You have a point - but a point is not enough: The case for distributional forecasts of earnings

Ilia D. Dichev
Goizueta Business School
Emory University

Xinyi Huang
Goizueta Business School
Emory University

Donald K.K. Lee
Goizueta Business School and Department of Biostatistics & Bioinformatics
Emory University

Jianxin (Donny) Zhao
Goizueta Business School
Emory University

November 15, 2023

We appreciate helpful comments from workshop participants at the Emory Accounting Brownbag, Southern Methodist University, the 2023 Tulane Mini Accounting Research conference, the 2023 Terry Shevlin University of Washington conference, University of Sao Paulo, Fucape, FGV Ebape, the 2023 Goizueta PhD program conference, and especially Sarah McVay and Andy Call.

You have a point - but a point is not enough: The case for distributional forecasts of earnings

Abstract: Existing forecasts of earnings are typically expressed as point estimates. The future earnings number is ex-ante uncertain, however, and is statistically represented by a probability distribution over all possible earnings outcomes. We use recent advances in statistical machine learning to estimate the distributions of future earnings right before earnings announcements, and investigate how these distributions can help managers, analysts, and investors make better decisions along three directions. First, we show that our distributional forecasts are well-calibrated to actual earnings realizations. Second, we document that management and financial analyst forecasts are way too narrow, severely underestimating the variability of future earnings. Critically, since our distributional estimates are available ex-ante at the firm-quarter level, they can be proactively used to identify and correct such miscalibration. Third, we use our distributional estimates to model the probability of beating or missing the consensus analyst forecasts. Going long (short) on stocks in the extreme decile probabilities of beating (missing) the consensus produces hedge returns of about 60 basis points over the three-day earnings announcement window.

Keywords: Earnings distribution, Statistical machine learning, Analyst forecasts, Stock returns

1. Introduction

In this paper, we explore how to derive and use distributional forecasts of earnings. We focus on earnings because existing research shows that earnings is the single most important number about firm performance, and forecasts of it play a decisive role in issuing stock recommendations. For example, Graham, Harvey, and Rajgopal (2005) find that Chief Financial Officers in U.S. public firms consider earnings to be by far the top firm performance measure. Brown, Call, Clement, and Sharp (2015) show that financial analysts also consider earnings the lynchpin in the evaluation of firm well-being. Given the importance of earnings, it is not surprising that there is a massive literature on earnings forecasting, including major strands around financial analyst forecasts, management forecasts, the use of time-series and cross-sectional models of earnings forecasting, and others (Bradshaw 2011; Kothari, So, and Verdi 2016).

Our main innovation is based on the observation that people typically use point estimates in their earnings forecasting and related decisions. For example, a typical analyst forecast is along the lines of "we expect EPS of $3.00 for the quarter ending March 31, 2023." The consensus forecasts, which are the most widely used analyst forecasts of earnings, are also point estimates. Extant time-series and cross-sectional models of earnings prediction also produce point estimates, e.g., OLS regressions in Abarbanell and Bushee (1997) and Penman and Zhang (2002), see also review of this research in Monahan (2018). More recently, several studies use machine learning techniques to predict earnings or the sign of earnings changes, e.g., Cao and You (2020) and Chen, Cho, Dou, and Lev (2022). Although machine learning offers substantial advantages in terms of allowing non-linear relations, complex interactions between predictive variables, and nonparametric estimation, the earnings outputs are still oriented toward sign or point estimates.

On some level, the widespread use of point estimates in earnings forecasting is not surprising. Point estimates are essentially compact summaries of a lot of information about future earnings, and they are easy to use, remember, and communicate. And yet, by their very nature, point estimates are limited because, ex-ante, the future earnings number is unknown and is therefore represented by a probability distribution over all possible earnings realizations.[1] Thus, point estimates of earnings can be useful, but they are insufficient statistics for the relevant probability distributions. The implication is that considering the full distributions of possible earnings provides more information than point estimates, and therefore allows for better forecasting and investing decisions.

