controlling for the contemporaneous change in coverage and dropping the firm fixed effects. The estimated coefficients of our diversity measures (0.413 – 0.477) are slightly smaller than those (0.480 – 0.578) of the level regressions in Table 3, but still highly significant at the 1% level. By contrast, the estimated coefficients of the change in the contemporaneous coverage are insignificant.

The results of the other control variables remain qualitatively similar. Noticeable exceptions are *SP500* and *MNMD*. *SP500*, the S&P 500 dummy, is positive in the level regressions, but statistically insignificant in the difference regressions. *MNMD*, skewness of EPS, is insignificant in the level regressions, but becomes positive and highly statistically significant in the difference regression. This suggests that when there is an increase in skewness of EPS, the mean consensus is less accurate.

It is well known that compared with the mean, the median is less affected by extreme values. It is possible that the mean consensus error reflects the influence of more extreme individual forecasts.[6] Therefore, we also consider median consensus forecasts. Specifically, for each firm year, we change the dependent variable to the first difference in the forecast error of

[6] Some studies (e.g., Clement and Tse 2005) find that bold forecasts are more accurate.

the median consensus forecast. We also change the control variables to the first difference in their medians. The last 3 columns of Table 4 report the results. The estimated coefficients of the location HHIs (0.450 – 0.546) for median consensus forecasts are larger than and as highly significant as those (0.413 – 0.477) for mean consensus forecasts in the first three columns of the same table. These suggest that diversity could give rise to more extreme forecasts that make the mean forecast less accurate than the median, mirroring the summary statistics in Table 1. As for the change in the contemporaneous coverage, its estimated coefficients remain insignificant at the standard levels.[7]

3.4 Controlling for absolute error of first forecast

Next, we adopt another approach to dealing with the possibility that forecast accuracy and analyst diversity both change together in response to any common unmodelled reason, e.g., a change in the firm's information environment. In particular, we include the absolute error of the first consensus forecast as an additional control variable. The idea is that the *incremental accuracy* of the last forecast over the first, where the group of the analysts is the same for both the last and first forecast, is less likely to be affected by a change in the firm environment that also causes diversity to change. Therefore, any remaining effect of diversity on the accuracy of the last forecast must reflect the information contained in forecast *revisions* by analysts. It is still possible, however, that a steadily improving information environment will improve forecast accuracy via forecast revisions. However, such an effect should manifest by reducing *the bias* in the forecasts.[8] However, unlike Merkley et al. (2017), we find no evidence that greater location diversity reduces bias. Thus, any effect of diversity on the forecast error must be happening through the diversification channel – the *MAD* is lower because greater diversity

[7] We get similar results for level regressions whether the consensus is defined as the deviation from the median or from the mean.
[8] The bias is defined as the signed difference between the consensus forecast and the realized earnings, scaled by the previous period's stock price.

means the errors are less correlated and therefore "cancel out" each other, improving forecast accuracy of the consensus forecast. It is difficult to argue that a change or improvement in the information environment should lead to less correlated errors.

The first 3 columns of Table 5 report the results for regressions in which the dependent variable is the absolute error of the mean consensus forecast, based on the last available forecast of each analyst of a firm year. Among the control variables, we include the absolute error of the consensus comprised of the first forecast of each analyst of the same firm year. All three of our location HHI measures have significantly positive effects on the consensus forecast errors. Their magnitude is approximately half that of those without controlling the absolute error of the first consensus forecast in Table 3.

The last three columns of Table 5 present the results for the forecast bias of the mean consensus forecast for exactly the same specification. None of the HHI indices affect forecast bias.[9] Thus, overall, our results suggest that analyst diversity improves consensus forecast accuracy through the diversification channel. Unlike Merkley et al. (2017), there is no effect on forecast bias, which is also reassuring because this argues against the possibility that we are picking up a reverse causality running from improvement in firm environment, change in coverage, and change in diversity. As regards the untabulated results for the control variables, controlling for the first consensus forecast error renders *COVERAGE*, *LNBM*, *RET* and *SIGMA* insignificant.

3.5 Location diversity and absolute errors of individual forecasts

We next examine how analyst diversity affects individual forecasts. Diversity can affect individual forecasts if and only if individual analysts assign some positive weights to the forecasts/opinions of the other analysts who cover the same firm over the same fiscal year when

[9] HHI does not have any significant effect on forecast bias if the regressions are specified in first differences. We do not report these results in a table.

producing their individual forecasts. Based on our discussion in Section 2, we expect the individual forecasts of more diverse groups to be more accurate. Other than the overall HHI, we now consider two additional diversity measures. One is a slightly modified HHI measure based on the locations of other analysts in regions excluding that of the analyst in question. The other new measure is the proportion of analysts following the firm in the given year that come from the same geographical region as the analyst in question.

We examine the diversity effect on the very last forecast (by any analyst) during the forecasting period, as well as the last forecasts of all analysts. Our model is essentially the same as (3.1) above, but has the following modifications. First, when we can construct the variables at the analyst level, we replace the firm-level variables in (3.1) by the corresponding analyst-level variable. Second, we further include analyst fixed effects as controls. Finally, the robust standard errors are based on clustering at the firm and analyst levels. Table 6 reports the results for the absolute errors of the individual forecasts of all firm years. Column (1) reports the results for the last forecast of any analyst of each firm year. Columns (2) – (4) report the results for the last forecasts of all analysts of each firm year.

Consistent with our expectation, in Columns (1) and (2), the estimated coefficients of *BEA_HHI* are positive and statistically significant at the 1% level, but smaller than those for the consensus forecasts. To gauge the economic significance, we make similar comparisons as those for the absolute mean consensus errors. Consider again the homogeneous and heterogeneous groups of two analysts. Based on the estimated coefficients in Columns (1) and (2) of Table 6, compared with the homogeneous group, the heterogeneous group have, on average, 0.186% and 0.133% (half the HHI estimated coefficients) lower absolute individual errors, compared with the unconditional medians of the absolute all and last individual errors of 0.210% and 0.238% (in Table 1), respectively. They amount to 2.8¢ and 2.0¢ (4.1¢ and 2.9¢) per share based on the median (mean) share price.

In Column (3) of Table 6, the estimated coefficient of the proportion of analysts from the analyst's region (*OWN REGION'S PROPORTION*) is also positive and statistically highly significant. This suggests that an individual analyst generally benefits more from diversity when a larger proportion of analysts are located in regions different from his.

Finally, in Column (4) of Table 6, the estimated coefficient of BEA_HHI based on the remaining regions outside the analyst's region (*OTHER REGIONS' BEA_HHI*) is positive and statistically significant at the 1% level. This shows that an individual analyst forecasts more accurately when the diversity of the remaining analysts, following the same firm in the same forecasting period, outside his region is higher. This suggests that individual analysts pay attention to the other analysts; the group distribution of the other analysts affect their forecast accuracy – the larger the dispersion of the remaining distribution, the higher the accuracy of the individual forecasts.

To illustrate the economic impacts of *OWN REGION'S PROPORTION* and *OTHER REGIONS' BEA_HHI*, let us consider two group of analysts (Group A and Group B). Each group has the same number of analysts, 3 analysts. The only difference between these two groups is their location distributions. In Group A, two of the 3 analysts, including the analyst (analyst *i*) in the question, are located in one region and the third analyst located in another region. In Group B, each analyst, including an analyst identical to analyst *i* in all aspects, is in a different region. Hence, the difference in *OWN REGION'S PROPORTION* between Group A and Group B is -1/3 and the difference in *OTHER REGIONS' BEA_HHI* between Group A and Group B is -0.5. Based on the estimated coefficients in Table 6, the absolute error of analyst *i* in the more diverse group, Group B, is estimated to be 0.04% lower (associated with *OWN REGION'S PROPORTION*) and 0.065% lower (associated with *OTHER REGIONS' BEA_HHI*) than that of the otherwise identical analyst in Group A. The estimates come to

0.5952¢ and 0.9672¢ (0.8719¢ and 1.417¢) and sum up to 1.57¢ (2.29¢) per share based on the median (mean) share price.

2.5.1 First differences: location diversity and absolute errors of individual forecasts

As in the case of the consensus forecasts, we examine individual forecasts in first differences, control for the contemporaneous change in coverage and drop the firm and analyst fixed effects. Table 7 reports the results and shows that the first difference results are generally *economically stronger* than and statistically as strong as those of the level regressions in Table 6. The magnitude of the coefficients is larger up to 81.67%. Compared with the results of the level regressions, Δ*OWN REGION'S PROPORTION* has largest increase in the diversity effect while Δ*BEA_HHI* reveals the least increase.

2.5.2 Controlling for absolute error of first forecast for individual forecasts

As the analysis of the consensus forecasts, we also alleviate the endogeneity concern for our analysis of individual forecasts by controlling for the absolute error of the first forecast. We thus re-run the individual level regressions in Table 6 by including the first forecast error as an additional explanatory variable. We have three alternative first forecasts – the individual, median consensus, and mean consensus forecast (reported in the first, second, and third row in Column (5), respectively). Table 8 summaries the results for location diversity. Column (3) reports the estimated coefficients of our diversity measures, given in Column (2), after controlling for the absolute error of the first forecast made by the same individual or the consensus of all the analysts for the same firm year.

