WARWICK
THE UNIVERSITY OF WARWICK

Manuscript version: Author's Accepted Manuscript
The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:
http://wrap.warwick.ac.uk/166995

How to cite:
Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:
Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Do Differences in Analyst Quality Matter for Investors Relying on Consensus Information?

Roni Michaely, Amir Rubin, Dan Segal, and Alexander Vedrashko*

June 15, 2022

Abstract:

This study investigates whether investors can reap economic benefits from analyzing differences in analyst quality. Although high-quality analysts' average forecast is more accurate than the consensus forecast for firms with a large analyst following, the benefits of using high-quality analysts' average forecast are not economically significant. In contrast, the value of analyst quality differentiation exists in the second moment of forecasts. High-quality analysts' forecast dispersion gives investors an advantage in dealing with uncertainty by predicting return volatility and providing opportunities for economically significant returns using option straddle and post-earnings announcement drift investment strategies.

Keywords: consensus, analyst quality, forecasts, dispersion, post-earnings announcement drift

JEL: G10, G11, G14, G24, M41

*Michaely is from the The University of Hong Kong, (ronim@hku,hk) and SFI. Rubin is from Simon Fraser University and Reichman University (arubin@sfu.ca). Segal is from Reichman University and Warwick University (dsegal@idc.ac.il). Vedrashko is from Simon Fraser University (awv@sfu.ca).

It is well-recognized that there are persistent differences across analysts in ability and information sets, which affects their forecast accuracy.[1] The greater forecast accuracy of superior analysts implies that investors may be better off using the forecasts of only these analysts. While the distinction between high- and low-quality individual analysts is instructive and intuitively appealing, investors and academics alike routinely use the joint information provided by all analysts, as evidenced by the importance of "consensus estimates." This apparent contradiction leads us to examine whether investors should go beyond consensus-type information and rely only on aggregate signals produced by high-quality analysts, that is, whether information derived from analyst quality differences has incremental economic value relative to the information in consensus measures. In this study, we examine the economic benefits of using high-quality analysts' (i) average forecast and (ii) forecast dispersion relative to the corresponding measures derived from the forecasts of all analysts following the firm.

There are three main arguments for why investors may prefer the consensus forecast over the average forecast of superior analysts: error cancelation, persistence, and cognitive costs. First, a phenomenon known as the wisdom of the crowds suggests that aggregating multiple individual opinions reduces the collective error.[2] More specifically, it is well-established that the average signal of forecasters diversifies errors made by individual forecasters and constitutes the optimal choice for predictability (Ahlers and Lakonishok, 1983; Granger and Newbold, 1986). Since over- and under-estimations should cancel each other out, it is difficult to beat the consensus forecast in

[1] Individual analysts' forecast accuracy systematically differs for reasons such as analysts' varying experience (Mikhail, Walther, and Willis, 1997; Clement, 1999; Hirst, Hopkins, and Wahlen, 2004), aptitude (Jacob, Lys, and Neale, 1999), education (Maines, McDaniel, and Harris, 1997; De Franco and Zhou, 2009), brokerage house association and underwriting relationships (Lin and McNichols, 1998; Clement, 1999), proximity to the firm (Malloy, 2005), lead analyst and star status categorization (Stickel, 1992; Cooper, Day, and Lewis, 2001), or work habits (Rubin, Segal, and Segal, 2017).

[2] One of the tenets of the theory is that greater independence among the participants makes for better accuracy. Consistent with this idea, Merkley, Michaely, and Pacelli (2021) show that the error of the consensus forecast is negatively associated with cultural diversity among analysts.

terms of accuracy. Thus, the comparatively smaller number of high quality (HQ) analysts relative to the total number of analysts negatively affects the HQ analysts' average forecast accuracy relative to that of the consensus because of the lower diversification effect. Second, persistence in quality is not perfect (Sinha, Brown, and Das, 1997). Investors cannot know in advance which analysts will be more accurate next period because competition among analysts, private information, and incentives related to forecast accuracy lead to changes in analysts' relative forecast accuracy over time (Horton, Serafeim, and Wu, 2017).[3] Third, investors incur financial and cognitive costs to obtain and process information about analysts' quality variation, including identifying HQ analysts, while obtaining the consensus forecast is relatively easy and inexpensive. For these three reasons, it is unclear *a priori* whether there are economic benefits to using HQ analysts' average forecast instead of relying on the consensus forecast.

In contrast to the average, that is, the consensus forecast, the diversification effect is not a factor in forecast dispersion, which is a proxy for uncertainty and information asymmetry. Forecast dispersion based on the forecasts of all analysts following a firm (consensus dispersion) does not benefit from the diversification effect because it is based on the distribution of forecasts and not on their accuracy. Consequently, given that on average, HQ analysts individually possess superior information and ability, the variability of HQ analysts' forecasts could potentially be a superior measure of uncertainty regarding future firm performance than the consensus dispersion.

[3] Moreover, even if investors know with certainty who will be the most accurate analyst next period, they should not necessarily follow this analyst or a group of most accurate analysts. Since accuracy is measured in terms of absolute forecast errors, but individual forecast errors can be either positive or negative, the consensus can be more accurate than the average forecast of high-quality analysts. Consider a firm followed by four analysts, whose rankings in terms of forecast accuracy are known, that is, perfectly persistent over time. Let the top analyst's forecast error (the analyst's forecast minus actual earnings per share) be 2¢, and the remaining three analysts have forecast errors of -4¢, -5¢, and 8¢. The consensus forecast error is then 0.25¢ (=(2-4-5+8)/4), which is less, in absolute terms, than both the forecast error of the most accurate analyst and the forecast error of the average forecast of the analysts in the top half of the ranking. As an applied example, Coca-Cola had an actual annual EPS of $1.46 with the most accurate individual forecast of $1.45 and the consensus forecast of $1.46 (forecasts ranging from $1.38 to $1.54).

We start the empirical analysis with a comparison of the accuracy of HQ analysts' average and consensus forecasts.[4] It is important to point out that, in contrast to studies of the effects of analyst heterogeneity on individual forecasts or revisions, this study analyzes forecast *aggregates*, e.g., averages, whose accuracy is affected by diversification. Consequently, we expect that the results will be sensitive to the number of HQ analysts following the firm.[5] Indeed, we find that HQ analysts' average forecast is more accurate than the consensus forecast when a sufficient number of HQ analysts follow the firm.[6]

We next examine whether there are economic benefits from utilizing HQ analysts' superior forecasts. To address this question, we first assess the extent to which investors incorporate HQ analysts' average forecasts into prices when forming expectations about future earnings. The earnings response coefficient (ERC) for the consensus-based earnings surprise is greater than the ERC for the HQ average forecast both for firms whose HQ average forecast is more accurate than the consensus forecast and in the full sample. Since investors do not fully utilize the superior information available in HQ analysts' average forecasts, one can measure the degree of potential market inefficiency by forming a long-short trading strategy that uses HQ analysts' average forecast to predict "earnings surprise." Specifically, the strategy buys the stock a day before the earnings announcement when the HQ average forecast is greater than the consensus forecast (the

[4] Following the literature, high-quality analysts are a fraction of all analysts following the firm and are determined based on their past forecast accuracy ranking for this firm (e.g., Stickel, 1992; Sinha, Brown, and Das, 1997). We find significant (rather than perfect) persistence in forecasting performance across time and across firms covered by an analyst, indicating that forecasting performance does a good job capturing analyst quality. This study's results are robust to different definitions of HQ analysts; see Appendices A and B and Table 9.

[5] The relationship is straightforward. As the numbers of HQ and, proportionately, all analysts increase, the diversification advantage of the consensus forecast over HQ analysts' average forecast is diminished (since an additional HQ analyst provides a greater diversification benefit to the HQ analysts' average forecast than to the consensus forecast). This is analogous, for example, to the diversification benefit of adding an additional stock to a five-stock portfolio compared to a ten-stock portfolio.

[6] For example, if HQ analysts are defined as the top 50% of the most accurate analysts following a firm, HQ analysts' average forecast is more accurate than the consensus forecast in firms with four or more HQ analysts.

earnings surprise is expected to be positive) and shorts it when the HQ average forecast is lower than the consensus. This strategy yields a 0.26% abnormal return over the two-day announcement window, which potentially explains why investors tend to follow the consensus rather than HQ analysts' average forecast—the transaction and cognitive costs of identifying HQ analysts likely outweigh the benefits.[7]

While the comparison between the first moments of the HQ and all analysts' forecasts does not indicate a clear economic advantage of either, there can be more pronounced economic effects of differences between the second moments of HQ and all analysts' forecasts.[8] We start by considering HQ analysts' ability to predict return volatility. A positive relationship between forecast dispersion and return volatility following earnings announcements has received theoretical and empirical support in studies that assume analysts are homogeneous (Ajinkya and Gift, 1985; Daley, Senkow, and Vigeland, 1988; Abarbanell, Lanen, and Verrechia, 1995). However, to the best of our knowledge, analyst heterogeneity has not yet been considered in the literature in connection with forecast dispersion. Because HQ analysts are expected to be more skillful or have more access to private information, the difference between the dispersion of HQ analysts and that of all analysts (i.e., consensus dispersion) potentially provides information about uncertainty regarding a firm's future performance, which would manifest in future volatility. Indeed, we find that return volatility in the month following the earnings announcement month is positively associated with the difference in the dispersion of HQ analysts and the consensus dispersion.

[7] While statistically significant, its economic significance is questionable given the transaction costs associated with the trades, which are estimated to be between 0.4% and 0.8% for large firms and small firms, respectively (Novy-Marx and Velikov, 2016).

[8] There is limited evidence regarding analysts' ability to assess risk (e.g., Lui, Markov, and Tamayo, 2007; Joos, Piotroski, and Srinivasan, 2016), and even less on how one could aggregate analysts' mean forecast estimates to produce improved measures of uncertainty.

We assess the economic implications of this finding in two ways. First, we follow Eisdorfer, Sadka, and Zhdanov (2022) and construct at-the-money long and short straddle strategies depending on the sign of the relationship between the HQ dispersion and consensus dispersion. We form the straddle positions on the earnings announcement day and hold them to maturity. Such a strategy yields a 4.1% bi-monthly return (before transaction costs) or, equivalently, 24.5% on an annual basis. Second, we analyze the implications of the differences in dispersions for the post-earnings announcement drift (PEAD) anomaly. The PEAD anomaly is a natural way to examine the economic implications of differences in dispersion because both theoretical and empirical literature suggests that there is a positive relationship between uncertainty and price drift (Mendenhall, 2004; Zhang, 2006a; Francis et al., 2007; Hung, Li, and Wang, 2015). We conjecture a larger PEAD when uncertainty (proxied by HQ analysts' forecast dispersion) is relatively high. Indeed, we find a significantly high PEAD (9.4% after 11 months) when HQ analysts' forecast dispersion is higher relative to the consensus dispersion. Interestingly, when the dispersion of HQ analysts' forecasts is lower than the consensus dispersion, there is no PEAD. Overall, the results indicate that investors can earn economically significant returns if they utilize information about HQ analysts' forecast dispersion.

A potential criticism of this study is that our method relies on identifying a set of HQ analysts based on past forecast accuracy and ignores other analyst characteristics that may help improve the accuracy of the alternative forecast relative to the consensus forecast. We address this issue by considering a host of analyst characteristics and prediction models to define HQ analysts (see Appendix B). Further, it is important to note that advancing the literature on measuring analyst quality, determining the subset of analysts whose average forecast beats the consensus by the biggest margin, or analyzing the relative performance of high- and low-quality analysts is not among our study's objectives and, even if an optimal forecast could be found, this would not affect the study's conclusions. Rather, our method boils down to examining whether the market can utilize the difference between the average forecast (and dispersion) of a group of superior analysts and the consensus forecast (and dispersion) for an investment strategy around the earnings announcement date. Importantly, one can be agnostic about the magnitude of the forecast error of the HQ average because all what matters for investors' trading profits is the trade direction based on the sign of the difference between the HQ average forecast (dispersion) and the consensus forecast (dispersion). We elaborate on this issue in Appendix B.

Our findings contribute to the literature across several dimensions. First, this study is the first to show that the difference between the dispersion of HQ analysts and the dispersion of all analysts following the firm is informative about the future uncertainty of the firm's environment, consistent with the idea that high-quality analysts have superior cognitive skills and/or private information that is yet to manifest in equity prices. Further, we provide evidence that this improved identification of high- and low-uncertainty categories of firms and periods allows investors to earn economic profits through long-short straddle and PEAD strategies. This provides a path for future research on whether the difference in dispersion between high-quality analysts and dispersion of all analysts can be used in other applications related to stock returns, information asymmetry, and trading volume (e.g., Barron et al., 1998; Diether, Malloy, and Scherbina, 2002; Johnson, 2004; Barron, Stanford, and Yu, 2009).