The idea that distributional estimates of future events are better than point estimates has a long tradition in the statistical forecasting literature, and this acknowledgement has percolated in various guises throughout far-flung settings and literatures, including in accounting, law, and economics. For example, Manski (2015) points out that government agencies produce aggregate economic statistics like GDP and household income as point estimates, while they are subject to transitory and permanent statistical uncertainties. Cunningham (2005) argues that the insertion in company GAAP earnings of forward-looking estimates like provisions for bad debt expense creates a false impression of precision and certainty, and recommends using ranges for reported numbers as compared to single amounts. The dominant practice of using ranges in management forecasts of earnings is also an acknowledgment of this issue, and represents a crude attempt at addressing it. Most of these efforts, however, fall rather short of the ideal of producing viable full

[1] The same core intuition appears in a number of settings. For example, the famous Schrödinger's cat paradox from physics can be thought of as an illustration of the inherent probabilistic nature of an outcome that has not yet been observed. In this thought experiment, a cat is put inside an opaque sealed box, and a random amount of poisonous gas is injected inside. Before opening the box to reveal whether the cat is dead or alive, the cat is probabilistically both dead and alive.

distributional forecasts of earnings. Cunningham (2005) provides only a high-level argument, with no technology to implement his suggestions. Ranges in management forecasts are a step in the right direction, but they fall short of operationalizing the key point that some outcomes in these ranges are a lot more likely than others.

We use recent advances in statistical machine learning to produce sound distributional forecasts of earnings. Specifically, we use a distributional machine learning approach developed in Lee, Chen, and Ishwaran (2021) to estimate the earnings distribution right before earnings announcement, conditional on observable inputs like company fundamentals. We use these estimates to provide evidence on the utility of the distributional forecast approach in three directions.

First, we provide calibration evidence on the validity of our approach. Simply put, if our empirical estimation is good, the properties of the distributional forecasts will map snugly into the properties of actual realized earnings. This is exactly what we document, with the percentiles of the estimated distributions mapping tightly into the percentiles of the realizations. In addition, we find that our distributional forecasts of earnings are a much better predictor of earnings realizations than OLS with the same predictive variables. Such evidence confirms the promise of the distributional forecast approach, and provides a solid footing for the remainder of the applications.

Second, we show how the distributional forecasts of earnings can be used to diagnose and improve management and financial analyst forecasts. As mentioned above, management forecasts are primarily expressed in ranges, which correctly reflects the idea that ex-ante earnings represents a probability distribution over possible outcomes. These range forecasts, however, appear to be way too narrow, grossly underestimating the variability of earnings. Building on the preceding evidence that our distributional forecasts are well-calibrated, we estimate that the typical management forecast range has only a 30% chance of covering actual earnings. In addition, there is considerable variation in miscalibration across managers. The key point here is that our distributional estimates are well-calibrated and available ex-ante at the firm-quarter level, providing a systematic way for managers and others to identify and correct the miscalibration in these forecasts. Turning to analyst forecasts, we use the availability of multiple analyst forecasts per company to produce an analyst-implied range forecast of earnings. We find that such analyst range forecasts are also too narrow, although less so than the range forecasts for managers.

Third, we use the announcement of actual earnings and the corresponding stock market reaction to test the utility of our distributional earnings forecasts for investors. Based on our distributional forecasts, we construct a measure of the differential probability of beating/missing the consensus analyst forecasts. Our results indicate that a hedge position in firms belonging to the top (bottom) decile of this measure yields an average return of about 60 basis points (bps) during the three-day earnings announcement window over the period 2011–2021, which corresponds to about 50% annualized abnormal returns. These findings are fairly consistent over time, and are robust to reasonable research design permutations. In addition, we show that the superior stock returns earned at earnings announcements are due to the ability of our differential probability measure to predict consensus-defined earnings surprises, i.e., our differential probability measure is able to identify predictable errors in analyst forecasts. Overall, these results suggest that using distributional forecasts of earnings yields superior stock returns. More broadly, the totality of our findings illustrates the utility of using distributional forecasts of earnings vs. point estimates.

2. Estimating the distributional forecasts of earnings

2.1. The uncertainty about future earnings is completely characterized by a distribution

We start with a brief theoretical grounding, showing that the uncertainty about future earnings is completely characterized by the distribution of future earnings. Therefore, producing a *distributional forecast* of earnings completely characterizes and captures the uncertainty about future earnings.