By additionally controlling the first forecast error, the estimated coefficients of our diversity measures have smaller magnitude (approximately about half), but significance as strong as before. The results controlling for the first consensus error, which should capture better the reverse causality, are close to those controlling for the first individual error, both in terms of economic and statistical significance. Overall, our results show that the individual forecasts of groups of more geographically heterogeneous analysts generally have smaller errors. Individual errors are also smaller when the other locational categories of analysts make up a larger proportion of the group and when the geographical diversity of the remaining categories of the group is higher. These suggest that analysts incorporate the information of each other's specific signals when producing their forecasts, whereby improving their forecast accuracy; the more the analyst diversity, the higher the individual forecast accuracy.

3.6 Experience diversity and consensus and individual forecasts

We repeat the above analyses for experience diversity. In general, we find qualitatively very similar results as for location diversity. Tables 9, 10 and 11 report the major results. Table 9 presents the regression in levels using firm and year fixed effects for all regressions; in addition, we include analyst fixed effects for regressions of individual forecasts. The HHI based on analyst experience has a significant positive coefficient for the mean and median consensus forecast errors, as well as all individual last forecast errors.

Table 10 reports results of regressions in first differences. Consistent with the results of location diversity, Δ*EXP_HHI*, Δ*OWN EXP_GROUP'S PROPORTION* and Δ*OTHER GROUPS' EXP_HHI* are all positive and statistically significant, except for the mean consensus with p-value of 0.14. Compared with the estimated coefficients of location diversity, the corresponding coefficients of experience diversity are generally smaller, excluding Δ*EXP_HHI* and Δ*OTHER GROUPS' EXP_HHI* for individual forecasts.

In Table 11, we repeat the analyses for experience diversity for individual forecasts after controlling for the first (mean consensus, median consensus and individual) forecast error. We reach qualitatively the same conclusion as for location diversity in Table 8. One noticeable difference is that the results of *OWN GROUP'S PROPORTION* is weaker, being marginally insignificant when controlling for first individual forecast or median consensus error. While we do not report these results, we find no effect of experience diversity on forecast bias.

3.7 Proportions of subgroups constituting HHIs

One concern is whether the above HHI results are driven by forecast accuracy of a particular group. For example, analysts in regions with more social capital may produce forecasts that are more accurate because of better external knowledge acquisition (Uzzi 1997; Laursen at el. 2012). This raises the possibility that the location-HHI actually captures the forecast accuracy of analysts in the regions with more social capital. To address this concern, we additionally control for the proportions of the analysts in all the subgroups composing the HHIs. Tables A1 and A3 report the results of regressions of mean forecast consensuses for location and experience diversity, respectively. In general, we find that the estimated coefficients of our diversity measures have similar magnitude and levels of significance whether we include the proportions of the analysts in the subgroups. Meanwhile, all of the estimated coefficients of the proportions are largely small and statistically insignificant.

3.8 Gender and ethnic diversity

Finally, we consider diversity based on two *visible demographic characteristics* – gender and ethnicity. As discussed in the introduction, previous research does not find diversity based on such characteristics to improve group or individual performance; however, since interpersonal relationships are less important in the context of analyst performance, it is interesting to explore whether our setting yields different results. We do not find any effect of diversity along these dimensions on analyst forecast error. Below, we briefly describe our attempts.

It is difficult to accurately identify the gender and ethnicity of all analysts. We employ a number of approaches. The gender of analysts is classified, based on their given names, as follows. First, we obtain analysts' surnames and initials, abbreviations of names of the brokerage firms in which the analysts worked, and time of the above data items from the I/B/E/S. The analysts' given names are missing in the I/B/E/S. In addition to the above I/B/E/S data items, the annual volumes of Nelson's Directory of Investment Research have the given names of analysts.[10] We then match the surnames, the initials, the names of the brokerage firms and the time of the data of I//B/E/S and Nelson's Directory. When there is insufficient information for matching, we conduct internet searches. In this way, we obtain the given names of 7838 analysts filed in I/B/E/S. The Social Security Administration of the United States (SSA) provides information of the frequency of given names of the new-born babies, together with their gender, in the United States since 1880.

We then assume that analysts filed in I/B/E/S are aged between 20 and 65. We thus use the corresponding years of the SSA data to calculate the proportion of female of all babies with a particular given name.[11] We classify an analyst as a female if the proportion of female in the SSA data for the analyst's given name is above 99%. The remaining analysts are either assumed to be male (in one classification) or classified as a male if the female proportion is below 1% (in another classification). We also use the 90% or 50% female proportion as the cut-off, in place of the cut-off of 99%.[12]

We next construct different gender-HHIs based on the above gender classifications, and use them as the diversity measures in the regressions. However, the estimated coefficients of these gender-HHIs for regressions of both forecast consensuses and individual forecasts are generally insignificant. For example, the estimated coefficient of the gender-HHI based on the

[10] We use the Nelson's Directories of 1997, 1999, 2001, 2003 and 2005.
[11] For example, for an analyst filed in 1997, we assume that he/she is born in a year between 1932 and 1977. We thus use the SSA data of these years.
[12] The 99% and 90% classifications yield 15.87% and 16.01% as the proportion of female of all analysts. Kumar (2010) and Green et al. (2007) report similar corresponding female proportions (16%; 13.9% - 16.1% with an average of 15.6%).

cut-offs of the 99% and 1% for the regression of mean consensus errors is 0.256, with a *t* statistics of 0.59.

As for ethnic diversity, we classify the ethnicity of analysts based on their surnames as follows. First, we retrieve analysts' surnames from the I/B/E/S database. Next, we obtain the proportions of population of different ethnic groups associated with different surnames of the 2000 Census from the U.S. Census Bureau. There are six ethnic groups: Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian and Pacific Islander, Non-Hispanic American Indian and Alaskan Native, Non-Hispanic of two or more races, and Hispanic. We then classify the ethnicity of analysts in several ways. First, like the gender classification, we classify an analyst's ethnicity as that of the group with at least 90% of the population who have the analyst's surname, based on the Census data. Alternatively, we group an analyst into an ethnic group with the highest proportion of the population who have the analyst's surname, among all ethnic groups. We then construct different ethnic-HHIs based on these ethnic classifications. Finally, instead of classifying analysts into a specific ethnic group, we use the proportions of population of all ethnic groups associated with each analyst's surname in computing the HHI. We separately employ these ethnic-HHIs as the diversity measure and re-run regressions. Generally, the estimated coefficients of these ethnic-HHIs for regressions of both consensus and individual forecasts are typically insignificant.

4 Conclusion

Using US data over the 1982 to 2014 period, we study the effect of diversity in terms of location and experience of the group of analysts forecasting earnings of the same firm over the same fiscal year on the accuracy of both the consensus forecast as well as individual analyst forecasts. We first document that the more heterogeneous the group of analysts, the smaller are the errors of their mean and median forecasts. We show that this is not because greater diversity reduces forecast bias. Rather, the effect is consistent with more diverse forecasts averaging out the individual errors. We further show that the individual forecasts of these more heterogeneous analysts also generally have smaller errors, reflecting that analysts take into consideration of idiosyncratic components of each other's information when producing their forecasts. Moreover, an individual analyst on average makes smaller forecasting errors when the proportion of analysts not in the same category as that analyst is larger, and when there is greater diversity over the remaining categories (excluding his own category). In contrast to diversity along the dimensions of location and experience, gender and ethnicity-based diversity measures have no significant effects on analyst forecast accuracy.

References

Ashraf, Q., & Galor, O. (2013). The "out of Africa" hypothesis, human genetic diversity, and comparative economic development. American Economic Review, 103, 1–46.

Bae, K-H., Stulz, R. M., & Tan, H. (2008). Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts. *Journal of Financial Economics*, *88*, 581–606.

Bae, K-H., Tan, H, & Welker, M. (2008). International GAAP differences: The impact on foreign analysts. *Accounting Review*, *83*, 593-628.

Byrne, D.E. (1971). *The attraction paradigm*. New York, Academic Press.

Christensen, T. E. (2007). Discussion of "letting the `tail wag the dog': the debate over GAAP versus street earnings revisited". *Contemporary Accounting Research, 24*(3), 741-762.

Clement, M. B. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics*, *27*(3), 285-303.

Clement, M. B., Hales, J., & Xue, Y. (2011). Understanding analysts' use of stock returns and other analysts' revisions when forecasting earnings. *Journal of Accounting and Economics, 51*(3), 279-299.

Clement, M. B., & Tse, S. Y. (2005). Financial analyst characteristics and herding behavior in forecasting. *Journal of Finance*, *60*(1), 307-341.

Cohn, J. and J. Juergens, 2014, How Much Do Analysts Influence Each Other's Forecasts?, Quarterly Journal of Finance, 2014, 4(3):1-35.

Denis, D. J., Denis, D. K., & Sarin, A. (1997). Agency problems, equity ownership, and corporate diversification. *Journal of Finance*, *52*(1), 135-160.