Second, our focus on the economic merits of comparing HQ analysts' average forecasts to the widely and freely available consensus forecasts advances study of the stock price effects of quality variation among individual analysts, such as its impact on individual earnings forecast revisions (Stickel, 1992; Clement and Tse, 2003; Gleason and Lee, 2003; Chen, Francis, and Jiang, 2005).[9] Our finding that HQ analysts provide a more accurate average forecast than the consensus only for firms with large enough analyst followings incidentally contributes to the long literature on the statistical optimization of forecast accuracy (Clemen, 1989). While beating an equal-weighted measure is not easy because of the diversification effect, we show that using a simple definition of analyst quality one can beat the consensus in terms of forecast accuracy. Yet, more importantly, the improved accuracy turns out to be of little value for investors.

In summary, the major informational advantage of HQ forecasts is in determining the level of uncertainty, which entails profit opportunities for investors. Further, our results indicate that it behooves future researchers to demonstrate that any strategy of distinguishing among analysts is informationally superior to the freely available consensus.

2. Literature review

The literature related to this study consists of two separate strands. First, there are studies whose objective is to improve on the consensus forecast accuracy, as well as those whose objective is to use the dispersion (of all analysts) to assess uncertainty. Second, there are studies on the differences in individual analyst quality. Importantly, the latter studies do not consider whether considering such differences is better for investors than using the consensus output. This study bridges the two literatures and applies its insights to forecast errors and forecast dispersion.

Numerous studies in the last four decades propose different ways of aggregating forecasts to achieve greater forecast accuracy than that of the consensus forecast, including Makridakis and

[9] Practitioner white papers that consider strategies based on individual forecast revisions, e.g., Refinitiv StarMine, Bonne et al. (2007), also report sizeable profit gains after taking into account differences in analyst quality. While related practitioner papers on the consensus forecast show that their forecast measures can consistently beat the consensus by high margins (e.g., Vieira, Genin, and Birman, 2018), as far as we know, none of them provide direct evidence or report estimates of profits that can be generated around the earnings announcement date by predicting the actual earnings more accurately than the consensus does.

Winkler (1983); Conroy and Harris (1987); Brown, Gay, and Turac (2008); Vieira, Genin, and Birman (2018); and literature reviews in Clemen (1989) and Armstrong (2001). However, none of these studies examine whether and to what extent achieving such greater forecast accuracy has economic implications and benefits for investors relative to the widely available consensus forecast.

While there is significant literature on how one can beat the consensus, there is virtually no literature that tries to assess the extent to which differences in dispersion among analyst groups have economic implications. The literature on the overall dispersion (that is, the dispersion among all analysts who cover a firm) focuses on its determinants and consequences. Theory demonstrates that forecast dispersion is predominantly composed of uncertainty and information asymmetry (Barry and Jennings, 1992; Abarbanell, Lanen, and Verrecchia, 1995; Barron et al., 1998). [10] In that regard, earnings announcements, which provide the market with information, may lead to changes in the information asymmetry and uncertainty components (Lundholm, 1988; Kim and Verrecchia, 1994, 1997) and hence can have implications for future volatility. Recent empirical evidence suggests that analysts and investors indeed obtain new private information just prior to an earnings announcement (Lee, Mucklow, and Ready, 1993; Krinsky and Lee, 1996; Barron, Byard, and Kim, 2002; Barron, Harris, and Stanford, 2005; Frankel, Kothari, and Webber, 2006). Our evidence suggests that differences between the dispersion of HQ analysts and that of all analysts predict future volatility and consequently can be used to gain economic benefits. This

[10] The dispersion in forecasted earnings is an outcome of differences in opinions about future earnings. This difference in opinions arises due to several factors, such as uncertainty, information asymmetry, and cultural diversity (e.g., Merkley, Michaely, and Pacelli, 2021). Still, the most basic primitive of dispersion is uncertainty, since without uncertainty, there is no difference of opinion. At the extreme, if the future were known, no other factor would matter, and the forecast dispersion would be zero. In less extreme cases, other factors can also matter, but the literature agrees that a large component of dispersion is uncertainty (e.g., Imhoff and Lobo, 1992; Lang and Lundholm, 1996; Barron et al., 1998; Barron and Stuerke, 1998; Diether, Malloy, and Scherbina, 2002; Zhang, 2006).

implies that evaluating dispersion based on high-quality analysts' forecasts relative to the dispersion based on the forecasts of all analysts following a firm prior to its earnings announcement may be informative to investors.

A separate stream of studies proposes replacing analysts' consensus outputs altogether. Specifically, So (2013) and Wahlen and Wieland (2011) analyze whether investors would be better off if they forecast earnings using firms' financial fundamentals instead of using analysts' consensus outputs. So (2013) finds that investors can use information in firms' financial statements to obtain an earnings forecast superior to the consensus forecast, resulting in a trading strategy that generates long-term abnormal returns. Similarly, Wahlen and Wieland (2011) propose a method of predicting earnings increases that uses a set of financial ratios, leading to a trading strategy that produces greater returns than a strategy based on the consensus recommendation. Our approach differs from these two studies in that we do not dismiss the information in analysts' outputs. Analysts are sophisticated users of financial and non-financial information, and their outputs tend to incorporate information that is not yet reflected in firms' fundamentals.

Somewhat more distant from our study are studies that consider whether analysts' heterogeneity can lead to a differential market reaction to forecast revisions (Stickel, 1992; Clement and Tse, 2003; Gleason and Lee, 2003; Chen, Francis, and Jiang, 2005). These studies analyze *individual* forecast revisions. These studies thus differ from our research question, which focuses on the consensus information—they do not examine the economic benefit of using the *aggregated* outputs of the supposedly superior analysts relative to the consensus forecast or dispersion. As discussed in this study, the optimality of the consensus forecast relative to average forecasts of subsets of analysts is driven by the diversification benefit of averaging forecasts, which is obviously not applicable to individual forecast revisions. Similarly, while we analyze how forecast dispersion, which, by definition, aggregates forecasts, is related to return volatility, the notion of dispersion is not applicable to an individual forecast revision. As a result, neither our findings on the first and second moments of forecasts nor the findings of this literature can be generalized as extensions of each other.

3. Data and variables

We use the sample of annual earnings per share (EPS) estimates and earnings announcements in I/B/E/S during January 1992-December 2015 for companies with daily return data provided by the Center for Research in Security Prices (CRSP) database.[11] The earnings estimates and actual earnings are adjusted for splits using the daily cumulative adjustment factor from CRSP (Glushkov and Robinson, 2006).

Each year, we rank analysts following a firm based on the closest absolute forecast error, which is the absolute difference between an analyst's earnings forecast closest to the earnings announcement (made at least a day prior to the announcement day) and the actual earnings, divided by the share price at the beginning of the calendar year.[12] We define HQ analysts as those with an absolute forecast error below the median absolute forecast error for the firm-year.[13] The firm-year

[11] We focus on annual rather than quarterly earnings for two main reasons. First, fewer analysts provide quarterly forecasts than annual forecasts. Second, annual earnings announcements are typically more informative, including that they are more often supplemented with a conference call and followed by recommendation changes.

[12] This ranking has been used in Sinha, Brown, and Das (1997), Loh and Mian (2006), and Ertimur, Sunder, and Sunder (2007). In fact, ranking by forecast accuracy measured over more than one year could make the rankings less reliable because this would quickly reduce the pool of analysts available for ranking in each firm by requiring analysts to follow the firm for a number of years in a row. More generally, rankings based on past performance are common for analyst forecast persistence studies (Stickel, 1992) and in a number of other areas, such as in the mutual fund and pension fund performance forecasting (Hendricks, Patel, and Zeckhauser, 1993; Carhart, 1997; Tonks, 2005) and economic forecasting (Aiolfi and Timmerman, 2006) literatures. In Appendix A, we provide evidence that there is persistence in analysts being classified as HQ no matter how exactly we define HQ analysts, as long as our ranking relies on past accuracy. Appendix B conducts extensive sensitivity analyses for different proxies used to define HQ, such as assigning other weights to the non-HQ analysts' forecast or using other criteria based on analyst characteristics. The evidence suggests that our method of classifying HQ based on the previous year's accuracy outperforms other characteristics and methods for classifying HQ.

[13] Since the 50% cut-off in the HQ analysts' definition could be considered arbitrary, we analyze the sensitivity of our main findings to other cut-off for HQ analysts in Appendix C.

ranking measure is better suited for our study purpose compared to various "star" analyst, cross-firm classifications (e.g., the Wall Street Journal's "Best on the Street" or the Thomson Reuters StarMine's "Top Earnings Estimators") for three main reasons. First, our measure of analyst quality is persistent not only over time but also across the firms an analyst follows, as we report below. Second, our measure is simpler (it uses analysts' prior accuracy only in a given firm) and preserves the sample size better than star analyst classifications because many firms are not followed by star analysts.[14] Third, the results of our robustness tests indicate that our findings are not affected by different cutoff values in the definitions of HQ analysts.

From the initial sample, we generate 861,349 firm-year-analyst rankings based on the closest forecast error; the number drops to 804,003 observations once we require firms to have Compustat data. Next, to avoid small sample bias in our ranking when only a small number of analysts follow a firm, we exclude firm-years with fewer than four analysts following, reducing the sample to 750,295 observations.[15] In addition, analysts must appear in the data in two consecutive years for a given annual announcement to be ranked based on the first of the two years, reducing the sample to 484,125 observations.

In the firm-level regressions, we control for the following firm characteristics: size, the annual stock return, book-to-market ratio, number of analysts following, and leverage. *Firm size* is the market value of the firm's equity at the end of the month prior to the earnings announcement month. The *annual stock return* is measured based on monthly equity returns in the 12 months prior to the earnings announcement month. The *book-to-market ratio* is computed as stockholder equity minus preferred stock plus deferred taxes at the end of the fiscal year for which the earnings

[14] Specifically, the all-star ranking is at the industry/sector level, while we need a way to partition analysts at the firm level (to beat the consensus estimate). Furthermore, according to Clarke et al. (2007), the percentage of all-star analysts is small and has decreased from 12.4% in 1988 to 7.6% in 1999. Similarly, Guan et al. (2019) find that the percentage of all-star analysts is 8.5%.
[15] Loh and Mian (2006) exclude firm-years with fewer than five analysts, and Horton, Serafeim, and Wu (2017) and Clement and Tse (2003) exclude firms followed by fewer than three analysts.

are announced, divided by firm size. The *number of analysts* is the number of analysts who made at least one earnings forecast for the given announcement. *Leverage* is the book value of total liabilities divided by total assets at the end of the fiscal year for which the earnings are announced. Some of the regression models also control for analyst characteristics: (i) *overall tenure* is the number of years since the analyst first appeared in the I/B/E/S file; (ii) *firm-specific tenure* is the number of years since the analyst began covering the company in the I/B/E/S file; (iii) *brokerage house size* is the number of analysts employed by the brokerage firm; (iv) *firm coverage* is the number of firms covered by the analyst; and (v) *boldness* is the absolute value of the difference between the analyst's forecast and the consensus of forecasts in the previous 90 days (De Franco and Zhou, 2009).

4. Individual analysts

4.1. Sample description and tests of persistence in forecasting performance

Our preliminary analysis determines whether individual HQ analysts' forecast accuracy is persistent, which is a prerequisite for using groups of analysts ranked according to past accuracy. Table 1 reports the summary statistics for individual HQ and all analysts and examines the persistence of their forecast accuracy. Panel A shows that relative to all analysts, HQ analysts tend to be more experienced overall and within individual firms, are employed by larger brokerage firms, and cover a greater number of firms. However, HQ analysts are not bolder in their forecasts. The univariate analysis of the persistence of analysts' forecast accuracy shows that the absolute forecast errors of HQ analysts remain smaller than the errors of an average analyst one year after the analysts were ranked.

The regression analyses in Panel B of Table 1 examine the persistence in analysts' quality classification over time (columns 1 – 4) and across firms (columns 5 and 6). In the probit models (columns 1 and 2), the dependent variable is the HQ status in year t, which equals one if the analyst is categorized as HQ and zero otherwise. In columns (3) and (4), the dependent variable is the absolute forecast error, a continuous variable that allows us to control for firm fixed effects in the regression.[16] The main coefficient of interest is the HQ classification in year t-1. The results show that the coefficient of *HQ status (t-1)* is highly significant in all specifications, indicating that analysts' rankings and forecast accuracy are persistent in consecutive years.