Consider the following question - given a set of predictors X, what can a researcher say about future earnings $Earn$ of a firm in quarter q? Here, $Earn$ is defined as the earnings per share divided by stock price at the beginning of the quarter:

$$Earn = EPS_q / P_{q-1}, \tag{2.1}$$

and X is a set of variables observed by the researcher before the earnings number is announced. To explain how uncertainty arises in future earnings, we employ a simple statistical framework. $Earn$ depends on X and possibly also on variables ε that are not observed by the researcher at the time of the analysis. The unobserved variables ε include things like insider information or events that have not yet occurred, for example a future pandemic that affects the firm's supply chain and hence earnings. Therefore, the actual earnings number is a function of both the observed and unobserved variables:

$$Earn = g(X, \varepsilon). \tag{2.2}$$

A simple example of Equation (2.2) is the linear regression framework $g(X, \varepsilon) = X'\beta + \varepsilon$, with ε being normally distributed with mean 0 and variance σ^2. Note that $Earn$ is not a fixed deterministic number, due to the inherent uncertainty in the realized value of

ε. To be specific, the uncertainty that remains in $Earn$, after conditioning on the information conveyed by X, is the normal distribution $N(X'\beta, \sigma^2)$ in this example.

Returning to the general case of Equation (2.2) where we do not assume a normal distribution for ε or a linear functional form for $g(\cdot,\cdot)$, the uncertainty remaining in $Earn$ after conditioning on X is completely characterized by the *predictive distribution*:

$$P(Earn \leq y|X) = P(\{\varepsilon : g(X, \varepsilon) \leq y\}|X). \tag{2.3}$$

In this expression, the right-hand side represents the probability that ε takes on the set of values that satisfy the inequality $g(X, \varepsilon) \leq y$.

The key point here is that the predictive distribution distills both sources of uncertainty (knowledge of $g(\cdot,\cdot)$ and the variability in ε) into a single uncertainty measure that tells the researcher everything they need to know about future earnings, such as its mean, variance, skewness, probability of meeting or beating certain thresholds, probability of earnings being confined to a certain range, etc. To link to the existing literature, and to emphasize our main interest in forecasting, we use the term *distributional forecasts* as an alternative for *predictive distributions* as we move to the more practical applications.

2.2. Obtaining practical distributional forecasts of earnings

Since $P(Earn \leq y|X)$ in Equation (2.3) is unknown in practice, an estimate of it

$$\hat{P}(Earn \leq y|X),$$

is needed, and this constitutes our distributional forecast for future earnings. (From $Earn$ it is also straightforward to obtain a distributional forecast in terms of EPS_q, because P_{q-1} is known at the time of forecast, so $Earn$ is a simple rescaling of EPS_q.) One approach to obtain $\hat{P}(Earn \leq y|X)$ is to assume that it belongs to some class of parametric family (e.g., normal distribution), and then fit the parameters of this distribution to data. Such an approach is rather restrictive, however, given existing evidence that earnings distributions tend to be ill-behaved (Burgstahler and Dichev 1997; Givoly and Hayn 2000; Gu and Wu 2003), e.g., they are typically left-skewed and heavy-tailed. We avoid restrictive assumptions that can lead to model misspecification by using a distributional machine learning approach called BoXHED (Wang, Pakbin, Mortazavi, Zhao, and Lee 2020; Pakbin, Wang, Mortazavi, and Lee 2021) to directly estimate Equation (2.3) nonparametrically. BoXHED is an open-source implementation of the method established mathematically in Lee, Chen, and Ishwaran (2021), originally used to estimate hazard functions from time-to-event data.[2] We describe the details of BoXHED in Appendix A.

2.3. Predictor variables

In this paper, we aim to estimate the predictive distribution of future earnings right before earnings announcements. Relying on the voluminous prior literature on the prediction of earnings (e.g., Hou, van Dijk, and Zhang 2012; So 2013; Call, Hewitt, Shevlin, and Yohn 2016; Monahan 2018; Chen et al. 2022), the set of plausible predictors include the last consensus (median) analysts forecast before earnings announcement divided by stock price at the beginning of the quarter, *Consensus*; analyst forecast dispersion divided by stock price at the beginning of the quarter, *Dispersion*; the number of analysts following the firm, *Analyst;* quarterly revenue divided by market capitalization, *Revenue*; book-to-market ratio, *BTM*; net operating cash flow divided by beginning market capitalization, *CFO*; gross profit divided by market capitalization, *GrossProfit*; research and development expenses divided by market capitalization, *R&D*; selling, general and administrative expenses divided by market capitalization, *SG&A*; the natural log of market capitalization, *Size;* quarterly changes in non-cash working capital accounts plus

[2] The BoXHED package is available from https://github.com/BoXHED.

depreciation expense divided by market capitalization, *WC Accruals*; median analyst cash flow forecast before earnings announcements divided by beginning stock price, *CPS*. All variables are measured at the beginning of the quarter except those related to analyst forecasts. Please see Appendix B for a full list of variables used in this study, and their definition and source.