Einhorm, H., R. Hogarth, and E. Klempner (1977), Quality of Group Judgment. *Psychological Bulletin, 84*, 158-172.

Erhardt, N. L., Werbel, J. D., & Shrader, C. B. (2003). Board of director diversity and firm financial performance. *Corporate governance: An international review*, *11*(2), 102-111.

Green, T. C., Jegadeesh, N., & Tang, Y. (2007). Gender and job performance: Evidence from Wall Street (No. w12897). National Bureau of Economic Research.

Gruenfeld, D., Mannix, E., Williams, K., & Neale, M. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational Behavior and Human Decision Processes*, *67*, 1-15.

Gu, Z., & Wu, J. S. (2003). Earnings skewness and analyst forecast bias. *Journal of Accounting and Economics*, *35*, 5-29.

Hansen, Z., Owan, H., & Pan, J. (2006). *The Impact of Group Diversity on Performance and Knowledge Spillover--An Experiment in a College Classroom* (No. w12251). National Bureau of Economic Research.

Herfindahl, O. C. (1950). Concentration in the steel industry. Doctoral dissertation, Columbia University.

Hirschman, A. O. (1964). The paternity of an index. American Economic Review, 54(5), 761-762.

Hong, H., & Kacperczyk, M. (2010). Competition and bias. *Quarterly Journal of Economics*, *125*, 1683-1725.

Hong, H., & Kubik, J. E. (2003). Analyzing the analysts: career concerns and biased earnings forecasts, *Journal of Finance, 58*, 313 – 351.

Hong, H., Kubik, J. D., & Solomon, A. (2000). Security analysts' career concerns and herding of earnings forecasts. *Rand Journal of Economics*, 121-144.

Hotchkiss, E. S., & Strickland, D. (2003). Does shareholder composition matter? Evidence from the market reaction to corporate earnings announcements. *Journal of Finance*, *58*(4), 1469-1498.

Jegadeesh, N., & Kim, W. (2009). Do analysts herd? An analysis of recommendations and market reactions. *Review of Financial Studies*, *23*(2), 901-937.

Ke, B., & Yu, Y. (2006). The effect of issuing biased earnings forecasts on analysts' access to management and survival. *Journal of Accounting Research*, *44*(5), 965 – 999.

Kramer, B., & MacKinnon, A. (December 1993). Localization: theory and experiment. *Reports on Progress in Physics*. *56*(12), 1469.

Kumar, A. (2010). Self-Selection and the Forecasting Abilities of Female Equity Analysts. *Journal of Accounting Research*, *48*(2), 393-435.

Laakso, M., & Taagepera, R. (1979). "Effective" number of parties: a measure with application to West Europe. *Comparative Political Studies*, *12*(1), 3-27.

Lang, L. H., & Stulz, R. M. (1994). Tobin's q, corporate diversification, and firm performance. *Journal of Political Economy*, *102*(6), 1248-1280.

Laursen, K., Masciarelli, F., & Prencipe, A. (2012). Regions matter: How localized social capital affects innovation and external knowledge acquisition. *Organization Science*, *23*(1), 177-193.

McDonald, D. G., & Dimmick, J. (2003). The conceptualization and measurement of diversity. *Communication Research*, *30*(1), 60-79.

Palan, N. (2010). Measurement of specialization: The choice of indices. FIW Working Paper, No. 62.

Pelled, L.H. (1996). Demographic diversity, conflict and work group outcomes: An intervening process theory. *Organization Science*, *7*(6), 615-631.

Pelled, L. H., Eisenhardt, K. M., & Xin, K. R. (1999). Exploring the black box: An analysis of work group diversity, conflict and performance. *Administrative science quarterly*, *44*(1), 1-28.

Rhoades, S. A. (1993). The herfindahl-hirschman index. *Federal Reserve Bulletin*, *79*, 188.

Richter, A. W., Hirst, G., Van Knippenberg, D., & Baer, M. (2012). Creative self-efficacy and individual creativity in team contexts: Cross-level interactions with team informational resources. Journal of Applied Psychology, 97(6), 1282.

Simpson, E.H. (1949). Measurement of diversity. *Nature*, *163*, 688.

Thatcher, S. M. (1999). The contextual importance of diversity: The impact of relational demography and team diversity on individual performance and satisfaction. *Performance Improvement Quarterly*, *12*(1), 97-112.

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, 35-67.

Van Dijk, H., Van Engen, M. L., & Van Knippenberg, D. (2012). Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes*, *119*(1), 38-53.

Van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual review of psychology*, *58*.

Walther, B. R., & Willis, R. H. (2013). Do investor expectations affect sell-side analysts' forecast bias and forecast accuracy? *Review of Accounting Studies*, *18*(1), 207-227.

Welch, I. (2000). Herding among security analysts. *Journal of Financial Economics*, *58*(3), 369-396.

Williams, K. Y., & O'Reilly, C.A., III (1998). *The complexity of diversity: A review of forty years of research.* In E. Gruenfeld & M. Neale (Eds.), Research on managing in groups and teams. Thousand Oaks, CA: Sage.

Wise, L. R., & Tschirhart, M. (2000). Examining empirical evidence on diversity effects: how useful is diversity research for public-sector managers? *Public Administration Review*, *60*(5), 386-394.

Appendix A. Variables

This appendix lists the definition of all variables used in this paper.

A.1. Dependent variables

Absolute forecast error (in percentage terms):

$$ABS\ ERROR = \frac{|F-A|}{P} \times 100 \quad \text{(A1.1)}$$

where *ABS ERROR* is absolute forecast error, estimated as the absolute difference between *F* and *A*, scaled by *P*. *F* is the median or mean annual EPS (earnings per share) forecast of all analysts for firm *j* in year *t*. *A* is the actual value of the EPS being forecasted. *P* is the market share price of firm *j* as of the end of the last fiscal year. In the analyses of individual forecasts, *ABS ERROR* is the absolute different between the last annual EPS forecast of the analyst and *A*, scaled by *P*.

Forecast bias (in percentage terms):

$$BIAS = \frac{F-A}{P} \times 100 \quad \text{(A1.2)}$$

where *BIAS* is forecast bias, estimated as the difference between *F* and *A*, scaled by *P*. *F* is the median or mean annual EPS (earnings per share) forecast of all analysts for firm *j* in year *t*. *A* is the actual value of the EPS being forecasted. *P* is the market share price of firm *j* as of the end of the last fiscal year.

A.2. Key explanatory variables

BEA_HHI:

$$BEA\_HHI_{j,t} = \sum_{q=1}^{9} \left(\frac{Number\ of\ analysts\ in\ location\ q_{j,t}}{Total\ number\ of\ analysts_{j,t}} \right)^2 \quad \text{(A2.1)}$$

$BEA\_HHI_{j,t}$ is the location-Herfindahl-Hirschman Index for firm j in year t. $Total\ number\ of\ analysts_{j,t}$ is the total number of analysts producing forecasts for firm j in year t. *Number of analysts in location* $q_{j,t}$ is the number of analysts, producing forecasts for firm j in year t, in the q^{th} location where q is one of the 8 Bureau of Economic Analysis regions or any non-US country.

CENSUS_HHI:

$$CENSUS\_HHI_{j,t} = \sum_{q=1}^{5} \left(\frac{Number\ of\ analysts\ in\ location\ q_{j,t}}{Total\ number\ of\ analysts_{j,t}} \right)^2 \quad \text{(A2.2)}$$

$CENSUS\_HHI_{j,t}$ is the location-Herfindahl-Hirschman Index for firm j in year t. $Total\ number\ of\ analysts_{j,t}$ is the total number of analysts producing forecasts for firm j in year t. *Number of analysts in location* $q_{j,t}$ is the number of analysts, producing forecasts for firm j in year t, in the q^{th} location where q is one of the 4 Census Bureau-designated regions or any non-US country.

STATES_HHI:

$$STATES\_HHI_{j,t} = \sum_{q=1}^{52+} \left(\frac{Number\ of\ analysts\ in\ location\ q_{j,t}}{Total\ number\ of\ analysts_{j,t}} \right)^2 \quad \text{(A2.3)}$$

$STATES\_HHI_{j,t}$ is the location-Herfindahl-Hirschman Index for firm j in year t. $Total\ number\ of\ analysts_{j,t}$ is the total number of analysts producing forecasts for firm j in year t. *Number of analysts in location* $q_{j,t}$ is the number of analysts, producing forecasts for firm j in year t, in the q^{th} location where q is one of the US states or a non-US country.

OWN REGION'S PROPORTION:

$$OWN\ REGION'S\ PROPORTION_{i,j,t} = \frac{Number\ of\ analysts\ in\ analyst\ i's\ location_{j,t}}{Total\ number\ of\ analysts_{j,t}} \quad \text{(A2.4)}$$

$OWN\ REGION'S\ PROPORTION_{i,j,t}$ is the proportion of the analysts following firm j in year t located in analyst i's location. $Total\ number\ of\ analysts_{j,t}$ is the total number of analysts producing forecasts for firm j in year t. *Number of analysts in analyst i's location*$_{j,t}$ is the number of analysts, producing forecasts for firm j in year t, in analyst i's location where the location is one of the 8 Bureau of Economic Analysis regions or any non-US country.