We next examine whether forecasting performance is persistent not only over time but also across the firms that analysts follow. Affirmative findings will reinforce the argument that some analysts provide information that is superior to that of others; that is, the HQ designation is not a firm-specific attribute but is rather an analyst characteristic. We define an analyst's performance in other firms as HQ if the analyst is classified in the HQ category for the majority of the other firms that the analyst follows during the year (excluding this firm).[17] Columns (5) and (6) of Panel B test whether being ranked as an HQ analyst in other firms in year t-1 can predict an analyst's HQ classification in year t over and above his or her HQ classification in year t-1 in the same firm. That is, the independent variables of interest are the HQ status indicator of this analyst in the same firm in year t-1 and the *HQ status in other firms* indicator, which equals one if this analyst is also HQ in the majority of the other firms they followed in year t-1. We find that analysts who were HQ in the majority of other firms they followed in year t-1 are more likely to be HQ in a given firm in year t, and the coefficient of the firm-specific HQ designation in year t-1 remains positive and significant. Hence, the cross-firm findings suggest that analysts' forecasting performance

[16] Note that we do not control for analyst fixed effects or analyst-firm fixed effects in Panel B of Table 1 because its objective is to analyze the degree of persistence (autocorrelation) in forecast accuracy rather than persistence in forecast accuracy changes. Including analyst-level fixed effects also leads to a "forward looking" estimation of analyst quality as it eliminates the average accuracy differences among analysts across the entire sample (including future years).

[17] If the numbers of high- and low-quality rankings of the analyst in other firms are the same, this analyst-year-firm observation cannot be categorized as either high- or low-quality in the other firms and is thus excluded from this analysis (approximately 9% of the observations).

transcends across the stocks they follow and, further, that HQ analysts are indeed better than their peers in a persistent manner.

4.2. What does persistence in individual forecasting performance mean for the average of HQ forecast relative to the Consensus?

Our finding that individual HQ analysts tend to provide more accurate earnings forecasts than an average individual analyst following a firm may or may not extend to aggregate forecasts. Whether and how much more accurate the average of HQ analysts' forecasts is than the consensus may depend on the number of HQ analysts following a firm because of the diversification effect.

Table 2 investigates this issue and provides statistical tests comparing the absolute standardized unexpected earnings (SUE) of the consensus, which equals the difference between actual earnings and the consensus forecast, with the absolute SUE of HQ analysts, which is based on the difference between actual earnings and the HQ analysts' average forecast. Both SUE measures are normalized by the stock price at the beginning of the year. We find that as the number of HQ analysts following a firm increases, HQ analysts as a group become more accurate overall and relative to the consensus, consistent with the diversification effect. Further, when there are four or more HQ analysts, the absolute forecast error of HQ analysts is smaller than that of the consensus. Therefore, it is in these firms that investors seeking more accurate earnings forecasts should forego the consensus forecast in favor of the average of HQ analysts' forecasts. For the same reason, our sample in our main analyses, Tables 4-8, consists of firms with four or more HQ analysts when we examine whether the market can exploit analyst quality differences in the following analyses.[18]

In Table 3, we investigate the sample of firms with four or more HQ analysts in year t. First, we compare firm characteristics for this sample and firms with fewer than four HQ analysts in year t. As expected, firms with four or more HQ analysts are significantly larger. While these firms represent less than 30% of all firms, they account for 61% of the market value of the full sample. This finding is important because it suggests that our results cannot be attributed to small firms or firms followed by a few or inexperienced analysts.

In the lower part of Table 3, we provide the descriptive statistics of the main variable of interest for our focus firms. The absolute SUE of the consensus has a mean of 0.0042, while the SUE of HQ analysts is 0.0041 or 2.4% percent less. The difference in the median SUE is more substantial—close to 6%. Also of interest is the fact that the dispersion of HQ analyst forecasts is smaller by 13% (0.34/0.39-1) compared to the regular dispersion estimate. Jointly, it indicates that HQ analysts are not only more accurate but are also less dispersed in their forecast estimates.

5. Earnings announcements

The previous section demonstrates that relying on HQ analysts' earnings forecasts can produce an earnings forecast that is superior to the consensus forecast. The next question is whether the market is aware of this empirical regularity. To test this, we first examine whether investors are aware of quality differences among analysts. Specifically, we examine whether the immediate

[18] There is another technical condition for the sample. Because some analysts may stop covering a firm after year t-1 and new, unranked analysts may commence coverage, the number of HQ analysts in year t can become too small or too large (e.g., one of six analysts or five of six analysts are HQ), leading to small sample bias and a lack of robustness when the average accuracies of HQ and all analysts as groups are compared in the firm-level analysis. To mitigate this issue, we restrict the sample in all firm-level analyses (Table 2 and subsequent tables) to firm-years in which the proportion of HQ analysts in year t is not too different from the proportion in year t-1. Specifically, we require that the proportion of HQ analysts stays between 25% and 75% of all analysts providing forecasts for a given announcement.

reaction to an earnings surprise based on the mean forecast of HQ analysts is greater than the reaction to an earnings surprise based on the consensus forecast. Table 4 reports the regression results in which the dependent variable is the buy-and-hold cumulative abnormal return (BHAR) for the earnings announcement day and the following trading day based on the four-factor model (Fama and French, 1993; Carhart, 1997). The main variables of interest are the SUE calculated based on the consensus and SUE based on HQ analysts' average forecast. Although the two SUE variables are highly correlated at 0.92 (untabulated), the table shows that the reaction to the earnings surprise based on the consensus forecast is greater than the reaction to the SUE of HQ analysts, with a highly statistically significant difference between the coefficients (p-value<0.01) according to the chi-squared test in the full sample and in the sample of firms with four or more HQ analysts. When we include both variables in the regressions (columns 3 and 6), the results show that the market tends to react to surprise based on the consensus. Hence, the evidence indicates that the market does not sufficiently recognize differences in forecast accuracy among analysts because its reaction to the consensus forecast is significantly stronger, even for the sample where HQ analysts are, on average, more accurate than the consensus.

The finding that the market does not give more weight to HQ analysts' forecasts than the consensus may have meaningful economic implications. To gauge their magnitude, we first construct a simple measure of predicted earnings surprise based on the difference between HQ analysts' mean forecast and the consensus forecast, labeled *predicted surprise*. The intuition is to replace the actual earnings in the SUE formula with HQ analysts' average forecast,

$$Predicted\ surprise = \frac{Avg.Forecast^{HQ} - Avg.Forecast^{consensus}}{Price_{t-1}}, \qquad (1)$$

so that *predicted surprise* can be used to predict the SUE of the consensus one day before the announcement. Specifically, given that HQ analysts are more accurate than the consensus and the market overweighs the consensus forecast when it reacts to earnings surprise, one can expect positive (negative) abnormal returns to the earnings announcement when the mean forecast of HQ analysts is greater (smaller) than the consensus. A simple hedging strategy is to buy the stock on the day before the earnings announcement when the predicted surprise is positive and short it when it is negative.

We report the empirical results for the hedging strategy in Table 5. The out-of-sample analysis uses two variations of the signal based on *predicted surprise*: *positive predicted surprise* and *big positive predicted surprise* indicators. *Positive predicted surprise* equals one if *predicted surprise* is positive and zero otherwise. A potentially stronger signal, the *big positive predicted surprise* indicator equals one (zero) if the *predicted surprise* is above (below) the median of its positive (negative) values in the previous year and is set to missing otherwise. We regress the cumulative two-day announcement BHAR on each of these indicators and the control variables.

Consistent with our expectations, the coefficients of the *predicted surprise* indicators are positive and significant, suggesting that one can indeed measure the predicted earnings surprise ex-ante. The bottom line of the table reports the two-day abnormal returns of the hedging strategy that is long if the predicted surprise indicator in that column equals 1 and short if it equals 0. The returns to the hedging strategy are statistically significant and reach 0.26% for *big positive predicted surprise* in the sample of firms covered by four or more HQ analysts. Nonetheless, although the "4 or more HQ" sample consists of relatively large firms associated with low trading costs, the reported returns are only marginally significant and unlikely to exceed round-trip trading costs, which are estimated to be over 0.4% for large stocks (Novy-Marx and Velikov, 2016) and the cognitive and other costs associated with determining analysts' quality for each annual earnings announcement. Thus, the overall conclusion from Tables 4 and 5 is that the benefits of using HQ analysts' more accurate forecasts are not economically significant, potentially explaining why the market seemingly does not fully utilize the superior information embedded in HQ analysts' forecast.

6. Forecast dispersion and implications for option trading and PEAD

The persistence in analysts' forecasting performance over time and across stocks suggests that HQ analysts have superior ability. In this section, we empirically examine whether the dispersion of HQ analysts' forecasts (henceforth, "HQ dispersion") explains future uncertainty better than the dispersion of all analysts' forecasts ("consensus dispersion") and if so, whether and to what extent this has economic consequences.

6.1. Analysts' forecast dispersion and volatility

Analysts' forecast dispersion has been widely used as a proxy for uncertainty about a firm's future prospects. We conjecture that, just as the average forecast of HQ analysts provides superior earnings forecasts, their forecast dispersion better predicts future uncertainty. As mentioned above, the superiority of HQ analysts with reference to future uncertainty can be attributed to two factors: first, they may be better informed, and second, they have better cognitive skills, which allow them to process (public and private) information better. Hence, we start by examining whether disagreement about the firm's prospects among HQ analysts is a superior predictor of uncertainty surrounding the firm's future performance, measured by future return volatility, relative to the consensus dispersion.[19]

To test the relationship between dispersion and volatility, we regress stock return volatility during the month following the earnings announcement month on the standard deviation of analysts' forecasts before the earnings announcement. To avoid stale forecasts and make forecasts

[19] A possible caveat to the interpretation that high levels of HQ dispersion reflect high uncertainty is that, ceteris paribus, a high level of dispersion among HQ analysts may reflect their higher cultural diversity relative to that of other analysts. To mitigate this concern, we obtained the analysts' cultural diversity measures of Merkley, Michaely, and Pacelli (2020), which allowed us to test whether there is a difference in dispersion between HQ analysts and other analysts. The results indicate that there is no significant relationship between forecast dispersion and diversity. We thank Merkley, Michaley, and Pacelli for providing us with their data.

more comparable in terms of their proximity to the announcement, forecast dispersion is calculated using only forecasts made during the 60 days prior to the announcement.[20] The sample consists of firms in which HQ analysts' earnings forecasts are more accurate than the consensus, that is, firms followed by four or more HQ analysts.

In Table 6, apart from the firm characteristics used throughout the study (firm size, annual return, leverage, book-to-market, number of analysts), we also include three variables that are relevant for the analysis: lagged volatility, implied volatility derived from options, and management forecast range. *Implied volatility* is from Option Metrics and is based on traded options on the firm's shares during the month with the earnings announcement. Columns (1) – (4) include all traded options during the earnings announcement month as long as the options are traded no later than the date of the earnings announcement, as that is the time in which the dispersions variables are measured. Columns (5) and (6) include all options traded on the date of the earnings announcement (or the closest prior date); and columns (7) and (8) include only the most traded options on the date of the earnings announcement (or the closest prior date). *Management forecast range,* which also reflects uncertainty about future earnings, is the difference between the upper and lower EPS guidance values (of the earnings month) divided by the share price; it equals zero if there is no EPS guidance or if management provides an exact EPS estimate. The regressions include firm and year fixed effects, and standard errors are clustered at the firm level.

The results of Table 6, columns (1) and (2), show that neither the consensus dispersion nor HQ dispersion, respectively, are statistically significant in explaining next month's volatility,

[20] The length of the measurement period for forecast dispersion significantly varies in the literature. For instance, it can be one month (Krishnaswami and Subramaniam, 1999), four months (Zhang, 2006b), six months (Babenko, Tserlukevich, and Vedrashko, 2012), and up to one year since the previous earnings announcement (Diether, Malloy, and Scherbina, 2002). Our choice of 60 calendar days ensures that we use only the annual earnings forecasts made after the previous quarterly earnings announcement. Our results are unaffected by using a different period length.

while both lagged volatility and implied volatility are highly significant. These results suggest that the actual level of dispersion (of either HQ or all analysts) is not informative with respect to future volatility once we control for implied volatility and lagged volatility. However, according to columns (3) through (8), future volatility is positively related to the difference between HQ analysts' dispersion and consensus dispersion. Furthermore, the results are stronger when HQ dispersion is greater than the consensus dispersion. This can be seen from columns (4), (6), and (8), where we use the variable *Positive difference in dispersion*, which equals zero when the difference in dispersion is negative.

We interpret the results as follows. Implied volatility incorporates all publicly known information about future volatility, and hence the actual level of dispersion among analysts, whether HQ or all analysts, does not provide additional information about future volatility. However, when HQ analysts' uncertainty about future earnings is larger than that of all analysts, it implies that HQ analysts possess private information that is yet to manifest in the market. Hence, we observe a positive association between the difference in dispersion and future volatility. We next use straddle strategies and the PEAD to examine whether this relation has economic implications.

6.2. Straddle strategies

The finding that differences between the dispersion of HQ analysts and all analysts are predictive of future volatility, over and above the information provided by lagged volatility and implied volatility, suggests that one can profit from option trading by going long (short) a straddle position when HQ analysts' dispersion is higher (lower) than the dispersion of all analysts. A long (short) straddle strategy is formed by buying (selling) a call option and a put option that have the same strike price and expiration date. The position of a straddle is formed using at-the-money options (the straddle has a relatively low naïve price), and the strategy is expected to generate a trading profit if equity prices change significantly.