While there may be additional variables that can improve the model's accuracy, we emphasize that the pursuit of the "best" possible model is not the primary thrust of our investigation, and we make no claims in this regard. Rather, our main goal is to provide a direct and uncluttered illustration of the utility of the distributional forecast approach. Hence, the method used to estimate the predictive earnings distribution, and the variables used in the model, may or may not be the best possible. We leave the refinement of these components for future research.

2.4. Data and sample

We obtain earnings and analyst forecast data from IBES, accounting information from Compustat, and stock information from CRSP. To be included in the sample, each firm-quarter observation must have non-missing data about the earnings announcement date, actual EPS, consensus analyst earnings forecast, and beginning stock prices for 2001 through 2021.[3] Our test period is 2011-2021. For each year in the test period, we use the preceding ten years as the training period to fit Equation (2.3).[4] The fitted distribution is then used to compute the probabilities for the test year. To remove the effects of outliers and/or problematic data entries, we delete observations with (1) a stock price of less than $1 or with a market capitalization of

[3] Observations with missing values in the other variables are included in our sample because our machine learning technique is able to handle missing values.

[4] We also used the preceding five years as the training period. Our main inferences remain unchanged using this alternative training period. We chose to use a long training period because forecasting the entire distribution of earnings is more data-intensive when we do not require any model assumptions.

less than $5 million at the beginning of the quarter; (2) earnings announcement dates in IBES and Compustat that are more than one day apart; or (3) *Earn* larger than 0.5 in absolute value. The final sample includes 283,356 firm-quarter observations between 2001-2021.

Table 1 presents the summary statistics for the sample. The average cumulative abnormal returns over the 3-day earning announcement window are close to zero, which is consistent with existing evidence that short-horizon consensus forecasts are close to unbiased. The average firm in our sample is covered by 7 analysts, and the average log of market capitalization is 6.83 (corresponding to a market capitalization of $925 million). Thus, the average firm in our sample is sizable, and likely enjoys a more transparent information environment than the average firm in the Compustat population. The statistics on the other variables are also generally in line with existing evidence from these widely used data sources.

2.5. Validating the estimated distributional forecasts of earnings

Panels A and B in Appendix C present some examples of the estimated predictive distributions for a subset of firms. Panel A presents distributional forecasts of EPS for 9 prominent firms, including Apple, Boeing, Coca Cola, and Chevron, all in the same quarter (Q3 2021). All distributions are single-peaked, but there is considerable variation in the shape, slope, and girth of the tails. Panel B presents distributional forecasts for Apple EPS over 8 consecutive quarters (2019-2020). While the graphs seem to share a family resemblance, there are also visible differences over time, e.g., entries in the middle have noticeably more subdued peaks. Overall, while the evidence in Appendix C is purely for illustrative purposes, it does provide some ground-level feel that the outputs of the BoXHED estimator seem “reasonable”.

Next, we turn to more formal evaluation of the quality of the estimated distributional forecasts, using actual realized earnings as the benchmark. Note that some commonly used goodness-of-fit statistics like mean squared error (MSE) are not directly applicable for our purposes because they are designed for the evaluation of point estimates. Instead, we turn to the statistical literature for two approaches that fit our distributional setting.

2.5.1 Probability calibration plot

The probability calibration plot is an intuitive way to visualize goodness-of-fit (Rice 2006). We plot the empirical probability of *Earn* being less than or equal to some value *y* against the corresponding probability from the predictive distribution. If the predictive distribution agrees with the actual data, the plot should line up close to the 45-degree diagonal line, i.e., the predicted probabilities equal the actual probabilities.