OTHER REGIONS' BEA_HHI:

$$OTHER\ REGIONS'\ BEA\_HHI_{i,j,t} = \sum_{q=1}^{8}\left(\frac{Number\ of\ analysts\ in\ location\ q_{j,t}}{Total\ number\ of\ remaining\ analysts_{j,t}}\right)^2 \quad \text{(A2.5)}$$

where the q^{th} location is not analyst i's location. $OTHER\ REGIONS'\ BEA\_HHI_{i,j,t}$ is the BEA_HHI of the analysts in the locations where analyst i is not located for firm j in year t. $Total\ number\ of\ remaining\ analysts_{j,t}$ is the total number of the remaining analysts producing forecasts for firm j in year t, excluding the number of analysts in analyst i's location. *Number of analysts in location* $q_{j,t}$ is the number of analysts, producing forecasts for firm j in year t, in the q^{th} location where q is one of the 8 remaining locations outside analyst i's location.

EXP_HHI:

$$EXP\_HHI_{j,t} = \sum_{k=1}^{5}\left(\frac{Number\ of\ analysts\ in\ group\ k_{j,t}}{Total\ number\ of\ analysts_{j,t}}\right)^2 \quad \text{(A2.6)}$$

$EXP\_HHI_{j,t}$ is the experience-Herfindahl-Hirschman Index for firm j in year t. $Total\ number\ of\ analysts_{j,t}$ is the total number of analysts producing forecasts for firm j in year t. *Number of analysts in group* $k_{j,t}$ is the number of analysts, producing forecasts for firm j in year t, in the k^{th} group based on the quintiles of the length of experience of all analysts in the same year, where k = {1, 2, 3, 4, 5}.

OWN EXP_GROUP'S PROPORTION:

$$OWN\ EXP\_GROUP'S\ PROPORTION_{i,j,t} = \frac{Number\ of\ analysts\ in\ analyst\ i's\ group_{j,t}}{Total\ number\ of\ analysts_{j,t}} \qquad (A2.7)$$

$OWN\ EXP\_GROUP'S\ PROPORTION_{i,j,t}$ is the proportion of the analysts following firm j in year t in analyst i's experience group. $Total\ number\ of\ analysts_{j,t}$ is the total number of analysts producing forecasts for firm *j* in year *t*. *Number of analysts in analyst i's group*$_{j,t}$ is the number of analysts, producing forecasts for firm *j* in year *t*, in analyst *i*'s experience group where the experience group is one of the 5 groups based on the quintiles of the length of experience of all analysts in the same year.

OTHER GROUPS' EXP_HHI:

$$OTHER\ GROUPS'\ EXP\_HHI_{i,j,t} = \sum_{k=1}^{4}\left(\frac{Number\ of\ analysts\ in\ group\ k_{j,t}}{Total\ number\ of\ remaining\ analysts_{j,t}}\right)^2 \qquad (A2.8)$$

where the k^{th} group is not analyst *i*'s experience group. $OTHER\ GROUPS'\ EXP\_HHI_{i,j,t}$ is the EXP_HHI of the remaining 4 experience groups that are not analyst *i*'s group. $Total\ number\ of\ remaining\ analysts_{j,t}$ is the total number of remaining analysts producing forecasts for firm *j* in year *t*, excluding the number of analysts in analyst *i*'s experience group. *Number of analysts in group* $k_{j,t}$ is the number of analysts, producing forecasts for firm *j* in year *t*, in the k^{th} group based on the quintiles of the length of experience of all analysts in the same year and to which analyst *i* does not belong.

A.3. Control variables

Following the literature (e.g., Hong and Kacperczyk 2010; Ke and Yu 2006; Gu and Wu 2003; Clement 1999), we employ the following firm-specific, analyst-specific and time-specific control variables. When the dependent variable is based on the median (mean) of the last forecasts of all analysts for firm *j* and year *t*, the medians (means) of the following control variables with (*), of the observations of the last forecasts of all analysts, for firm *j* and year *t* are used.

Firm-Specific Control Variables:

CONTEMPORANEOUS COVERAGE$_{j,t}$

A measure of analyst coverage, defined as the number of analysts covering firm *j* for fiscal year *t* and constructed using I/B/E/S data.

COVERAGE$_{j,t-1}$

A measure of analyst coverage, defined as the number of analysts covering firm *j* for fiscal year *t*-1 and constructed using I/B/E/S data.

FIRST MEAN OR MEDIAN CONSENSUS FORECAST ABS ERROR j,t

The absolute difference between the first annual EPS mean or median consensus forecast of all analysts, published after the end of the last fiscal year, and the actual value of the EPS being forecasted for firm *j* in year *t*, scaled by the market share price of firm *j* as of the end of the last fiscal year – constructed using I/B/E/S and Compustat data.

LNBM$_{j,t-1}$

The natural logarithm of firm *j*'s book value of equity divided by its market capitalization at the end of fiscal year *t*-1.

LNBM = ln(ceq/(csho × prcc_f)) (A3.1, from Compustat)

LNSIZE$_{j,t-1}$

The natural logarithm of firm *j*'s market capitalization at the end of fiscal year *t*-1.

LNSIZE = ln(csho × prcc_f) (A3.2, from Compustat)

$MNMD_{j,t}$

A skewness measure, defined as the difference between the mean and median of EPS, scaled by the market share price for the end of fiscal year *t*-1 for firm *j* for the period between fiscal year *t*-4 and fiscal year *t* + 4, excluding fiscal year *t*, using I/B/E/S data.

$PROFIT_{j,t-1}$

The operating income of firm *j* for fiscal year *t*-1 over the book value of firm assets as of the end of fiscal year *t*-2.

PROFIT = ib/lagged ceq (A3.3, from Compustat)

$RET_{j,t-1}$ (*)

The average monthly stock returns for firm *j* for the past 12 months in relation to the date of the last actual annual earnings, constructed using I/B/E/S and CRSP data.

$SIGMA_{j,t}$ (*)

The variance of the raw monthly stock returns for firm *j* for the past 12 months in relation to the month in which the forecast is released, constructed using I/B/E/S and CRSP data.

$SP500_{j,t}$

An indicator that equals one if firm *j* is in the S&P 500 index when the forecast is released and equals 0 otherwise.

$VOLROE_{j,t-1}$

The variance of the residuals from an AR(1) model for firm *j*'s annual ROE using the past ten-fiscal-year series. ROE is calculated as the ratio of earnings to the beginning book value of equity.

ROE = ib/lagged ceq (A3.4, from Compustat)

Analyst-Specific Control Variables:

FIRST INDIVIDUAL FORECAST ABS ERROR i,j,t

The absolute difference between the first annual EPS forecast of analyst *i*, published after the end of the last fiscal year, and the actual value of the EPS being forecasted for firm *j* in year *t*, scaled by the market share price of firm *j* as of the end of the last fiscal year – constructed using I/B/E/S and Compustat data.

$LNGENERAL\_EXP_{i,t}$ (*)

The natural logarithm of 1 plus the number of days since the release day of the first annual EPS forecast made by analyst *i*, constructed using I/B/E/S data.

$LNEXP\_WITH\_FIRM_{i,j,t}$ (*)

The natural logarithm of 1 plus the number of days since the release day of the first annual EPS forecast for firm j made by analyst *i*, constructed using I/B/E/S data.

$LOSS_{i,j,t}$ (*)

An indicator that equals one if the forecast made by analyst *i* is negative and equals 0 otherwise.

$NCO_{i,t}$ (*)

The number of firms for which analyst *i* generated annual EPS forecasts in fiscal year *t*, constructed using I/B/E/S data.

$NSIC3_{i,t}$ (*)

The number of 3-digit SIC industries for which analyst *i* made annual EPS forecasts in fiscal year *t*, constructed using I/B/E/S data.

Time-Specific Control Variables:

$LNHORIZON_{i,j,t}$ (*)

The natural logarithm of 1 plus the number of days between the release date of analyst *i*'s earnings forecast for firm *j* and the data date of the earnings being forecast, constructed using I/B/E/S data.

$RETTODATE_{i,j,t}$ (*)

The cumulative stock returns (using monthly data) for firm *j* between the data date of the last annual earnings and the date on which the earnings forecast by analyst *i* is released, constructed using I/B/E/S and CRSP data.