We follow the procedure of Eisdorfer, Sadka, and Zhdanov (2022) and create straddle strategies on the earnings announcement date and hold the positions until maturity. The sample includes options traded on the earnings announcement date whose expiration date is more than 14 days and less than 90 days from the earnings announcement date. The straddle strategies include call and put pairs whose underlying stock price (at the earnings announcement date) divided by the strike price is between 0.95 and 1.05. We show the results for earnings announcement cases where the HQ analysts' dispersion is greater (HQ>All) or smaller (HQ<All) than the consensus dispersion. The dollar payoff of a straddle is the payoff of the option positions at the expiration date minus the initial cost of the straddle. Short (long) maturity are straddles with expiration dates less (more) than 45 days after the earnings announcement date.

The results in Table 7 show that long (short) straddle pairs yield \$0.22 (-\$0.29) per one unit of the underlying, implying that a long/short strategy yields \$0.51 per one unit of the underlying (since options contracts are for 100 shares, this translates to a gain of \$51 per trade). The results seem to be larger for longer maturities than for shorter maturities, although the statistical significance of the results based on the smaller samples for the latter is lower. When we examine the percentage return (relative to the cost of the straddle), we find that on average, each straddle generates 4.09% (significant at the 10% level). Given an average maturity of 60 days for a straddle position, this translates to a 24.5% annual return. While this may appear to be sufficiently large, one should emphasize that transaction costs for options trading are expected to be relatively high, and furthermore, the low significance suggests that the return on the strategy is rather volatile.

6.3. The post-earnings announcement drift

Our results for the differences in the properties of forecast dispersion of all and HQ analysts' forecasts also have important implications for the PEAD anomaly. The motivation is based on the prior literature. First, Abarbanell, Lanen, and Verrecchia (1995) and Zhang (2006a) argue that investor underreaction to public information is more significant when uncertainty is high and find that analysts' forecast dispersion can predict a price drift following information releases. In related studies, Francis et al. (2007) find a positive relationship between uncertainty, which they measure with the unexplained portion of working capital accruals, and the PEAD. Hung, Li, and Wang (2015) find that exogenously reduced information uncertainty leads to a lower PEAD. Second, Mendenhall (2004) documents a relationship between return volatility and the PEAD. Combined, these studies posit a link between forecast dispersion and the PEAD. Note that these past studies focus on analyzing the PEAD in the cross-section or following an exogenous shock, while our contribution is in creating a dynamic measure at the firm level (dispersion of HQ minus dispersion of all analysts) that relies on how analysts' differences of opinion are related to their quality.

Specifically, our finding that the HQ dispersion in excess of the consensus dispersion predicts return volatility suggests the same relationship between the HQ dispersion in excess of the consensus dispersion and the PEAD. Benchmarking the HQ dispersion relative to the consensus dispersion in the *same firm* for the calendar-time PEAD analysis has the same objective as comparing the HQ and consensus dispersions for the same firm-announcement in the straddle trading analysis in Table 7, which is to control for cross-firm differences in the base level of uncertainty. Thus, a high HQ dispersion relative to the consensus dispersion represents a high uncertainty state for a firm, predicting high PEAD. In essence, our claim is that analysts' quality differences can be used to proxy for the PEAD because the HQ analysts' dispersion compared to the consensus dispersion seems to capture future changes in uncertainty well.

We examine whether investors can exploit the PEAD by utilizing the HQ dispersion relative to the consensus dispersion as a measure of uncertainty. The analysis compares the magnitudes of the PEAD when the HQ dispersion is high or low relative to the consensus dispersion, and the full-sample PEAD. We calculate the PEAD using the calendar-time approach. To be able to relate our results to the PEAD literature, we use the earnings surprise measure based on the consensus forecast to assign announcing stocks to a long or short portfolio at the end of each month, depending on whether the earnings surprise (SUE of consensus) is positive or negative, respectively.[21] The stocks are then held in the two portfolios for horizons from 1 to 11 months to avoid overlapping with the next annual earnings announcement. The monthly PEAD is the alpha from regressing the monthly value-weighted portfolio returns on the four Fama-French and Carhart factors.[22] The cumulative PEAD is the monthly alpha for a given horizon (the number of months the stock is held in the long or short calendar-time portfolio) multiplied by the horizon length in months. The resulting relationship between forecast dispersion and PEAD is presented in Figure 1 and Table 8.

Figure 1 reports the cumulative long PEAD portfolio return minus the short PEAD portfolio return for the sample of announcements with high uncertainty, defined as announcements for which the HQ dispersion is greater than the consensus dispersion; the low-uncertainty sample, in which the HQ dispersion is lower than the consensus dispersion; and the full-sample, that is, regular PEAD. The high-uncertainty PEAD is clearly greater than the full-sample PEAD. The low-uncertainty PEAD is below the full-sample PEAD and fluctuates around zero during most of the

[21] The PEAD phenomenon is known as an underreaction to an earnings surprise based on the consensus forecast. Since the market does not follow HQ analysts sufficiently and tends to react mainly to earnings surprises based on the consensus rather than HQ analysts' average forecast in Tables 3 and 4, it is unlikely that a drift exists following an earnings surprise measured relative to HQ analysts' average forecast. Indeed, if we measure the drift following earnings surprises calculated based on HQ analysts' SUE, the drift is not statistically significant.

[22] We obtain similar results using equal-weighted portfolios. Separately, we also note that because this sample consists of firms followed by four or more HQ analysts, the sample consists of relatively large firms, thereby reducing the likelihood that any findings may be attributed to illiquidity (Sadka, 2006).

post-announcement horizon. The largest cumulative long-short abnormal return of the PEAD strategy is 11.2%, reached eight months following high-uncertainty announcements.

Table 8 reports the monthly alphas on long, short, and long-minus-short strategies for the subsamples with high- and low-uncertainty announcements. We find that the low-uncertainty PEAD alpha (approximately 60% of the announcements) is not statistically significant except for the 11-month horizon. In contrast, when the HQ dispersion is greater than the consensus dispersion, that is, HQ analysts are more uncertain about the firm's future performance than all analysts following the firm (which consists of newcomers and all ranked analysts), the abnormal return of the long-minus-short portfolio is highly statistically significant in almost all horizons. To determine the additional return that investors can earn from this strategy over and above the return from the simple PEAD strategy, we regress the long-short post-announcement raw return for the high-uncertainty portfolio on the four Fama-French and Carhart factors and the long-short post-announcement raw return for the portfolio including all announcements (the regular PEAD). The resulting monthly alphas for different horizons are reported in column (7). Investors can earn up to 9.4% cumulative alpha (at the eight-month horizon) over and above the regular PEAD alpha. Overall, the PEAD is observed primarily during periods of high information uncertainty, as determined by the HQ dispersion relative to the consensus dispersion. This allows investors who exert effort and incur the transaction costs associated with processing information about analyst quality differences to achieve better investment performance than investors fixated on the consensus outputs.[23]

8. Conclusion

[23] Core et al. (2006) do not find corroborative evidence that insiders trade on the PEAD (SUE) anomaly. The results presented here are consistent with that and take the next step by suggesting that in most cases, the PEAD anomaly may not be sufficiently profitable, and investors (and insiders) would be better off trading on this anomaly only when the HQ analysts' forecast dispersion is relatively high.

This study analyzes whether investors can benefit from disregarding the consensus forecast and, instead, using a signal with greater information content based on analyst heterogeneity. Prior literature suggests that differentiating between high- and low-quality analysts' individual forecast revisions can be profitable. However, we find conclusions about the consensus forecast and aggregated analysts' forecasts that are quite different and have different implications for investors.

Although one can predict the earning surprise based on the difference between the HQ analysts' average forecast and the consensus forecast, investors react more strongly to deviations from the consensus earnings forecast than they do to deviations from the HQ analysts' forecasts. This provides modest announcement return predictability, but it is likely insufficient to be a worthwhile opportunity for investors or to indicate market inefficiency.

In contrast, differences between HQ analysts' forecast dispersion and the dispersion based on all analysts predicts the second moment of stock returns even after controlling for implied and lagged volatility. This allows potential investors to achieve modest returns in option straddle strategies and provides significantly greater returns based on the PEAD than if analyst quality differences are ignored.

To conclude, although analysts are heterogenous in terms of ability, and differences in ability and performance are persistent, the market seems to underappreciate the differences in ability. Nevertheless, the economic implications of these findings are quite small, suggesting that investors are better off focusing on the consensus and ignoring differences in quality among analysts. This conclusion is supported by the wisdom of the crowd theory, which implies that differences in ability are diversified away. Investors, however, can earn economic profit by trading based on differences in dispersion. This is because the second moment of the forecasts does not diversify away differences in ability.

Appendix A: Persistence in the definition of an HQ analyst

This appendix provides details on how several alternative ways of classifying analysts into the high-quality (HQ) group affect persistence in analyst forecasting performance. The ranking procedure sorts analysts in a given firm-year based on their absolute forecast error. In general, HQ analysts are those who are ranked in the top p percent of analysts, while low-quality (LQ) analysts are those in the bottom $(1-p)$ percent. If analysts' forecasting performance were uncorrelated across years, then the fractions of analysts who preserve their ranking in two consecutive years as HQ and LQ would be p^2 and $(1-p)^2$, respectively, or $p^2 + (1-p)^2$ of all analysts.

Figure A1 plots the fraction of analysts who retain their rankings in consecutive years and the expected fraction assuming no performance correlation across years. In this figure, to rank analysts up to the decile on the horizontal axis, the sample is restricted to firms that are followed by ten or more analysts. We find that with all cutoff values of p on the horizontal axis, the actual fraction of persistent forecasting performance is above the expected fraction and all differences are statistically significant (p-value<0.01). For example, when we classify the top 10% of analysts following a firm in a given year as HQ (p=10%) and the bottom 90% as LQ, the expected fraction given a random assignment is $0.9^2 + 0.1^2 = 0.82$. The figure shows that the actual fraction is greater than that at 0.843. The overall finding is that for all cutoff values, there is a sizeable persistent component; thus, for accuracy persistence, which exact cutoff value we choose to partition HQ and LQ analysts should make little difference.