Specifically, the probability plot in Figure 1 is produced in the following way. For each firm-quarter earnings announcement in the test period, we determine the 5^{th} percentile of the predictive distribution $\hat{P}(Earn \leq y|X)$. We then compute the fraction of all firm-quarters whose realized earnings were less than or equal to their corresponding 5^{th} percentile estimates. We plot this fraction on the vertical axis against 0.05 on the horizontal axis.[5] Intuitively, if the estimate of the 5^{th} percentile of the forecast distribution is "good", realized earnings will fall at or below this estimated 5^{th} percentile in about 5% of the cases, so a "good" 5th-percentile estimate will appear as a dot on or close to the plotted 45-degree line. The rest of the plot is produced by repeating this procedure for the 10^{th}, 15^{th}, …, 95^{th} percentiles. Thus, the probability plot provides an intuitive graphical device for assessing the accuracy of our distributional forecasts.

---

[5] The percentiles are computed to a numerical precision of 0.001. That is, we seek a value *y* that is the smallest multiple of 0.001 for which $\hat{P}(Earn \leq y|X) \geq 0.05$. As a result, the value on the horizontal axis is not exactly 0.05, but the average of the cumulative probabilities of being less than or equal to the approximate percentiles, across all firm-quarters.

Figure 1 shows that the percentile dots adhere very closely to the 45-degree diagonal line, which suggests that the predictive distributions are well-calibrated to actual earnings in the test period. We also regress the actual probabilities (values on the vertical axis) onto the predicted probabilities (values on the horizontal axis). The regression coefficient is 0.973 and significant (*t*-statistic = 136.04), and the estimated intercept is 0.003 and insignificant, which again validates the calibration of our fitted distribution.

2.5.2 Continuous ranked probability score (CRPS)

Next, we use the CRPS as the criterion for evaluating the accuracy of our distributional forecasts. The CRPS directly extends the mean squared error (MSE) accuracy measure for point forecasts to distributional forecasts (Gneiting and Raftery 2007).

As a brief primer, suppose we have a point forecast $\hat{y}$ for an observation whose realized outcome is y. The squared error of the point prediction is $(y-\hat{y})^2$, and the MSE for a set of observations is the average of the squared errors over the set. In the distributional setting, we forecast a cumulative distribution function (CDF) $\hat{F}(t)$ instead of a point $\hat{y}$. Note that the realized outcome y, while being a deterministic point, is also a special type of probability distribution that can only have one value. Its corresponding CDF $P(y \leq t)$ equals

$$I_{[y,\infty)}(t) = \begin{cases} 0 & t \text{ is less than } y \\ 1 & t \text{ is greater than or equal to } y \end{cases} . \tag{2.4}$$

Appendix D provides a graphical representation of $I_{[y,\infty)}(t)$. In words, it says that the probability mass is 100% concentrated at the value y, and has zero probability of being anywhere else. In terms of the cumulative distribution, there is zero chance of the realized value of the distribution being *less than* y, but 100% chance that it is *less than or equal to* y, i.e., the realized value is always y.

The integrated squared error between the CDF associated with the realized outcome and the forecasted CDF is

$$\int_{-\infty}^{\infty} \left\{ I_{[y,\infty)}(t) - \hat{F}(t) \right\}^2 dt, \tag{2.5}$$

and the CRPS for a set of observations is the average of the integrated squared errors over the set. In other words, the CRPS is the mean integrated squared error (MISE) between the CDFs for the realized outcomes and the forecasted CDFs. Just like the MSE, the CRPS is always non-negative, and a smaller value indicates a more accurate distributional forecast. Please see Appendix D for a visualization of CRPS.

A major advantage of the CRPS metric is that it allows us to compare the performance of our distributional forecasts to some commonly-used alternatives. Specifically, we compute the CRPSs for three types of distributional forecasts:

i) The forecasted mean of earnings based on an OLS regression model, $\hat{y}_{OLS}$. OLS regression model is perhaps the most common existing technology for predicting earnings, and thus it provides a natural benchmark for our distributional forecasts. Note that while an OLS regression model produces a point forecast, recall from the discussion for Equation (2.4) that a point is a special type of probability distribution. We therefore set the forecasted CDF in Equation (2.5) as $\hat{F}(t) = I_{[\hat{y}_{OLS},\infty)}(t)$.