Table 1. Summary statistics

ABS ERROR (Mean or Median consensus) is the absolute error of the mean or median of the last forecasts of all analysts for a firm year. *ABS ERROR* (All analysts) is the absolute error of the last forecasts of all analysts for a firm year. *ABS ERROR* (Last analyst) is the absolute error of the last forecast of any analyst for a firm year. MEAN or MEDIAN DIFF is the mean or median of the absolute errors of the last forecasts of all analysts minus the mean or median consensus absolute error in the same firm year. *STATES_HHI*, *BEA_HHI* and *CENSUS_HHI* are the Herfindahl-Hirschman Indices of all analysts for a firm year for their geographical locations based on the US states and non-US countries, based on the 8 Bureau of Economic Analysis regions and non-US countries, and based on the 4 Census Bureau-designated regions and non-US countries, respectively. *EXP_HHI* is the Herfindahl-Hirschman Index of the 5 groups of all analysts of the corresponding firm year, based on the quintiles of the length of experience of all analysts in the same year. Variables with (*) is the mean of the last observation of all analysts for a firm year. *COVERAGE* is the number of analysts covering firm *j* in fiscal year *t*-1. *LNSIZE* is the natural logarithm of firm *j*'s market capitalization at the end of fiscal year *t*-1. *LNBM* is the natural logarithm of firm *j*'s book value divided by its market capitalization at the end of fiscal year *t*-1. *RET* is the average monthly stock returns for firm *j* for the past 12 months in relation to the data date of the last actual annual earnings. *SIGMA* is the standard deviation of the raw monthly stock returns for firm *j* for the past 12 months in relation to the forecast publication month. *VOLROE* is the variance of the residuals from an AR(1) model for firm *j*'s annual ROE using the past ten-year series. ROE is calculated as the ratio of earnings to the beginning book value of equity. *PROFIT* is the operating income of firm *j* for fiscal year *t*-1 over the book value of assets of the firm as of the end of fiscal year t-2. *SP500* equals one if firm j is in the S&P 500 index on the day that the forecast is released and equals 0 otherwise. *MNMD* is the difference between the mean and median of EPS, scaled by the market share price for the end of fiscal year *t*-1 for firm *j* for the period between fiscal year *t*-4 and fiscal year *t* + 4, excluding fiscal year *t*. *LOSS* equals one if the forecast made by analyst *i* is negative and equals 0 otherwise. *RETTODATE* is the cumulative stock returns (using monthly data) for firm *j* between the last annual earnings and the release of the earnings forecast by analyst *i*. *NCO* is the number of firms for which analyst *i* has generated annual EPS forecasts in fiscal year *t*. *NSIC3* is the number of 3-digit SIC industries for which analyst *i* has generated annual EPS forecasts in fiscal year *t*. *EXP_WITH_FIRM* is the number of years since the release of the first annual EPS forecast for firm *j* generated by analyst *i*. *GENERAL_EXP* is the number of years since the release of the first annual EPS forecast generated by analyst *i*.

	Mean	S.D.	25%	Median	75%
ABS ERROR (Mean consensus)	1.163	2.895	0.097	0.291	0.861
ABS ERROR (Median consensus)	1.042	2.787	0.073	0.228	0.703
ABS ERROR (All analysts)	0.889	2.593	0.069	0.210	0.628
ABS ERROR (Last analysts)	1.066	3.060	0.078	0.238	0.714
MEAN DIFF (individual - consensus)	0.166	0.773	0.000	0.009	0.098
MEDIAN DIFF (individual - consensus)	0.086	0.554	0.000	0.000	0.031
STATES_HHI	0.462	0.228	0.296	0.401	0.556
BEA_HHI	0.526	0.225	0.360	0.500	0.625
CENSUS_HHI	0.558	0.220	0.388	0.500	0.636
EXP_HHI	0.344	0.150	0.246	0.288	0.375
COVERAGE	10.902	8.966	4.000	8.000	15.000
LNSIZE	6.743	1.723	5.480	6.613	7.864
LNBM	-0.791	0.692	-1.196	-0.728	-0.333
RET	0.014	0.038	-0.007	0.012	0.032
SIGMA (*)	0.116	0.065	0.070	0.101	0.145
VOLROE	0.131	0.652	0.002	0.007	0.029
PROFIT	0.090	0.249	0.038	0.122	0.192
SP500	0.200	0.399	0.000	0.000	0.000
MNMD	-0.002	0.023	-0.004	0.000	0.005
LOSS (*)	0.126	0.315	0.000	0.000	0.000
RETTODATE (*)	0.025	0.335	-0.135	0.046	0.208
NCO (*)	32	55	15	18	24
NSIC3 (*)	8	8	4	7	10
EXP_WITH_FIRM (*), in years.	1.674	1.463	0.619	1.328	2.314
GENERAL_EXP (*), in years	5.283	2.688	3.422	4.950	6.737

Table 2. Variance inflation factors (VIF) and correlation of explanatory variables

Refer to Appendix A for variable definition.

	VIF	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) STATES_HHI	5.68									
(2) BEA_HHI	7.74	0.90								
(3) CENSUS_HHI	5.47	0.85	0.90							
(4) EXP_HHI	1.58	0.34	0.32	0.32						
(5) COVERAGE	2.83	-0.25	-0.24	-0.25	-0.49					
(6) LNSIZE	3.79	-0.19	-0.18	-0.20	-0.48	0.75				
(7) LNBM	1.48	0.13	0.14	0.14	0.18	-0.20	-0.34			
(8) RET	1.28	-0.02	-0.03	-0.02	-0.01	-0.07	0.09	-0.38		
(9) SIGMA	1.39	0.01	-0.03	-0.01	0.10	-0.16	-0.35	0.01	0.00	
(10) VOLROE	1.07	0.03	0.02	0.01	0.04	-0.06	-0.08	-0.14	0.03	0.14
(11) PROFIT	1.57	-0.08	-0.05	-0.05	-0.13	0.15	0.30	-0.14	0.15	-0.35
(12) SP500	2.16	-0.01	-0.01	-0.02	-0.28	0.62	0.68	-0.17	-0.04	-0.22
(13) MNMD	1.04	-0.02	-0.02	-0.02	-0.06	0.08	0.12	-0.10	0.07	-0.10
(14) LOSS	1.71	0.06	0.03	0.03	0.11	-0.12	-0.27	0.03	-0.14	0.41
(15) RETTODATE	1.1	0.00	0.01	0.01	-0.01	-0.01	-0.04	0.11	-0.08	-0.09
(16) NCO	6.54	0.13	0.12	0.13	0.02	-0.05	-0.07	0.03	0.02	0.02
(17) NSIC3	6.83	0.17	0.17	0.16	0.06	-0.11	-0.10	0.03	0.01	0.02
(18) LNEXP_WITH_FIRM	1.61	0.07	0.08	0.06	0.09	0.13	0.13	0.13	-0.12	-0.12
(19) LNGENERAL_EXP	1.64	0.03	0.04	0.03	0.13	0.03	0.09	0.12	-0.07	-0.10

Table 2. (continued)

	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(11) PROFIT	-0.13								
(12) SP500	-0.06	0.17							
(13) MNMD	-0.02	0.13	0.06						
(14) LOSS	0.18	-0.55	-0.15	-0.13					
(15) RETTODATE	-0.02	0.04	0.01	0.06	-0.20				
(16) NCO	0.00	0.01	-0.01	0.01	-0.03	0.02			
(17) NSIC3	-0.01	0.06	-0.02	0.02	-0.07	0.03	0.91		
(18) LNEXP_WITH_FIRM	-0.06	0.02	0.15	0.01	-0.03	-0.02	-0.04	0.01	
(19) LNGENERAL_EXP	-0.03	0.02	0.08	-0.01	-0.04	0.03	0.08	0.14	0.58

Table 3. Analyst Location Diversity and Absolute Error of Mean Consensus Forecasts

The dependent variable is the absolute error of the mean of the last forecasts of all analysts of a firm in a year (*ABS ERROR*). *STATES_HHI*, *BEA_HHI* and *CENSUS_HHI* are the Herfindahl-Hirschman Indices of all analysts for a firm year for their geographical locations based on the US states and non-US countries, based on the 8 Bureau of Economic Analysis regions and non-US countries, and based on the 4 Census Bureau-designated regions and non-US countries, respectively. Refer to Appendix A for the definition of the variables. Year and firm fixed effects are included. Standard errors are based on clustering at the firm level. Estimated coefficients and the robust standard errors (in parentheses) are reported. ***, ** and * indicate the 1%, 5% and 10% levels of significance, respectively.