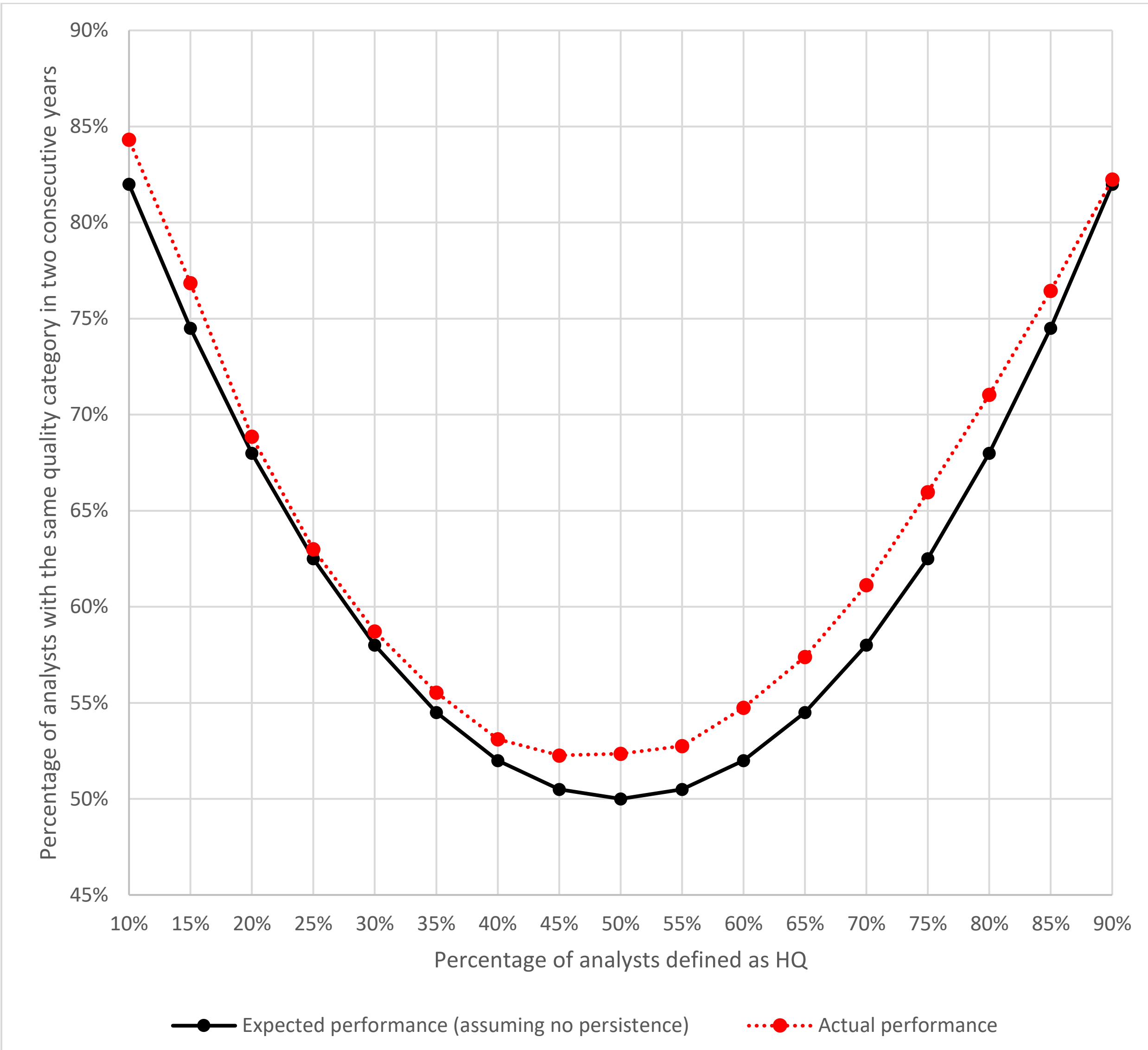

Figure A.1: Persistence in analysts' forecasting performance. This figure depicts how the fraction of analysts retaining their ranking of either high- or low-quality (HQ and LQ) in terms of forecast accuracy in two consecutive years depends on the cutoff percentile in the definition of HQ analysts. HQ analysts are those whose closest absolute forecast errors are less than the absolute forecast error at the cutoff percentile (the horizontal axis) of the distribution of forecast errors for a firm's annual earnings announcement in year t-1. The closest absolute forecast error is the absolute difference between an analyst's forecast estimate closest to the earnings announcement prior to the announcement day and the actual annual earnings, divided by the share price at the beginning of the calendar year. *Expected performance assuming no persistence* is the fraction of analysts who would have the same forecast performance category in two consecutive years if their performance were uncorrelated between years.

Appendix B: Other ways to define HQ analysts

In Table 2 in the main text, we compare the forecast error of the HQ group with the consensus forecast error for different numbers of HQ analysts. The purpose of this Appendix is to examine the relationship between the forecast errors of HQ analysts' average forecast and the consensus forecast for different definitions of HQ analysts.

Our alternative to the consensus forecast uses only analysts with forecast errors below the median, ignoring the forecasts of analysts that are "low quality" (LQ). One can consider that a more accurate measure could be created by assigning some positive weight (but less than 50%, since the LQ are less accurate than HQ by definition) to the LQ group. Clearly, the closer the weight assigned to the LQ group is to 50%, the closer the weighted measure would be to the consensus.

Indeed, Panel A1 of Table B1 follows the approach in Table 2 and reports accuracy improvements when we predict future earnings by assigning a 60% weight to the HQ group and a 40% weight to the LQ group. The accuracy improvement drops in absolute terms (as well as statistical significance) compared to the forecast with a 100% weight to the HQ group, but the improvement's signs remain the same. This result is expected because placing a 50% weight on the HQ analysts' average and 50% weight on the non-HQ analysts' average is by definition the consensus estimate.[24] To summarize, the analysis is not affected by the zero-weight assigned to non-HQ analysts' forecasts. That is, using any weighted scheme that assigns some weight to LQ analysts (but less than 50% because LQ are inferior) does not change the trading decision compared to the trading decision for the 100% HQ measure, since the trading decision is

[24] In Panel A1, we observe that the 60% HQ and 40% LQ weighted average forecast dominates the consensus when the firm is followed by five or more analysts and is inferior to the consensus when there are fewer than three HQ analysts. When we therefore use the sample of firms followed by five or more HQ analysts to see whether investors can profit from the superior forecast, the trading strategy yields a return of 0.2%, which is economically indistinguishable from the 0.26% return for the 100% HQ forecast reported in Table 5. We obtain similar results for other weights on LQ (from 5% to 45%, in increments of 5%).

determined by the sign of the difference between the alternative forecast and the consensus forecast.[25]

In Panel A2, analysts are ranked as HQ if they were HQ (above the median based on accuracy ranking in a firm) in the majority of the other firms they followed in the previous year. This alternative definition is motivated by the results in Table 1, Panel B, where we find that the HQ status in the other firms the analyst covered in year t-1 is positively associated with the analysts' HQ status in the given firm this year. The results show that the consensus beats this ranking in most cases. A possible explanation for why the cross-firm HQ measure is inferior to the firm-specific HQ designation is that according to the cross-firm measure, analysts can be considered HQ in a firm even if their forecast accuracy in that firm is not high enough to be ranked as HQ.

In Panel B of Table B1, we control for the "age" of the forecast used to identify HQ analysts. Specifically, analysts' rankings are based on the residual from a firm fixed-effects cross-sectional regression where the dependent variable is the absolute forecast error, and the independent variable is the number of days to the earnings announcement. The results are like those reported (i.e., four or more HQ analysts are needed to beat the consensus), but are somewhat less significant, suggesting that a residual model that considers forecast timing does not improve on our method of determining HQ analysts.

Panels C1–C4 of Table B1 show the results when we rank analysts based on a single characteristic (C1: brokerage size, C2: overall tenure, C3: coverage, C4: boldness) being above or below the median. As is evident, these definitions of HQ all fail to improve on the consensus, irrespective of the number of HQ analysts we consider. Finally, in C5, we rank analysts based on

[25] The trading decision is not affected by the weight on the HQ group because any linear combination of HQ and LQ group forecasts lies between the consensus and 100% HQ group forecast, since the weight on the HQ group must be greater than 50% due to its superior accuracy.

the predicted absolute forecast error from a regression of the absolute forecast error on the aforementioned characteristics. Specifically, we estimate the regression by year with firm-fixed effects and classify those analysts with below the median predicted forecast errors as HQ for the following year. This predictive model also fails and is not able to produce a group of HQ analysts whose average forecasts are more accurate than the consensus.

Overall, the analyses suggest that a simple definition of HQ analysts based on previous year accuracy outperforms, with respect to future accuracy, more elaborate definitions that include other characteristics and predictive models.

Table B1: The number of HQ analysts and the accuracy of their average forecast relative to the consensus forecast

The table compares the accuracy of HQ analysts' average forecast and the consensus forecast sorted by the number of HQ analysts following a firm in a given earnings announcement year. In Panel A1, analysts are ranked as HQ based on their previous year's absolute forecast errors, defined in Table 1. The panel reports the percentage accuracy improvement and t-statistics for differences between either the absolute SUE of a 100% HQ group average forecast or the absolute SUE of a weighted average forecast (with a 40% weight on the low-quality analysts (LQ) group and 60% weight on the HQ group average forecasts) and the absolute SUE of the consensus forecast. In Panel A2, we define HQ analysts in a given firm-year as those who were ranked as the top 50% based on accuracy in the majority of all other firms they followed in year t-1. In Panel B, the ranking is based on the residual from a cross-sectional regression in which the dependent variable is the absolute forecast error, and the independent variable is the number of days from the forecast to the earnings announcement date (with firm fixed-effects); the ranking is predictive as the regression is run at year t-1 and used for ranking in year t. Panels C1–C4 rank analysts based on various characteristics (brokerage size, overall tenure, number of firms covered, and boldness, respectively), and HQ analysts are defined as those with a characteristic value above the median characteristic of all analysts following the firm. Panel C5 ranks analysts based on the predicted absolute forecast error from a regression of the absolute forecast error on the aforementioned characteristics and firm-fixed effects at t-1. The SUE of consensus (SUE of HQ analysts) is standardized unexpected earnings, equal to the difference between the actual earnings and the average forecast provided by all analysts (HQ analysts) following a firm normalized by the stock price at the beginning of the year. Absolute SUE is the absolute value of SUE. *Accuracy improvement* is the percentage reduction from the absolute SUE of the consensus to the absolute SUE of HQ analysts. *t*-statistics in the last column are for the difference in the means test between the absolute SUEs of the consensus and HQ analysts. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A1. Ranking based on accuracy:
HQ defined as the top 50% based on accuracy in followed firm

Number of HQ analysts	Accuracy improvement: 0% weight to LQ analysts	Accuracy improvement: 40% weight to LQ analysts	t-statistics: Difference between Abs. SUE of consensus and Abs. SUE for 0% LQ	t-statistics: Difference between Abs. SUE of consensus and Abs. SUE for 40% LQ
1 or more	-3.24%	-1.03%	-8.63***	-5.26***
2 or more	-1.01%	-0.55%	-3.19***	-3.05***
3 or more	0.19%	-0.12%	0.54	-0.61
4 or more	0.19%	0.31%	2.99***	1.52
5 or more	1.32%	0.38%	3.51***	1.64*
6 or more	2.02%	0.46%	3.95***	1.73*
7 or more	2.39%	0.59%	4.47***	2.15**
8 or more	2.72%	0.72%	4.60***	2.51**
9 or more	2.60%	0.69%	3.91***	2.16**
10 or more	2.67%	0.60%	3.44***	1.75**

Panel A2: Ranking based on accuracy:
HQ in a given firm defined as being ranked as the top 50% based on accuracy in the majority of other firms followed

Number of HQ analysts	Absolute SUE of Consensus	Absolute SUE of HQ analysts	Accuracy improvement	t-statistics Abs. SUE difference
1 or more	0.00827	0.00878	-5.20%	-6.60***
2 or more	0.00662	0.00672	-3.46%	-1.79*
3 or more	0.00561	0.00567	-2.53%	-1.07
4 or more	0.00548	0.00556	-1.61%	-1.31
5 or more	0.00503	0.00504	-1.47%	-0.16
6 or more	0.00446	0.00446	-1.32%	-0.04
7 or more	0.00397	0.00384	-1.14%	1.72*
8 or more	0.00462	0.00451	-0.88%	0.91
9 or more	0.00459	0.00445	-0.63%	0.90
10 or more	0.00433	0.00392	0.34%	2.04**

Panel B. Ranking based on abnormal accuracy:
HQ defined as the top 50% - abnormal accuracy (annual rolling predictive)

Number of HQ analysts	Absolute SUE of consensus	Absolute SUE of HQ analysts	Accuracy improvement	t-statistics Abs. SUE difference
1 or more	0.00657	0.00686	-4.23%	-11.11***
2 or more	0.00614	0.00600	2.33%	-6.33***
3 or more	0.00535	0.00539	-0.74%	-2.25***
4 or more	0.00491	0.00489	0.41%	0.71
5 or more	0.00458	0.00454	0.88%	1.72*
6 or more	0.00453	0.00449	0.89%	1.88*
7 or more	0.00450	0.00443	1.58%	2.78***
8 or more	0.00463	0.00455	1.76%	2.49**
9 or more	0.00448	0.00441	1.59%	2.21**
10 or more	0.00448	0.00436	2.75%	2.68***

Panel C1. Ranking based on analyst characteristics:
HQ are above median brokerage size among analysts following the firm-year

Number of HQ analysts	Absolute SUE of consensus	Absolute SUE of HQ analysts	Accuracy improvement	t-statistics Abs. SUE difference
1 or more	0.00650	0.00693	-6.20%	-14.88***
2 or more	0.00570	0.00599	-4.84%	-12.95***
3 or more	0.00485	0.00504	-3.77%	-9.27***
4 or more	0.00421	0.00434	-3.00%	-6.26***
5 or more	0.00395	0.00403	-1.99%	-3.75***
6 or more	0.00371	0.00379	-2.11%	-3.29***
7 or more	0.00343	0.00349	-1.72%	-2.32**
8 or more	0.00338	0.00342	-1.17%	-1.55
9 or more	0.00322	0.00326	-1.23%	-1.18
10 or more	0.00301	0.00303	-0.66%	-0.40

Panel C2. Ranking based on analyst characteristics:
HQ are above median tenure as analysts among analysts following the firm-year

Number of HQ analysts	Absolute SUE of consensus	Absolute SUE of HQ analysts	Accuracy improvement	t-statistics Abs. SUE difference
1 or more	0.00651	0.00686	-5.10%	-12.25***
2 or more	0.00575	0.00597	-3.69%	-10.33***
3 or more	0.00499	0.00513	-2.73%	-7.25***
4 or more	0.00430	0.00439	-2.05%	-4.54***
5 or more	0.00405	0.00409	-0.98%	-2.21**
6 or more	0.00380	0.00382	-0.52%	-0.71
7 or more	0.00360	0.00360	0.00%	-0.24
8 or more	0.00353	0.00352	0.28%	0.18
9 or more	0.00327	0.00326	0.31%	0.24
10 or more	0.00316	0.00315	0.32%	0.23