ii) The mean of the estimated distributional forecast using BoXHED, $\hat{y}_{BoXHED}$. While this is also a point forecast, it potentially improves upon i) because it is derived nonparametrically. Per i), we set $\hat{F}(t) = I_{[\hat{y}_{BoXHED},\infty)}(t)$.

iii) The estimated distributional forecast using BoXHED. The forecasted CDF $\hat{F}(t)$ is the predictive distribution $\hat{P}(Earn \leq t|X)$. Note that a comparison of the results between ii) and iii) helps to distinguish whether the predictive gains between i) and iii) are due to the use of BoXHED, to the use of the distributional approach in forecasting, or both.

All forecasts use the same set of predictors as described in Section 2.3.[6] Table 2 presents the CRPS results for each year in the test period. Column 1 reports the CRPS for the OLS mean forecasts $\hat{y}_{OLS}$, which serves as the baseline for comparison. Column 2 reports the percentage reduction in CRPS when switching over to the nonparametric mean forecasts $\hat{y}_{BoXHED}$. This switch results in a strict improvement in each year, with an average reduction of 18%. When switching from point forecasts to the full distributional forecasts produced by BoXHED, the average reduction doubles to 36%, and strictly improves upon its own mean forecasts in every year (Column 3). The improvements are all statistically significant. Perhaps most importantly, the magnitude of the improvements points to the potential of utilizing distributional forecasts to enhance accuracy in earnings forecasting.

Overall, using the CRPS metric for goodness-of-fit shows that our distributional forecasts lead to significant improvements over an OLS-based model. These gains are due to both using nonparametric machine learning, and to the distributional approach to forecasting.

3. Evaluating the quality of management and analyst forecast ranges

In this section, we demonstrate how to use the estimated distributional forecasts to create prediction ranges for future earnings, and to diagnose and improve management and analyst earnings forecasts.

3.1. Prediction ranges for future earnings

[6] When estimating the OLS regression models, for each year in our test period, we use data from past ten years to estimate the regressions and then use the estimated coefficients to project earnings next period.

Distributional forecasts can be used to provide investors and financial analysts with a *prediction range* within which an upcoming earnings number would fall. More importantly, we can attach a degree of certainty to this interval based on statistical theory. For example, we can provide a prediction range that has an 80% or 95% chance of covering *Earn*.[7] To illustrate, assume that the point estimate of EPS for Company A is $1.90, which corresponds to the mean of the true distribution of future earnings. This, however, tells us nothing about the spread of the range of possible earnings outcomes around the mean. It would be clearly valuable for an investor to also know that there is 80% chance that EPS will be between $1.70 and $2.10, and/or there is 95% chance that EPS will be between $1.50 and $2.30. While management often provide forecasts in a range format, they rarely tie a degree of certainty to these range forecasts, resulting in a significant loss of information to investors.

We can construct prediction ranges for any level of coverage probabilities from our distributional forecasts, be it 80% or 95% or otherwise. Since users probably desire a range with a high degree of coverage, we construct $(100 \times \alpha)\%$ prediction ranges for $\alpha \in \{0.8, 0.85, 0.9, 0.95, 0.99\}$. Of course, the larger the desired coverage α, the wider the prediction range will be. To obtain a $(100 \times \alpha)\%$ prediction range $[l, u]$ such that[8]

$$\hat{P}(l \leq EPS_q \leq u|X) = \hat{P}\left(\frac{l}{P_{q-1}} \leq Earn \leq \frac{u}{P_{q-1}}\middle| X\right) = \alpha, \tag{3.1}$$

[7] Note that what we define here as a prediction range is technically known as a prediction interval in statistics. We use the term prediction range to remain consistent with the management forecast literature. The concept of a prediction interval is similar to that of a confidence interval. However, with a prediction interval, we are concerned with estimating the likely range of a future draw from the distribution, whereas a confidence interval provides a range within which the true parameter of a distribution (e.g., the mean) might fall.
[8] Here, we present earnings in the form of *EPS* instead of *Earn* to be consistent with the format used by investors, financial analysts, and managers. Recall that the stock price at the beginning of the quarter is known at the time of forecast.

we need to find the value l corresponding to $\hat{P}(EPS_q \leq l|X) = \alpha/2$, and the value u corresponding to $\hat{P}(EPS_q \leq u|X) = 1 - \alpha/2$. Thus, there is a $(100 \times \alpha)$% chance that the range $[l, u]$ will contain the actual earnings number, given the observed data X. In practice, Equation (3.1) may not have an exact coverage of α due to the computational cost of finding the exact values of l and u for every firm-quarter. Instead, it suffices to seek approximate values of l and u so that the coverage is at least α.