	(1)	(2)	(3)
STATES_HHI	0.578***		
	(0.113)		
BEA_HHI		0.521***	
		(0.110)	
CENSUS_HHI			0.480***
			(0.113)
NCO	-0.003**	-0.003**	-0.003**
	(0.001)	(0.001)	(0.001)
NSIC3	0.020**	0.020**	0.020**
	(0.010)	(0.010)	(0.010)
LNEXP_WITH_FIRM	0.037*	0.037*	0.038*
	(0.020)	(0.020)	(0.020)
LNGENERAL_EXP	0.031	0.028	0.025
	(0.043)	(0.043)	(0.043)
COVERAGE	0.017***	0.016***	0.016***
	(0.004)	(0.004)	(0.004)
LNSIZE	-0.882***	-0.887***	-0.889***
	(0.057)	(0.056)	(0.057)
LNBM	0.305***	0.306***	0.309***
	(0.060)	(0.060)	(0.060)
RET	-3.123***	-3.118***	-3.105***
	(0.574)	(0.573)	(0.574)
SIGMA	6.763***	6.765***	6.730***
	(0.580)	(0.580)	(0.579)
VOLROE	-0.085	-0.086	-0.085
	(0.065)	(0.065)	(0.065)
PROFIT	-0.210*	-0.214*	-0.216*
	(0.122)	(0.122)	(0.122)
SP500	0.170**	0.172**	0.176**
	(0.085)	(0.085)	(0.085)
MNMD	-2.686	-2.685	-2.667
	(1.977)	(1.978)	(1.979)
LOSS	2.097***	2.100***	2.099***
	(0.134)	(0.134)	(0.134)

Table 3. (continued)

	(1)	(2)	(3)
LNHORIZON	0.469***	0.468***	0.468***
	(0.036)	(0.036)	(0.036)
RETTODATE	-0.740***	-0.743***	-0.744***
	(0.068)	(0.068)	(0.068)
Observation	34,612	34,612	34,612
Adjusted R^2	0.426	0.426	0.426

Table 4. First Difference in Analyst Location Diversity and Absolute Error of Mean and Median Consensus Forecasts

The dependent variable is the first difference in the absolute error of the mean or median of the last forecasts of all analysts of a firm in a year (Δ*ABS ERROR*). Δ*STATES_HHI*, Δ*BEA_HHI* and Δ*CENSUS_HHI* are the first difference in the Herfindahl-Hirschman Indices of all analysts for a firm year for their geographical locations based on the US states and non-US countries, based on the 8 Bureau of Economic Analysis regions and non-US countries, and based on the 4 Census Bureau-designated regions and non-US countries, respectively. Refer to Appendix A for the definition of the variables. Year fixed effects are included. Standard errors are based on clustering at the firm level. Estimated coefficients and the robust standard errors (in parentheses) are reported. ***, ** and * indicate the 1%, 5% and 10% levels of significance, respectively.

Consensus forecast:	Mean	Mean	Mean	Median	Median	Median
	(1)	(2)	(3)	(4)	(5)	(6)
ΔSTATES_HHI	0.477***			0.546***		
	(0.139)			(0.138)		
ΔBEA_HHI		0.468***			0.512***	
		(0.139)			(0.137)	
ΔCENSUS_HHI			0.413***			0.450***
			(0.137)			(0.136)
ΔCONTEMPORANEOUS COVERAGE	0.007	0.006	0.006	-0.003	-0.004	-0.004*
	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.002)
ΔNCO	-0.003*	-0.003*	-0.003*	-0.005***	-0.004*	0.029*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.016)
ΔNSIC3	0.023**	0.025**	0.025**	0.033**	0.029*	0.049**
	(0.011)	(0.013)	(0.013)	(0.014)	(0.016)	(0.022)
ΔLNEXP_WITH_FIRM	0.029	0.032	0.032	0.043**	0.049**	-0.013
	(0.023)	(0.025)	(0.025)	(0.021)	(0.022)	(0.051)
ΔLNGENERAL_EXP	0.040	0.023	0.019	0.045	-0.005	-0.005
	(0.048)	(0.052)	(0.051)	(0.046)	(0.051)	(0.005)
ΔLNSIZE	-1.511***	-1.493***	-1.497***	-1.403***	-1.420***	-1.422***
	(0.120)	(0.117)	(0.117)	(0.120)	(0.117)	(0.117)
ΔLNBM	0.329***	0.364***	0.365***	0.271***	0.319***	0.319***
	(0.104)	(0.103)	(0.103)	(0.105)	(0.103)	(0.103)
ΔRET	0.456	0.389	0.393	0.636	0.760	0.767
	(0.672)	(0.692)	(0.692)	(0.655)	(0.672)	(0.673)
ΔSIGMA	4.174***	4.116***	4.099***	3.322***	3.343***	3.315***
	(0.585)	(0.589)	(0.588)	(0.512)	(0.518)	(0.518)
ΔVOLROE	-0.371***	-0.389***	-0.388***	-0.390***	-0.399***	-0.390***
	(0.134)	(0.122)	(0.122)	(0.134)	(0.120)	(0.119)
ΔPROFIT	0.316**	0.240	0.241	0.169	0.100	0.103
	(0.155)	(0.149)	(0.149)	(0.145)	(0.141)	(0.141)
ΔSP500	0.074	0.016	0.019	0.083	0.044	0.047
	(0.155)	(0.162)	(0.162)	(0.142)	(0.146)	(0.146)
ΔMNMD	19.859***	17.763***	17.784***	17.843***	15.464***	15.571***
	(3.030)	(2.902)	(2.903)	(2.994)	(2.907)	(2.908)
ΔLOSS	1.201***	1.268***	1.268***	0.784***	0.834***	0.831***
	(0.136)	(0.135)	(0.135)	(0.121)	(0.118)	(0.118)

Table 4. (continued)

Consensus forecast:	Mean	Mean	Mean	Median	Median	Median
	(1)	(2)	(3)	(4)	(5)	(6)
∆LNHORIZON	0.396***	0.383***	0.382***	0.351***	0.343***	0.343***
	(0.034)	(0.036)	(0.036)	(0.030)	(0.030)	(0.030)
∆RETTODATE	-0.742***	-0.764***	-0.766***	-0.577***	-0.608***	-0.612***
	(0.073)	(0.074)	(0.074)	(0.064)	(0.065)	(0.065)
Observation	28,507	28,507	28,507	28,037	28,507	28,507
Adjusted R^2	0.118	0.118	0.118	0.100	0.100	0.100

Table 5. Analyst Location Diversity and Absolute Error/Bias of Mean Consensus Forecasts, Controlling for Absolute Error/Bias of First Mean Consensus Forecast

This table reports the results of the regressions for which we additionally include the absolute error (*FIRST CONSENSUS FORECAST ABS ERROR*) or bias (*FIRST CONSENSUS FORECAST BIAS*) of the mean of the first forecasts of all analysts of a firm in a year. The dependent variable is the absolute error or bias of the mean of the last forecasts of all analysts of the same firm year. *STATES_HHI*, *BEA_HHI* and *CENSUS_HHI* are the Herfindahl-Hirschman Indices of all analysts for a firm year for their geographical locations based on the US states and non-US countries, based on the 8 Bureau of Economic Analysis regions and non-US countries, and based on the 4 Census Bureau-designated regions and non-US counties, respectively. Year and firm fixed effects are included in all regressions. Standard errors are based on clustering at the firm level. For the sake of brevity, the results for the other control variables are not reported. Estimated coefficients and the robust standard errors (in parentheses) are reported. *** indicates the 1% level of significance.

Mean consensus forecast:	ABS ERROR	ABS ERROR	ABS ERROR	BIAS	BIAS	BIAS
	(1)	(2)	(3)	(4)	(5)	(6)
STATES_HHI	0.268*** (0.062)			0.066 (0.065)		
BEA_HHI		0.236*** (0.062)			0.048 (0.062)	
CENSUS_HHI			0.236*** (0.064)			0.043 (0.064)
FIRST CONSENSUS FORECAST ABS ERROR	0.496*** (0.008)	0.495*** (0.008)	0.495*** (0.008)			
FIRST CONSENSUS FORECAST BIAS				0.505*** (0.007)	0.505*** (0.007)	0.505*** (0.007)
Observations	33,773	33,773	33,773	33,773	33,773	33,773
Adjusted R^2	0.780	0.779	0.780	0.750	0.750	0.750

Table 6. Analyst Location Diversity and Absolute Error of Individual Forecasts

The dependent variable is the absolute error of the individual last forecast of a firm in a year (*ABS ERROR*). *BEA_HHI* is the Herfindahl-Hirschman Index of the geographical locations, based on the 8 Bureau of Economic Analysis regions and non-US countries, of all analysts of the corresponding firm year. *OWN REGION'S PROPORTION* is the proportion of the analysts located in the region where the analyst is located. *OTHER REGIONS' BEA_HHI* is *BEA_HHI* based on the remaining Bureau of Economic Analysis regions and non-US countries where the analyst is not located. For the sake of brevity, the results for the other control variables are not reported. Year, firm and analyst fixed effects are included. Standard errors are based on clustering at the firm and analyst levels. Estimated coefficients and the robust standard errors (in parentheses) are reported. *** indicates the 1% level of significance.

Absolute forecast error:	Last analyst	All analysts	All analysts	All analysts
	(1)	(2)	(3)	(4)
BEA_HHI	0.371***	0.266***		
	(0.083)	(0.065)		
OWN REGION'S PROPORTION			0.120***	
			(0.026)	
OTHER REGIONS' BEA_HHI				0.130***
				(0.044)
Observations	31,739	379,761	304,971	232,210
Adjusted R^2	0.388	0.401	0.411	0.408

Table 7. First Difference in Analyst Location Diversity and Absolute Error of Individual Forecasts

The dependent variable is the first difference in the absolute error of the individual last forecast of all analysts of firm *j*, i.e., the change from year *t*-1 to year *t*. Δ*BEA_HHI* is the first difference in the Herfindahl-Hirschman Index of the geographical locations, based on the 8 Bureau of Economic Analysis regions and non-US countries, of all analysts of the corresponding firm year. Δ*OWN REGION'S PROPORTION* is the first difference in the proportion of the analysts located in the region where the analyst in question is located. Δ*OTHER REGIONS' BEA_HHI* is the first difference in *BEA_HHI* based on the remaining Bureau of Economic Analysis regions and non-US countries where the analyst is not located. Δ*CONTEMPORANEOUS COVERAGE* is the first difference in the number of analysts covering firm *j*, i.e., the change from year *t*-1 to year *t*. For the sake of brevity, the results for the other control variables are not reported. Year fixed effects are included. Standard errors are based on clustering at the firm and analyst levels. Estimated coefficients and the robust standard errors (in parentheses) are reported. *** indicates the 1% level of significance.