Panel C3. Ranking based on analyst characteristics: HQ are below median coverage among analysts following the firm-year				
Number of HQ analysts	Absolute SUE of consensus	Absolute SUE of HQ analysts	Accuracy improvement	t-statistics Abs. SUE difference
1 or more	0.00651	0.00686	-5.10%	-12.64***
2 or more	0.00573	0.00595	-3.70%	-9.94***
3 or more	0.00491	0.00505	-2.77%	-7.18***
4 or more	0.00429	0.00437	-1.83%	-4.27***
5 or more	0.00402	0.00408	-1.47%	-3.15***
6 or more	0.00377	0.00381	-1.05%	-1.75*
7 or more	0.00349	0.00353	-1.13%	-1.71*
8 or more	0.00341	0.00345	-1.16%	-2.00**
9 or more	0.00333	0.00333	0.00%	-0.24
10 or more	0.00314	0.00317	-0.95%	-0.85

Panel C4: Ranking based on analyst characteristics: HQ are above median boldness among analysts following the firm-year				
Number of HQ analysts	Absolute SUE of consensus	Absolute SUE of HQ analysts	Accuracy improvement	t-statistics Abs. SUE difference
1 or more	0.00674	0.00711	-5.20%	-12.73***
2 or more	0.00586	0.00607	-3.46%	-9.53***
3 or more	0.00501	0.00514	-2.53%	-6.84***
4 or more	0.00429	0.00436	-1.61%	-3.99***
5 or more	0.00401	0.00407	-1.47%	-2.94***
6 or more	0.00374	0.00379	-1.32%	-2.65***
7 or more	0.00348	0.00352	-1.14%	-1.97**
8 or more	0.00337	0.00340	-0.88%	-1.37
9 or more	0.00313	0.00315	-0.63%	-0.84
10 or more	0.00297	0.00296	0.34%	0.32

Panel C5: Ranking based on all characteristics at t-1: HQ are above median predictive absolute FE based on regression on characteristics at t-1				
Number of HQ analysts	Absolute SUE of consensus	Absolute SUE of HQ analysts	Accuracy improvement	t-statistics Abs. SUE difference
1 or more	0.006558	0.006887	-4.78%	-13.25***
2 or more	0.005883	0.006084	-3.30%	-9.93***
3 or more	0.005094	0.005205	-2.12%	-5.91***
4 or more	0.004594	0.004672	-1.67%	-4.24***
5 or more	0.004231	0.004289	-1.35%	-3.04***
6 or more	0.003953	0.003991	-0.95%	-1.87*
7 or more	0.003853	0.003864	-0.27%	-0.50
8 or more	0.003628	0.003633	-0.14%	-0.24
9 or more	0.003387	0.003388	-0.04%	-0.09
10 or more	0.003269	0.003282	-0.41%	-0.57

Appendix C: Other cut-offs for defining an HQ analyst

The two alternative ranking procedures that we consider define HQ analysts as those in the top 70% and the top 30% of the forecast accuracy distribution. We repeat the study's key firm-level analyses using these two alternative quality definitions. The summary of the results is in Table C1. If HQ is defined as the top 70% of analysts, we find that HQ analysts provide more accurate forecasts than the consensus when the number of HQ analysts is greater than or equal to 3. This result is rather intuitive, as with a 70% fraction, the HQ average becomes more like the consensus, and hence, the diversification benefit of the consensus is smaller. In that subsample, the market reaction to the consensus forecast is greater than that for HQ analysts (the ERC difference is 0.023, significant at 10%). This suggests that, consistent with Tables 4 and 5, the market does not sufficiently recognize analyst quality differences and is fixated on the consensus forecast, resulting in predictable mispricing at the earnings announcement day. However, the returns of the hedging strategy in column (3) of Table 5 remain not economically significant at 0.18%.[26]

[26] These results are also robust to defining the top 30% of analysts as HQ. In this case, HQ forecasts become more accurate than the consensus only when the number of HQ analysts is greater than or equal to 6. The ERC of the consensus forecast exceeds the HQ analysts' ERC, but the difference is not statistically significant. However, this finding does not mean that the market sufficiently recognizes analyst quality differences because the market should react more strongly to the HQ average forecast than to the consensus forecast. The hedging strategy in column (3) of Table 5 remains not economically significant at 0.20%.

Table C1: Alternative definitions of HQ analysts and replication of results

Column (1) replicates Table 2 for the smallest number of HQ analysts following a firm, resulting in their average forecast having a lower absolute SUE than the consensus forecast. Column (2) replicates the difference between the earnings response coefficients (ERC) on HQ and consensus SUEs in columns (3) and (4) of Table 3, for the sample of firms followed by the number of HQ analysts provided in column (1) in this table. The statistical significance is based on the chi-squared test of coefficient differences as in the last line of Table 3. Column (3) corresponds to the long-short return in the last line of column (1) of Table 4. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Definition of HQ analysts	Table 2 Number of HQ analysts required for Abs. (SUE HQ) < Abs. (SUE consensus)	Table 4 Market reaction to earnings surprise	Table 5 Long-short strategy return on announcement day
	(1)	(2)	(3)
HQ: top 70%	3 or more	Consensus ERC greater by 0.023 (significant, *)	0.18%**
HQ: top 30%	6 or more	Consensus ERC greater by 0.069 (not significant)	0.20%*

References

Abarbanell, J.S., Lanen, W.N., Verrecchia, R.E., 1995. Analysts' forecasts as proxies for investor beliefs in empirical research. Journal of Accounting and Economics, 20(1), 31-60.

Aiolfi, M., Timmermann, A., 2006. Persistence in forecasting performance and conditional combination strategies. Journal of Econometrics, 135(1-2), 31-53.

Ajinkya, B.B., Gift, M.J., 1985. Dispersion of financial analysts' earnings forecasts and the (option model) implied standard deviations of stock returns. The Journal of Finance, 40(5), 1353-1365.

Armstrong, J.S., 2001. Combining forecasts. In Armstrong, J.S. ed., Principles of forecasting: A handbook for researchers and practitioners Vol. 30. Springer Science and Business Media, 417-439.

Babenko, I., Tserlukevich, Y., Vedrashko, A., 2012. The credibility of open market share repurchase signaling. Journal of Financial and Quantitative Analysis, 47(5), 1059-1088.

Barron, O.E., Byard, D. and Kim, O., 2002. Changes in analysts' information around earnings announcements. The Accounting Review, 77(4), pp.821-846.

Barron, O.E., Harris, D.G. and Stanford, M., 2005. Evidence that investors trade on private event-period information around earnings announcements. The Accounting Review, 80(2), pp.403-421.

Barry, C.B. and Jennings, R.H., 1992. Information and diversity of analyst opinion. Journal of Financial and Quantitative Analysis, 27(2), 169-183.

Barron, O.E., Kim, O., Lim, S.C., Stevens, D.E., 1998. Using analysts' forecasts to measure properties of analysts' information environment. Accounting Review, 421-433.

Barron, O.E., Stanford, M.H., Yu, Y., 2009. Further evidence on the relation between analysts' forecast dispersion and stock returns. Contemporary Accounting Research, 26(2), 329-357.

Barron, O.E. and Stuerke, P.S., 1998. Dispersion in analysts' earnings forecasts as a measure of uncertainty. Journal of Accounting, Auditing & Finance, 13(3), pp.245-270.

Bessembinder, H., 1999. Trade execution costs on Nasdaq and the NYSE: A post-reform comparison. Journal of Financial and Quantitative Analysis, 34(3), 387-407.

Bonne, G., Jahansouz, A., Lichtblau, D., and Malinak, S., 2007. StarMine Analyst Revisions Model (ARM): Global performance, StarMine white paper.

Brown, L.D., 2001. How important is past analyst forecast accuracy? Financial Analysts Journal, 57(6), 44-49.

Brown, L.D., Gay, G.D., Turac, M., 2008. Creating a "smart" conditional consensus forecast. Financial Analysts Journal, 64(6), 74-86.

Carhart, M.M., 1997. On persistence in mutual fund performance. The Journal of Finance, 52(1), 57-82.

Carpenter, J.N., Lynch, A.W., 1999. Survivorship bias and attrition effects in measures of performance persistence. Journal of Financial Economics, 54(3), 337-374.

Chen, Q., Francis, J., Jiang, W., 2005. Investor learning about analyst predictive ability. Journal of Accounting and Economics, 39(1), 3-24.

Clarke, J., Khorana, A., Patel, A., Rau, P.R., 2007. The impact of all-star analyst job changes on their coverage choices and investment banking deal flow. Journal of Financial Economics, 84(3), 713-737.

Clemen, R.T., 1989. Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, 5(4), 559-583.

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

Conroy, R., Harris, R., 1987. Consensus forecasts of corporate earnings: Analysts' forecasts and time series methods. Management Science, 33(6), 725-738.

Cooper, R.A., Day, T.E., Lewis, C.M., 2001. Following the leader: A. Journal of Financial Economics, 61(3), 383-416.

Core, J.E., Guay, W.R., Richardson, S.A., Verdi, R.S., 2006. Stock market anomalies: What can we learn from repurchases and insider trading? Review of Accounting Studies, 11(1), 49-70.

De Franco, G., Zhou, Y., 2009. The performance of analysts with a CFA designation: The role of human-capital and signaling theories. The Accounting Review, 84(2), 383-404.

Diether, K.B., Malloy, C.J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. The Journal of Finance, 57(5), 2113-2141.

Eisdorfer, A., Sadka, R., Zhdanov, A., 2022. Maturity driven mispricing of options. Journal of Financial and Quantitative Analysis, 57(2), 514-542.

Ertimur, Y., Sunder, J., Sunder, S.V., 2007. Measure for measure: The relation between forecast accuracy and recommendation profitability of analysts. Journal of Accounting Research, 45(3), 567-606.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3-56.

Francis, J., Lafond, R., Olsson, P., Schipper, K., 2007. Information uncertainty and post-earnings-announcement-drift. Journal of Business Finance and Accounting, 34(3-4), 403-433.

Frankel, R., Kothari, S.P. and Weber, J., 2006. Determinants of the informativeness of analyst research. Journal of Accounting and Economics, 41(1-2), pp.29-54.

Glushkov, D., Robinson, D., 2006. A note on IBES unadjusted data, WRDS documentation on IBES, https://wrds-web.wharton.upenn.edu/wrds/support/Data/_001Manuals%20and%20Overviews/_003I-B-E-S/index.cfm.

Guan, Y., Li, C., Lu, H., Wong, M.H.F., 2019. Regulations and brain drain: Evidence from Wall Street star analysts' career choices. Management Science, 65(12), 5766-5784.

Hendricks, D., Patel, J., Zeckhauser, R., 1993. Hot hands in mutual funds: Short-run persistence of relative performance, 1974-1988. The Journal of Finance, 48(1), 93-130.

Hirst, D.E., Hopkins, P.E., Wahlen, J.M., 2004. Fair values, income measurement, and bank analysts' risk and valuation judgments. The Accounting Review, 79(2), 453-472.

Hung, M., Li, X., Wang, S., 2015. Post-earnings-announcement drift in global markets: Evidence from an information shock. Review of Financial Studies, 28(4), 1242-1283.

Imhoff Jr, E.A. and Lobo, G.J., 1992. The effect of ex ante earnings uncertainty on earnings response coefficients. Accounting Review, pp.427-439.

Jacob, J., Lys, T.Z., Neale, M.A., 1999. Expertise in forecasting performance of security analysts. Journal of Accounting and Economics, 28(1), 51-82.

Johnson, T.C., 2004. Forecast dispersion and the cross-section of expected returns. The Journal of Finance, 59(5), 1957-1978.

Joos, P., Piotroski, J.D., Srinivasan, S., 2016. Can analysts assess fundamental risk and valuation uncertainty? An empirical analysis of scenario-based value estimates. Journal of Financial Economics, 121(3), 645-663.

Kim, O. and Verrecchia, R.E., 1994. Market liquidity and volume around earnings announcements. Journal of accounting and economics, 17(1-2), pp.41-67.

Kim, O. and Verrecchia, R.E., 1997. Pre-announcement and event-period private information. Journal of accounting and economics, 24(3), pp.395-419.

Krinsky, I. and Lee, J., 1996. Earnings announcements and the components of the bid-ask spread. The Journal of Finance, 51(4), 1523-1535.

Krishnaswami, S., Subramaniam, V., 1999. Information asymmetry, valuation, and the corporate spin-off decision. Journal of Financial Economics, 53(1), 73-112.

Lang, M.H. and Lundholm, R.J., 1996. Corporate disclosure policy and analyst behavior. Accounting review, pp.467-492.

Lee, C.M.C., Mucklow, B., Ready, M.J., 1993. Spreads, depths, and the impact of earnings information: An intraday analysis. Review of Financial Studies, 6(2), 345-374.

Lin, H.W., McNichols, M.F., 1998. Underwriting relationships, analysts' earnings forecasts and investment recommendations. Journal of Accounting and Economics, 25(1), 101-127.

Loh, R.K., Mian, G.M., 2006. Do accurate earnings forecasts facilitate superior investment recommendations? Journal of Financial Economics, 80(2), 455-483.