Figure 2 displays the probability plot for our prediction ranges. The plot is constructed as follows. The horizontal axis represents the desired coverage levels α for our prediction ranges.[9] For each value of α, we compute the fraction of all firm-quarters whose realized earnings fell within their associated $(100 \times \alpha)$% prediction range, and we plot this empirical coverage rate on the vertical axis. For example, for the 80% coverage point on Figure 2, we compute 80% coverage prediction ranges for every firm-quarter in our sample. Then, we compute the percentage of actual earnings realizations, which fall within these 80% ranges, which gives us the y-axis for the 80% coverage point. If our prediction ranges are "good", the y-axis for the 80% coverage will be close to 80%, which implies that the point will fall close to the diagonal on Figure 2. We repeat the same procedure for the 85% coverage, for the 90% coverage, etc. An inspection of the resulting plot in Figure 2 reveals that the derived points adhere closely to the 45-degree diagonal line. This is further validation evidence for our distributional approach, showing that our prediction ranges perform quite well when compared to actual earnings realizations.

[9] As noted above, the prediction ranges are computed to some level of numerical precision, so the values on the horizontal axis are not exactly α. Instead, they are the average of the coverage probabilities for the approximate intervals across all firm-quarters. For example, when we compute the 80% prediction ranges, we achieve a slightly higher coverage of 83% on average.

3.2. Imputing the coverage probabilities of management forecasts

Having validated the accuracy of our distributional forecasts, we now use them to impute the *coverage probabilities* implied by the ranges of management forecasts. During our sample period 2001-2021, more than 83% of individual management forecasts are issued in the form of ranges, which include a minimum value (lower bound) and a maximum value (upper bound). Rather than take a given coverage level α as input in order to produce a prediction range (as in Figure 2), we now go in the reverse direction by taking the management earnings forecast range as input in order to calculate the probability that the range will contain the actual earnings number, i.e., the coverage level α implied by a given management forecast range.

The coverage probability implied by a management forecast range is given by:

$$\hat{P}(EPS_q \leq \text{upper bound}|X) - \hat{P}(EPS_q \leq \text{lower bound}|X). \quad (3.2)$$

Bear in mind that a wide earnings forecast range does not necessarily imply a high coverage probability because coverage depends not only on the width of the raw range but also on the variability of earnings.[10]

3.2.1. Delta Airlines example

Before applying our coverage probability approach to the full sample of management forecasts, we first use the Delta Airlines 2018 Q1 management forecast as an illustrative example. Figure 3, Panel A depicts the BoXHED-derived distributional forecast of earnings for Delta for that quarter, showing an unimodal distribution where most of the probability mass is bound between \$0.50 and \$1 in EPS. Panel A also includes the Delta management forecast range for that quarter, and its corresponding coverage probability. Notice that the management forecast

[10] For example, for a normal distribution with zero mean and a standard deviation of 1, the range $[-1, +1]$ has a 68% coverage. On the other hand, for a normal distribution with a standard deviation of 0.25, the range $[-0.5, +0.5]$ will have a 95% coverage.

range of \$0.65-\$0.75 EPS seems quite narrow as compared to the spread of the BoXHED distribution, and that the estimated coverage probability is only 28%.

These impressions can be made more precise using the prediction ranges discussed above. Figure 3, Panel B extends the Delta example by taking the baseline graph in Panel A, and adding the 80% and 95% prediction ranges derived from our corresponding distributional forecast. As expected, these two ranges span most of the probability mass in the BoXHED-derived distribution, with the range stretching from \$0.57 to \$0.96 EPS to achieve 80% coverage, and stretching \$0.46-\$1.08 for 95% coverage.

Summing up, assessing the Delta 2018 Q1 management forecast range through the BoXHED distributional forecast brings sharp and actionable insights. The management forecast range spanning only 10 cents seems way too narrow given the projected variability of earnings, and needs to be stretched almost four-fold to achieve 80% coverage, and more than six-fold to achieve 95% coverage. Most importantly, these insights can be made available in real time to Delta managers and other users interested in proactive follow-up.