	(1)	(2)	(3)
ΔBEA_HHI	0.316***		
	(0.073)		
ΔOWN REGION'S PROPORTION		0.218***	
		(0.047)	
ΔOTHER REGIONS' BEA_HHI			0.225***
			(0.073)
ΔCONTEMPORANEOUS COVERAGE	-0.002	-0.005	-0.005
	(0.003)	(0.003)	(0.004)
Observation	242,788	180,894	140,523
Adjusted R^2	0.0933	0.0921	0.0931

Table 8. Analyst Location Diversity and Absolute Error of Individual Forecasts, Controlling for First Individual or Consensus Forecast Error

This table summarises the results of the regressions for which we additionally include the absolute error of the first individual, or mean or median consensus forecast as an explanatory variable. Column (1) shows the dependent variables. Column (2) gives the key variables of interest. *BEA_HHI* is the Herfindahl-Hirschman Index of the geographical locations, based on the 8 Bureau of Economic Analysis regions and non-US countries, of all analysts of the corresponding firm year. *OWN REGION'S PROPORTION* is the proportion of the analysts located in the region where the analyst is located. *OTHER REGIONS' BEA_HHI* is *BEA_HHI* based on the remaining Bureau of Economic Analysis regions and non-US countries where the analyst is not located. Estimated coefficients of these key variables are reported in Column (3). Column (4) gives the key control variable of *FIRST FORECAST ABS ERROR*, for which the estimated coefficients are reported in Column (5). For the sake of brevity, the results for the other control variables are not reported. Year, firm and analyst fixed effects are included. Robust standard errors are based on clustering at the firm and analyst levels. *** indicates the 1% level of significance.

LAST FORECAST ABS ERROR	Independent variable of interest	Coefficient of Column (2) variable	Key control variable (FIRST FORECAST ABS ERROR)	Coefficient of Column (4) variable
(1)	(2)	(3)	(4)	(5)
Last analyst	BEA_HHI	0.211***	Individual forecast	0.289***
Last analyst	BEA_HHI	0.196***	Median consensus	0.287***
Last analyst	BEA_HHI	0.192***	Mean consensus	0.312***
All analysts	BEA_HHI	0.149***	Individual forecast	0.325***
All analysts	BEA_HHI	0.138***	Median consensus	0.336***
All analysts	BEA_HHI	0.124***	Mean consensus	0.369***
All analysts	OWN REGION'S PROPORTION	0.064***	Individual forecast	0.391***
All analysts	OWN REGION'S PROPORTION	0.067***	Median consensus	0.408***
All analysts	OWN REGION'S PROPORTION	0.059***	Mean consensus	0.449***
All analysts	OTHER REGIONS' BEA_HHI	0.116***	Individual forecast	0.379***
All analysts	OTHER REGIONS' BEA_HHI	0.127***	Median consensus	0.400***
All analysts	OTHER REGIONS' BEA_HHI	0.111***	Mean consensus	0.442***

Table 9. Analyst Experience Diversity and Absolute Forecast Error

The dependent variable is the absolute error of the last forecast of a firm in a year (*ABS ERROR*). *EXP_HHI* is the Herfindahl-Hirschman Index of the 5 groups of all analysts of the corresponding firm year, based on the quintiles of the length of experience of all analysts in the same year. *OWN EXP_GROUP'S PROPORTION* is the proportion of the analysts in the analyst's group of the length of experience. *OTHER GROUPS' EXP_HHI of* is *EXP_HHI* based on the remaining 4 groups of the length of experience. Refer to Appendix A for the definition of the other variables. Year and firm fixed effects are included in all regressions. Analyst fixed effects are further included in Columns (3)-(5). Standard errors are based on clustering at the firm level in all regressions, and also at the analyst level in Columns (3)-(5). Estimated coefficients and the robust standard errors (in parentheses) are reported. ***, ** and * indicate the 1%, 5% and 10% levels of significance, respectively.

Absolute forecast error:	Mean consensus	Median consensus	Last analyst	All analysts	All analysts
	(1)	(2)	(3)	(4)	(5)
EXP_HHI	0.684***	0.892***	0.612***	0.661***	
	(0.149)	(0.149)	(0.169)	(0.139)	
OWN EXP_GROUP'S PROPORTION					-0.046
					(0.041)
OTHER GROUPS' EXP_HHI					0.505***
					(0.085)
NCO	-0.002*	-0.003**	-0.001	-0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
NSIC3	0.016*	0.026**	0.007	-0.002	-0.002
	(0.009)	(0.010)	(0.005)	(0.002)	(0.002)
LNEXP_WITH_FIRM	0.034*	0.025	0.005	-0.004	-0.004
	(0.018)	(0.017)	(0.007)	(0.002)	(0.002)
LNGENERAL_EXP	0.008	0.002	0.022	0.013**	0.013*
	(0.038)	(0.037)	(0.019)	(0.007)	(0.007)
COVERAGE	0.016***	0.015***	0.015***	0.013***	0.013***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
LNSIZE	-0.804***	-0.755***	-0.762***	-0.674***	-0.674***
	(0.048)	(0.047)	(0.059)	(0.041)	(0.044)
LNBM	0.286***	0.251***	0.214***	0.199***	0.206***
	(0.052)	(0.050)	(0.064)	(0.040)	(0.042)
RET	-3.686***	-3.410***	-3.325***	-3.291***	-3.212***
	(0.518)	(0.520)	(0.646)	(0.409)	(0.422)
SIGMA	7.143***	6.254***	7.280***	5.727***	5.866***
	(0.534)	(0.485)	(0.596)	(0.449)	(0.462)
VOLROE	-0.098*	-0.094*	-0.254**	-0.170**	-0.157**
	(0.050)	(0.048)	(0.116)	(0.075)	(0.079)
PROFIT	-0.186*	-0.327***	-0.558***	-0.491***	-0.451***
	(0.102)	(0.098)	(0.136)	(0.086)	(0.089)

Table 9. (continued)

	Mean consensus	Median consensus	Last analyst	All analysts	All analysts
	(1)	(2)	(3)	(4)	(5)
SP500	0.222***	0.221***	0.143	0.189***	0.159**
	(0.080)	(0.076)	(0.096)	(0.062)	(0.062)
MNMD	-3.513**	-3.197**	-1.452	-3.016*	-3.731**
	(1.646)	(1.603)	(1.989)	(1.721)	(1.864)
LOSS	1.937***	1.444***	1.718***	1.116***	1.119***
	(0.118)	(0.103)	(0.124)	(0.097)	(0.100)
LNHORIZON	0.466***	0.441***	0.121***	0.324***	0.349***
	(0.034)	(0.030)	(0.016)	(0.011)	(0.011)
RETTODATE	-0.609***	-0.479***	-0.405***	-0.317***	-0.330***
	(0.061)	(0.054)	(0.062)	(0.042)	(0.046)
Observations	41,315	41,315	38,267	462,649	372,429
Adjusted R^2	0.427	0.407	0.350	0.374	0.384

Table 10. First Difference in Analyst Experience Diversity and Absolute Forecast Error

The dependent variable is the first difference in the absolute error of the last forecast of firm *j*, i.e., the change from year *t*-1 to year *t*. This is of the mean consensus forecast in Column (1), the median consensus forecast in Column (2) and the last forecasts of all analysts in Columns (3) – (5). Δ*EXP_HHI* is the first difference in the Herfindahl-Hirschman Index of the 5 groups of all analysts of firm *j*, based on the quintiles of the length of experience of all analysts, i.e., the change from year *t*-1 to year *t*. Δ*OWN EXP_GROUP'S PROPORTION* is the first difference in the proportion of the analysts in the analyst's group of the length of experience. Δ*OTHER GROUPS' EXP_HHI* is Δ*EXP_HHI* based on the remaining 4 experience groups to which the analyst does not belong. Δ*CONTEMPORANEOUS COVERAGE* is the first difference in the number of analysts covering firm *j*, i.e., the change from year *t*-1 to year *t*. For the sake of brevity, the results for the other control variables are not reported. Year fixed effects are included. Standard errors are based on clustering at the firm level in all regressions, and also at the analyst level in Columns (3) - (5). Estimated coefficients and the robust standard errors (in parentheses) are reported. *** and ** indicate the 1% and 5% levels of significance, respectively.

Absolute forecast error:	Mean consensus	Median consensus	All analysts	All analysts	All analysts
	(1)	(2)	(3)	(4)	(5)
ΔEXP_HHI	0.244[1]	0.381**	0.453***		
	(0.166)	(0.168)	(0.167)		
ΔOWN EXP_GROUP'S PROPORTION				0.192***	
				(0.069)	
ΔOTHER GROUPS' EXP_HHI					0.405***
					(0.101)
ΔCONTEMPORANEOUS COVERAGE	0.002	-0.007	-0.005	-0.005	-0.003
	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)
Observations	33,978	33,978	305,217	226,930	225,354
Adjusted R^2	0.1055	0.0903	0.0765	0.0843	0.0848

[1]: The t statistic is 1.47 with p-value of 0.140.