Lundholm, R.J., 1988. Price-signal relations in the presence of correlated public and private information. Journal of Accounting Research, pp.107-118.

Lui, D., Markov, S., Tamayo, A., 2007. What makes a stock risky? Evidence from sell-side analysts' risk ratings. Journal of Accounting Research, 45(3), 629-665.

Maines, L.A., McDaniel, L.S., Harris, M.S., 1997. Implications of proposed segment reporting standards for financial analysts' investment judgements. Journal of Accounting Research, 35, 1-24.

Makridakis, S., Winkler, R.L., 1983. Averages of forecasts: Some empirical results. Management Science, 29(9), 987-996.

Malloy, C.J., 2005. The geography of equity analysis. Journal of Finance, 60(2), 719-755.

McInish, T.H., Wood, R.A., 1992. An analysis of intraday patterns in bid/ask spreads for NYSE stocks. Journal of Finance, 47(2), 753-764.

Mendenhall, R.R., 2004. Arbitrage risk and post-earnings-announcement drift. Journal of Business, 77(4), 875-894.

Merkley, K., Michaely, R., Pacelli, J., 2020. Cultural diversity on Wall Street: Evidence from consensus earnings forecasts. Journal of Accounting and Economics, 70(1), 101330.

Mikhail, M.B., Walther, B.R., Willis, R.H., 1997. Do security analysts improve their performance with experience? Journal of Accounting Research, 35, 131-157.

Novy-Marx, R., Velikov, M., 2016. A taxonomy of anomalies and their trading costs. Review of Financial Studies, 29(1), 104-147.

Rubin, A., 2007. Ownership level, ownership concentration and liquidity. Journal of Financial Markets, 10(3), 219-248.

Sadka, R., 2006. Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk. Journal of Financial Economics, 80(2), 309-349.

Sinha, P., Brown, L.D., Das, S., 1997. A re-examination of financial analysts' differential earnings forecast accuracy. Contemporary Accounting Research, 14(1), 1-42.

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

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

Tonks, I., 2005. Performance persistence of pension-fund managers. The Journal of Business, 78(5), 1917-1942.

Vieira, M., Genin, H., and Birman, S., 2018. An update on the performance of the StarMine SmartEstimate from Refinitiv and Predicted Surprise. StarMine white paper.

Wahlen, J.M., Wieland, M.M., 2011. Can financial statement analysis beat consensus analysts' recommendations? Review of Accounting Studies, 16(1), 89-115.

Zhang, X.F., 2006a. Information uncertainty and stock returns. Journal of Finance, 61(1), 105-137.

Zhang, X.F., 2006b. Information uncertainty and analyst forecast behavior. Contemporary Accounting Research, 23(2), 565-590.

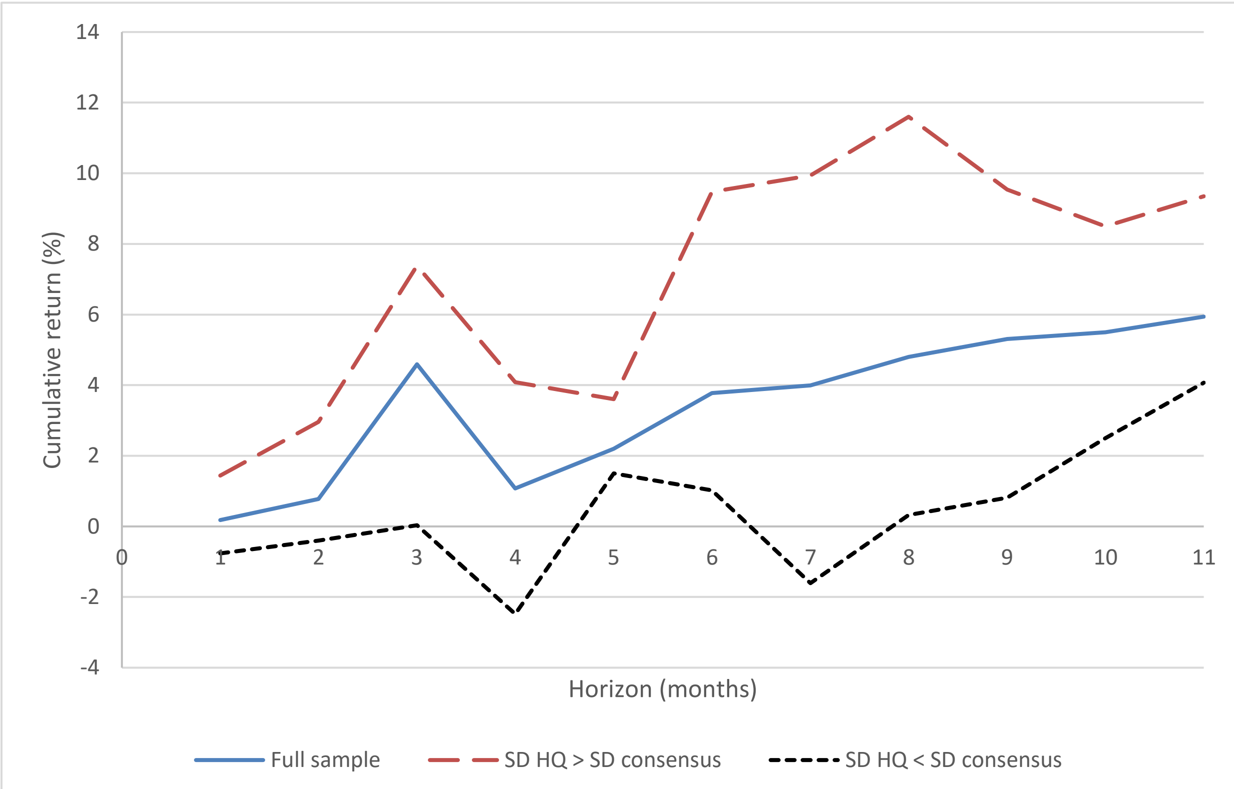

Figure 1: Cumulative post-earnings announcement drifts and analysts' uncertainty
This figure shows the cumulative returns for 1- to 11-month horizons following earnings announcements. Each month, stocks enter a calendar-time long (short) portfolio, depending on whether their earnings surprise is positive (negative), where earnings surprise is defined based on the consensus estimate in Table 2. The horizontal axis is the drift's horizon, which is the number of months a stock is held in the calendar-time portfolio. The monthly long-minus-short value-weighted portfolio return is regressed on the four Fama-French and Carhart factors. The cumulative drift for a given drift horizon on the vertical axis is calculated as the regression intercept (monthly alpha) multiplied by the portfolio's horizon length in months. The graphs are shown for the full sample and two subsamples of firms in which the standard deviation of HQ analysts' forecasts (*SD HQ*) is either greater or smaller than that of all analysts' forecasts (*SD consensus*) for the firm-year. Analysts are ranked as HQ based on their previous year's absolute forecast errors, defined in Table 1.

Table 1: Individual analyst analysis: characteristics and persistence of forecast accuracy

Panel A conducts univariate summary statistics for high-quality (HQ) analysts. HQ analysts are those whose closest absolute forecast errors are below (at or above) the median closest absolute forecast error for the firm's earnings announcement. The closest *absolute forecast error* for an analyst is the absolute difference between the analyst's forecast closest to the earnings announcement prior to the announcement day and the actual annual earnings, divided by the share price at the beginning of the calendar year. *Overall tenure* is the number of years since the analyst first appeared in the I/B/E/S file. *Firm-specific tenure* is the number of years since the analyst began covering the specific firm in the I/B/E/S file. *Brokerage house size* is the number of analysts in the analyst's brokerage house. *Firms covered* is the number of firms covered by the analyst. *Boldness* is the absolute difference between the analyst's forecast and the consensus forecast in the previous 90 days. Panel B reports probit models predicting the *HQ status* indicator, which equals one if the analyst is ranked as HQ in a given firm in the current year and zero otherwise (columns (1), (2), (5), and (6)) and the regressions for the analyst's closest absolute forecast error in a given firm in the current year in columns (3) and (4). *HQ status in other firms* equals one (zero) if the analyst is HQ in the majority of the other firms that the analyst follows during the year; analysts who have an equal number of HQ and non-HQ rankings in other firms are excluded. *Firm size* is the log of the firm's market value of equity, equal to the stock price times the number of shares outstanding at the end of the month prior to the annual earnings announcement. *Annual return* is the annual return of the firm's equity over the 12 months prior to the earnings announcement month. *Leverage* is the book value of total liabilities divided by the book value of total assets, and *Book-to-market* is the book value of common equity divided by the market value of equity at the end of the fiscal year. *Number of analysts* is the number of analysts following the firm. All independent variables are measured prior to the announcement date. The probit coefficients are reported as marginal probability effects. All models include the intercept. Robust standard errors are clustered by firm. *z*-statistics for probit models and *t*-statistics elsewhere are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

.

Panel A: HQ and all analyst characteristics

Analyst or announcement characteristic	HQ analysts	All analysts	Difference (t-statistic)
Overall tenure	7.07	7.04	0.03*** (3.39)
Firm-specific tenure	3.04	3.00	0.04*** (6.14)
Brokerage house size	65.75	64.35	1.41*** (13.61)
Firms covered	17.61	17.58	0.03* (1.30)
Boldness	7.74E-7	7.64E-7	1.0E-8 (0.13)
Absolute forecast error	0.0080	0.0084	-0.0035*** (-9.03)

Panel B: Persistence of analysts' forecasting performance

	HQ status		Absolute forecast error		HQ status	
	(1)	(2)	(3)	(4)	(5)	(6)
HQ status (year t-1)	0.0417***	0.0410***	-0.0007***	-0.0007***	0.0280***	0.0280***
	(25.66)	(25.18)	(-16.24)	(-15.91)	(15.55)	(15.44)
HQ status in other firms (year t-1)					0.0598***	0.0578***
					(32.45)	(31.12)
Firm size	0.0034***	0.0012***	-0.0061***	-0.0061***	0.0034***	0.0013***
	(10.54)	(3.30)	(-24.96)	(-24.73)	(8.37)	(3.00)
Annual return	-0.0003	0.0008	-0.0009***	-0.0008***	-0.0018**	-0.0008**
	(-0.44)	(0.93)	(-5.67)	(-5.54)	(-2.12)	(-0.96)
Leverage	0.0008	0.0000	0.0056***	0.0056***	0.00016	0.00004
	(0.51)	(0.47)	(6.52)	(6.53)	(0.08)	(0.02)
Book-to-market	0.00001	0.00004	0.00002	0.00002	0.0002***	0.0001**
	(1.53)	(0.79)	(1.45)	(1.44)	(3.14)	(2.29)
Number of analysts	0.0006***	0.0007***	0.0002***	0.0002***	0.0007***	0.0008***
	(12.33)	(13.87)	(10.75)	(10.56)	(10.09)	(11.25)
Overall tenure		0.0007***		-0.00004***		0.0002
		(4.32)		(-6.93)		(4.63)
Firm-specific tenure		0.0023***		0.00001		0.0021***
		(9.60)		(1.21)		(8.12)
Brokerage house size		0.0001***		-0.000001		0.0001***
		(10.56)		(-0.39)		(8.95)
Firm coverage		-0.0008***		0.00003***		-0.0007***
		(-12.37)		(8.51)		(-10.64)
Boldness		-6.708***		1.976***		-9.41***
		(-10.82)		(7.70)		(-10.66)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects			Yes	Yes		
Observations (analyst-years)	484,125	478,086	484,125	478,086	440,041	395,069
Adj. R-squared			0.344	0.334		

Table 2: The number of HQ analysts and accuracy of their average forecast relative to the consensus forecast

Panel A compares the accuracy of the HQ analysts' average forecast and the consensus forecast sorted by the number of HQ analysts following a firm in a given earnings announcement year. Analysts are ranked as HQ based on their previous year's absolute forecast errors, defined in Table 1. The consensus SUE (SUE of HQ analysts) is standardized unexpected earnings, equal to the difference between the actual earnings and average forecast provided by all analysts (HQ analysts) following a firm normalized by the stock price at the beginning of the year. Absolute SUE is the absolute value of SUE. *Accuracy improvement* is the percentage reduction from the absolute SUE of the consensus to the absolute SUE of the HQ analysts. *t*-statistics in the last column are for the difference in the means test between the absolute SUEs of the consensus and HQ analysts. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Number of HQ analysts	Absolute SUE of consensus	Absolute SUE of HQ analysts	Accuracy improvement	t-statistics Abs. SUE difference
1 or more	0.00656	0.00678	-3.31%	-8.63***
2 or more	0.00589	0.00595	-1.08%	-3.19***
3 or more	0.00514	0.00513	0.19%	0.54
4 or more	0.00461	0.00455	1.17%	2.99***
5 or more	0.00422	0.00415	1.52%	3.51***
6 or more	0.00404	0.00396	1.96%	3.95***
7 or more	0.00386	0.00377	2.35%	4.47***
8 or more	0.00377	0.00367	2.69%	4.60***
9 or more	0.00355	0.00346	2.61%	3.92***
10 or more	0.00346	0.00337	2.59%	3.44***

Table 3: Descriptive statistics

The table reports sample characteristics for variables defined in Table 1 and Table 2 in two samples sorted based on whether the firm is followed by fewer than four HQ analysts or four or more HQ analysts in a given announcement year.

	Number of HQ analysts		Difference
	Fewer than 4	4 or more	(t-statistics)
Mean number of firms per year	1466	589	
Firm characteristics:			
Firm size ($M)	2,675	10,721	8,046*** (44.44)
Number of analysts	4.5	13.1	8.6*** (230)
Book-to-market	1.60	0.80	-0.80*** (-4.69)
Leverage	0.54	0.57	0.03** (12.42)

Main variables of interest (firms with 4 or more HQ analysts)

	Mean	SD	P25	P50	P75
Absolute SUE of the consensus	0.0042	0.0997	0.00050	0.00138	0.00363
Absolute SUE of HQ analysts	0.0041	0.0100	0.00046	0.00130	0.00348
Dispersion of all analysts	0.0039	0.0070	0.00067	0.00157	0.00394
Dispersion of HQ analysts	0.0034	0.0068	0.00052	0.00121	0.00309

Table 4: Immediate reaction to earnings announcements

This table reports the earnings response coefficients for measures of earnings surprise (SUE) based on the forecasts of all analysts following a firm (consensus) and on the forecasts of high-quality (HQ) analysts. Analysts are ranked as HQ based on their previous year's absolute forecast errors as defined in Table 1. The dependent variable is the buy-and-hold abnormal return (based on the four-factor Fama-French and Carhart model) for the earnings announcement day and the following trading day. The SUE of Consensus and SUE of HQ analysts are defined in Table 2. All other variables are defined in Table 1. Columns (1)–(3) use the entire sample of earnings announcements, and columns (4)–(6) use the sample of earnings announcements of firms followed by at least four HQ analysts. All independent variables other than SUE are measured prior to the announcement date. The intercept and year fixed effects are included in all regressions. Robust standard errors are clustered by firm. *t*-statistics are provided in parentheses. The last two lines report the p-values for the chi-squared tests of the equality of the coefficients on the two SUE measures. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Full Sample			4 or more HQ analysts		
	(1)	(2)	(3)	(4)	(5)	(6)
SUE of Consensus	0.7245***		0.6925***	0.7526***		0.7614**
	(13.62)		(7.04)	(5.20)		(2.39)
SUE of HQ analysts		0.6211***	0.0318		0.6927***	-0.0087
		(12.93)	(0.37)		(5.12)	(-0.03)
Firm size	-0.0003	-0.0002	-0.0003	-0.0007	-0.0007	-0.0007
	(-0.71)	(-0.55)	(-0.71)	(-1.19)	(-1.15)	(-1.20)
Annual return	-0.0006	-0.0006	-0.0006	0.0007	0.0007	0.0007
	(-0.98)	(-0.89)	(-0.98)	(0.68)	(0.72)	(0.68)
Leverage	0.0059***	0.0057***	0.0059***	0.0030	0.0029	0.0030
	(3.73)	(3.55)	(3.73)	(1.24)	(1.19)	(1.24)
Book-to-market	0.00002	0.00001	0.00002	-0.0002***	-0.0002***	-0.0002***
	(0.38)	(0.29)	(0.38)	(-9.20)	(-7.67)	(-8.80)
Number of analysts	0.00002	0.00002	0.00002	-0.00004	-0.00003	-0.00004
	(0.25)	(0.28)	(0.25)	(-0.42)	(-0.40)	(-0.42)
Observations	44,709	44,709	44,709	20,221	20,221	20,221
Adj. R-squared	0.015	0.013	0.015	0.011	0.010	0.011
p-value: SUE of HQ analysts vs. SUE of consensus		0.000			0.009	

Table 5: Predicting market reaction on the earnings announcement day

The dependent variable is the buy-and-hold abnormal return (based on the four-factor Fama-French and Carhart model) for the earnings announcement day and the following trading day. Analysts are ranked as HQ based on their previous year's absolute forecast errors, defined in Table 1. *Predicted surprise* is equal to HQ analysts' average forecast minus the consensus, normalized by the stock price at the beginning of the year. *Positive predicted surprise* is an indicator that equals one if *predicted surprise* is positive and zero if it is negative. *Big positive predicted surprise* equals one if *predicted surprise* is greater than the median of positive values of *predicted surprise* and zero if *predicted surprise* is smaller than the median of the negative values of *predicted surprise* in year t-1. Columns (1)–(2) use the entire sample of earnings announcements, and columns (3)–(4) use the sample of earnings announcements of firms followed by at least four HQ analysts. All independent variables are measured prior to the announcement date, and the regressions include the intercept and year fixed effects. Robust standard errors are clustered by firm. *t*-statistics are provided in parentheses. The last line of the table provides the two-day holding returns of a trading strategy that is long if the predicted surprise indicator variable in that column equals 1 and short if it equals 0. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Full Sample		4 or more HQ analysts	
	(1)	(2)	(3)	(4)
Positive predicted surprise	0.0016**		0.0017*	
	(2.02)		(1.83)	
Big positive predicted surprise		0.0007*		0.0009
		(1.79)		(1.63)
Firm size	0.0002	-0.00005	-0.0002	-0.0007
	(0.52)	(-0.09)	(-0.35)	(-0.85)
Annual return	0.0005	0.00002	0.0003	0.0002
	(0.72)	(0.02)	(0.40)	(0.14)
Leverage	0.0039**	0.0064**	0.0017	0.0041***
	(2.33)	(2.54)	(0.79)	(1.23)
Book-to-market	0.00003	0.00001	-0.000006	0.00001
	(0.09)	(0.28)	(-0.11)	(0.05)
Number of analysts	0.00007	-0.00002	-0.00001	-0.0001
	(0.06)	(-0.19)	(-0.18)	(-0.13)
Observations	43,655	20,398	30,411	12,458
Two-day long-short strategy returns (%)	0.16**	0.22*	0.16*	0.26*
	(2.03)	(1.78)	(1.78)	(1.64)

Table 6: Predicting return volatility with forecast dispersion

The dependent variable is the standard deviation of a firm's daily returns in the month following the annual earnings announcement month. *Consensus dispersion* and *HQ analysts' dispersion* are the standard deviations of the consensus and HQ forecasts, respectively, normalized by the stock price, which use each analyst's forecast closest to the earnings announcement and issued during the 60 days prior to the announcement. *Difference in dispersion* is HQ analysts' dispersion minus HQ analysts' dispersion. *Positive difference in dispersion* is the same as *Difference in dispersion* if HQ analysts' dispersion is greater than the consensus dispersion and equals zero if HQ analysts' dispersion is smaller than the consensus dispersion. *Implied volatility* is from Option Metrics and includes traded options in the stock during the earnings announcement month, with sample specifications based on different types of these options provided in the top line. *Management forecast range* equals the difference between the upper and lower EPS guidance values (at the earnings announcement date) divided by the share price; it equals zero if there is no EPS guidance or only one mean value of guidance is provided. Analysts are ranked as HQ based on their previous year's absolute forecast errors, defined in Table 1. Other controls included in all regressions are *firm size*, *annual return*, *leverage*, *book-to-market*, and *number of analysts*, defined in Table 1 and measured prior to the announcement date. All regressions also include year and firm fixed effects. Robust standard errors are clustered by firm. *t*-statistics are provided in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	All options traded during the month up to (and inclusive) of EA date				Options traded on the date closest to EA date		Most traded calls and puts closest to EA date	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Consensus dispersion	-0.0436							
	(-0.57)							
HQ analysts' dispersion		0.0590						
		(0.80)						
Difference in dispersion			0.2084*		0.2827**		0.1121	
			(1.78)		(2.48)		(1.04)	
Positive difference in dispersion				0.8072***		0.8756**		0.8143**
				(2.76)		(2.40)		(2.14)
Lagged return volatility	0.5867***	0.5814***	0.5816***	0.5712***	0.5780***	0.5674***	0.4624***	0.4556***
	(11.43)	(11.26)	(11.44)	(11.35)	(11.47)	(11.63)	(10.91)	(10.99)
Implied volatility	0.0036***	0.0036***	0.0036***	0.0036***	0.0048***	0.0047***	0.0117***	0.0115***
	(5.68)	(5.65)	(5.81)	(6.04)	(6.28)	(6.28)	(6.84)	(6.76)
Management forecast range	0.0102	0.0205	0.0341	0.0414	0.1787	0.1957	0.0049	0.0159
	(0.10)	(0.19)	(0.33)	(0.38)	(1.13)	(1.19)	(0.05)	(0.15)
Observations	822,257	822,257	822,257	822,257	75,280	75,280	4,273	4,273
Adj. R-squared	0.860	0.860	0.861	0.861	0.854	0.853	0.795	0.798

Table 7: Forecast dispersion differences and options market strategies

The table provides the dollar payoff and return from straddle strategies, Call and Put option pairs, formed at the earnings announcement day and held until expiration. The sample includes traded options traded on the earnings announcement day and whose expiration is more than 14 days and less than 90 days from the earnings announcement. The straddle strategies include Call and Put pairs for which the ratio of the stock price (at the earnings announcement date) to the strike price is between 0.95 and 1.05. Earning announcements are divided into two types depending on whether the HQ analysts' dispersion is greater or smaller than the consensus dispersion. These are denoted in the table as HQ>All and HQ≤All, respectively. The dollar payoff of a straddle is the payoff of the option positions at the expiration date minus the initial cost of the straddle. Short (long) maturity are straddles whose expiration is less (more) than 45 days away. t-statistics are provided in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Obs.	Payoff (per unit of underlying)			Return on initial cost		
		HQ>All	HQ≤All	Difference	HQ>All	HQ≤All	Difference
All straddle pairs	4,402	0.217	-0.290**	0.507**	-0.027	-0.068***	0.041*
		(1.15)	(-2.49)	(2.42)	(-1.37)	(-4.52)	(1.65)
Short maturity	2,639	-0.236	-0.571***	0.335*	-0.060**	-0.095***	0.035
		(-1.39)	(-5.17)	(1.73)	(-2.47)	(-5.12)	(1.16)
Long maturity	1,763	0.937**	0.112	0.822*	0.024	-0.030	0.054
		(2.30)	(0.49)	(1.88)	(0.70)	(-1.18)	(1.27)

Table 8: Post-earnings announcement drift and analysts' uncertainty

The table reports the monthly abnormal returns for 1- to 11-month horizons following annual earnings announcements. Announcements are divided into two subsamples in which the standard deviation of HQ analysts' forecast errors is greater (the *high-uncertainty* sample) or smaller (the *low-uncertainty* sample) than the standard deviation of forecast errors of all analysts following a firm. Analysts are ranked as HQ based on their previous year's absolute forecast errors, defined in Table 1. A stock is assigned to the long or short portfolio depending on whether its earnings surprise is positive or negative, respectively, where earnings surprise is defined based on the consensus estimate in Table 2. A stock enters a portfolio at the beginning of the month following the month of the announcement and is held for the length of the horizon. In columns (1)–(6), the monthly value-weighted portfolio returns are regressed on the four Fama-French and Carhart factors to obtain the reported calendar-time drifts, which are the intercepts of the regressions (monthly alphas). For the monthly calendar-time alphas in column (7), the high-uncertainty long-short portfolio's raw return is regressed on the full-sample long-short portfolio's raw return and the four Fama-French and Carhart factors. *, **, and *** represent the 10%, 5%, and 1%, significance levels based on the regression *t*-statistics, respectively.

PEAD horizon (months)	High uncertainty			Low uncertainty			High uncertainty relative to full sample
	Long	Short	Long-Short	Long	Short	Long-Short	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1.09*	-0.35	1.44*	-0.46	0.30	-0.77	1.19**
2	1.36***	-0.12	1.48**	0.13	0.33	-0.20	1.17*
3	1.88***	-0.58	2.46***	0.20	0.19	0.02	1.48**
4	1.06***	0.04	1.02	-0.27	0.35	-0.62	0.81
5	0.90***	0.18	0.72	-0.05	-0.35	0.30	0.43
6	1.19***	-0.40	1.59***	-0.03	-0.21	0.18	1.19***
7	0.97***	-0.35	1.31***	-0.16	0.07	-0.24	1.00***
8	0.97***	-0.48*	1.45***	-0.18	-0.23	0.04	1.18***
9	0.50***	-0.56**	1.06***	-0.14	-0.22	0.08	0.80**
10	0.42***	-0.44**	0.85***	-0.08	-0.33	0.25	0.59**
11	0.41***	-0.44**	0.85***	0.03	-0.34**	0.37*	0.32*