Table 11. Analyst Experience Diversity and Absolute Error of Individual Forecasts, Controlling for First Individual or Consensus Forecast Error

This table summarises the results of the regressions for which we additionally include the absolute error of the first individual, median consensus, or mean consensus forecast as an explanatory variable. Column (1) shows the dependent variables. Column (2) gives the key variables of interest. *EXP_HHI* is the Herfindahl-Hirschman Index of the 5 groups of all analysts of the corresponding firm year, based on the quintiles of the length of experience of all analysts in the same year. *OWN EXP_GROUP'S PROPORTION* is the proportion of the analysts in the analyst's group of the length of experience. *OTHER GROUPS' EXP_HHI* is *EXP_HHI* based on the remaining 4 groups of the length of experience. Estimated coefficients of these key variables are reported in Column (3). Column (4) gives the key control variable of *FIRST FORECAST ABS ERROR*, for which the estimated coefficients are reported in Column (5). For the sake of brevity, the results for the other control variables are not reported. Year, firm and analyst fixed effects are included. Robust standard errors are based on clustering at the firm and analyst levels. ***, ** and * indicate the 1%, 5% and 10% levels of significance, respectively.

LAST FORECAST ABS ERROR	Independent variable of interest	Coefficient of Column (2) variable	Key control variable (FIRST FORECAST ABS ERROR)	Coefficient of Column (4) variable
(1)	(2)	(3)	(4)	(5)
Last analyst	EXP_HHI	0.195[1]	Individual forecast	0.340***
Last analyst	EXP_HHI	0.223*	Median consensus	0.339***
Last analyst	EXP_HHI	0.188[2]	Mean consensus	0.359***
All analysts	EXP_HHI	0.240***	Individual forecast	0.352***
All analysts	EXP_HHI	0.193***	Median consensus	0.377***
All analysts	EXP_HHI	0.137**	Mean consensus	0.414***
All analysts	OWN EXP_GROUP'S PROPORTION	0.048*[,3]	Individual forecast	0.390***
All analysts	OWN EXP_GROUP'S PROPORTION	0.036[4]	Median consensus	0.403***
All analysts	OWN EXP_GROUP'S PROPORTION	0.018	Mean consensus	0.442***
All analysts	OTHER GROUPS' EXP_HHI	0.163***	Individual forecast	0.389***
All analysts	OTHER GROUPS' EXP_HHI	0.159***	Median consensus	0.402***
All analysts	OTHER GROUPS' EXP_HHI	0.111***	Mean consensus	0.441***

[1]: t = 1.47 and p-value = 0.141. [2]: t = 1.39 and p-value = 0.164. [3]: t = 1.91 and p-value = 0.056. [4]: t = 1.39 and p-value = 0.165.

Table A1. Analyst Location Diversity & Subgroup Proportions

The dependent variable is the absolute error of the mean of the last forecasts of all analysts of a firm in a year (*ABS ERROR*). *BEA_HHI* is the Herfindahl-Hirschman Index of the geographical locations, based on the 8 Bureau of Economic Analysis regions and non-US countries, of all analysts of the corresponding firm year. *Proportion of "X"* is the proportion of analysts located in the "X" Region, where "X" = {New England, Mideast, Great Lakes, Plains, Southeast, Southwest, Rocky Mountain, Far West}. *Proportion of non-US locations* is the proportion of analysts located outside the US. Year and firm fixed effects are included. Standard errors are based on clustering at the firm level. For the sake of brevity, the results for the other control variables are not reported. Estimated coefficients and the robust standard errors (in parentheses) are reported. *** indicates the 1% level of significance.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
BEA_HHI	0.516***	0.574***	0.520***	0.526***	0.529***	0.520***	0.524***	0.529***	0.531***
	(0.112)	(0.126)	(0.111)	(0.111)	(0.112)	(0.110)	(0.110)	(0.112)	(0.112)
Proportion of New England	-0.094								
	(0.251)								
Proportion of Mideast		-0.117							
		(0.109)							
Proportion of Great Lakes			-0.018						
			(0.171)						
Proportion of Plains				0.093					
				(0.185)					
Proportion of Southeast					0.092				
					(0.163)				
Proportion of Southwest						-0.045			
						(0.321)			
Proportion of Rocky Mountain							0.191		
							(0.460)		
Proportion of Far West								0.077	
								(0.170)	
Proportion of non-US locations									0.154
									(0.256)
Observation	34,612	34,612	34,612	34,612	34,612	34,612	34,612	34,612	34,612
Adjusted R^2	0.426	0.426	0.426	0.426	0.426	0.426	0.426	0.426	0.426

Table A2. Analyst Location Diversity and Absolute Error of Median Consensus Forecasts

The dependent variable is the absolute error of the median of the last forecasts of all analysts of a firm in a year (*ABS ERROR*). *STATES_HHI*, *BEA_HHI* and *CENSUS_HHI* are the Herfindahl-Hirschman Indices of all analysts for the firm year for their geographical locations based on the US states and non-US countries, based on the 8 Bureau of Economic Analysis regions and non-US countries, and based on the 4 Census Bureau-designated regions and non-US countries, respectively. Refer to Appendix A for the definition of the other variables. Year and firm fixed effects are included. Standard errors are based on clustering at the firm level. Estimated coefficients and the robust standard errors (in parentheses) are reported. ***, ** and * indicate the 1%, 5% and 10% levels of significance, respectively.

	(1)	(2)	(3)
STATES_HHI	0.626***		
	(0.112)		
BEA_HHI		0.560***	
		(0.110)	
CENSUS_HHI			0.520***
			(0.113)
NCO	-0.004***	-0.004***	-0.004**
	(0.002)	(0.002)	(0.002)
NSIC3	0.036***	0.036***	0.036***
	(0.012)	(0.012)	(0.012)
LNEXP_WITH_FIRM	0.032*	0.031*	0.031*
	(0.019)	(0.019)	(0.019)
LNGENERAL_EXP	0.006	0.005	0.000
	(0.043)	(0.043)	(0.043)
COVERAGE	0.016***	0.015***	0.015***
	(0.004)	(0.004)	(0.004)
LNSIZE	-0.856***	-0.862***	-0.864***
	(0.056)	(0.056)	(0.056)
LNBM	0.262***	0.263***	0.266***
	(0.059)	(0.059)	(0.059)
RET	-2.733***	-2.726***	-2.714***
	(0.566)	(0.566)	(0.566)
SIGMA	5.906***	5.917***	5.881***
	(0.525)	(0.525)	(0.525)
VOLROE	-0.105	-0.109*	-0.109#
	(0.066)	(0.066)	(0.066)
PROFIT	-0.383***	-0.392***	-0.394***
	(0.119)	(0.119)	(0.119)
SP500	0.187**	0.190**	0.194**
	(0.081)	(0.081)	(0.081)
MNMD	-3.010	-3.007#	-2.988#
	(1.956)	(1.956)	(1.957)
LOSS	1.576***	1.578***	1.577***
	(0.117)	(0.117)	(0.117)

Table A2. (continued)

	(1)	(2)	(3)
LNHORIZON	0.441***	0.439***	0.440***
	(0.032)	(0.032)	(0.032)
RETTODATE	-0.598***	-0.601***	-0.603***
	(0.060)	(0.060)	(0.060)
Observation	34,612	34,612	34,612
Adjusted R^2	0.407	0.407	0.406

Table A3. Analyst Experience Diversity & Subgroup Proportion

The dependent variable is the absolute error of the mean of the last forecast of all analysts of a firm in a year (*ABS ERROR*). *EXP_HHI* is the Herfindahl-Hirschman Index of the 5 groups of all analysts of the corresponding firm year, based on the quintiles of the length of experience of all analysts in the same year. *Proportion of "X"* is the proportion of analysts in Group "X", where "X" = {1 (with least experience), 2, 3, 4, 5 (with most experience)}. Year and firm fixed effects are included. Standard errors are based on clustering at the firm level. For the sake of brevity, the results for the other control variables are not reported. Estimated coefficients and the robust standard errors (in parentheses) are reported. ***, ** and * indicate the 1%, 5% and 10% levels of significance, respectively.

	(1)	(2)	(3)	(4)	(5)
EXP_HHI	0.684***	0.681***	0.684***	0.687***	0.685***
	(0.149)	(0.149)	(0.149)	(0.149)	(0.150)
Proportion of Group 1	0.011				
(with least experience)	(0.143)				
Proportion of Group 2		-0.177*			
		(0.100)			
Proportion of Group 3			-0.004		
			(0.097)		
Proportion of Group 4				0.202**	
				(0.100)	
Proportion of Group 5					-0.001
(with most experience)					(0.121)
Observations	41,315	41,315	41,315	41,315	41,315
Adjusted R^2	0.427	0.427	0.427	0.427	0.427