3.2.2.Results for the full sample

The insights from the Delta example are extended to the full sample in the results in Figure 4. The upper box plot in Figure 4 summarizes the coverage probabilities imputed from about 15,000 available management earnings forecast ranges in the test period 2011-2021. As usual, the box plot illustrates the spread of the coverage probabilities through the spacing of the interquartile range; we also include and label the median and the 5^{th} and the 95^{th} percentiles of the distribution of coverage probabilities across firm-quarters. An inspection of Figure 4 reveals a stark message: management forecast ranges appear quite narrow, where the *median* value of coverage probabilities is only 30%, and the coverage probability even at the 95th percentile is far from 100%.

In fact, this coverage seems so low that it warrants some further clarification and comments. Simply put, the evidence in Figure 4 implies that only about 30% of earnings realizations are projected to fall within the range of management forecasts, which seems "too low" on some common-sense level. While this estimate is based on our ex-ante distributional forecasts, it is almost identical to the proportion of ex-post earnings realizations falling within their respective management forecast ranges, which is 30.2% in our sample. For additional validation, in the Call, Hribar, Skinner, and Volant (2023) survey of corporate managers 31.2% of ex-post earnings realizations fall within their management forecast ranges. Thus, our ex-ante estimate that on average only about 30% of earnings will fall within the management forecast ranges is quite close to actual ex-post results.

With coverage probability this low, a natural question is what coverage level managers have in mind in producing their range forecasts. Note that managers likely act motivated by a number of incentives beyond accurate forecasting. For example, existing research finds that manager forecasts tend to be pessimistic, aiming to avoid negative earnings surprises with respect to their forecasts (Ciconte, Kirk, and Tucker 2014). Perhaps managers aim to project confidence and expertise by using narrow forecast ranges, and more generally it is a question of whether their ranges reflect real miscalibration and/or some sort of strategic intent. While questions about intent are often difficult to answer, in this case we have some specific evidence.

The Call et al. (2023) survey documents that on average managers believe that they have a 78% likelihood of reporting earnings within their guidance range. Together with the preceding results, the combined impression is that indeed managers seem to be substantially miscalibrated with respect to the variability of earnings outcomes, and that their forecast ranges are too narrow. In the language of our paper, on average managers seem to believe that their forecast ranges have a coverage probability of about 80%, while their actual coverage probability is only about 30%.

Finally, circling back to the evidence in Figure 4 - and perhaps most importantly - note that the degree of miscalibration by managers exhibits great cross-sectional variation. The 5/25/75/95 percentile of coverage probability is 6%/18%/45%/67%, respectively. Since our distributional forecasts are available ex-ante, and can be tailored to the firm-quarter level, they can serve as a powerful real-time feedback and corrective mechanism for managers, along with other interested parties like financial analysts and investors. Indeed, the illustrative Delta example above already provides an outline for how such a corrective intervention might look like on the ground level.

To be clear, advocating for the use of our distributional forecasts as a corrective mechanism for management range forecasts does not imply that managers need a wholesale adoption of our distributional forecasts. Recall that our distributional forecasts are based on public information, while managers have access to private information, and may also be subject to forecast incentives beyond strictly accurate forecasting (e.g., existing evidence shows that managers have much stronger incentives to avoid negative earnings surprises as compared to positive earnings surprises, Ciconte, Kirk, and Tucker 2014). Thus, we advocate for using our distributional forecasts as one input in producing better management range forecasts rather than as providing the complete solution.

These findings also open up new research opportunities. For example, future research can delve deeper into understanding the factors that contribute to these cross-sectional variations in miscalibration, and their implications for investment decision-making, firm performance, and financial outcomes.

3.3. Imputing the coverage probabilities of analyst forecasts

Next, we utilize our distributional forecasts to analyze the likelihood of the actual earnings number falling within the range of analyst earnings forecasts. Unlike management earnings forecasts, analyst earnings forecasts are typically in the form of point estimates. However, firms are typically covered by multiple analysts; thus, the minimum and maximum values of these earnings forecasts establish a range. We impute the coverage probability of the range for a group of analyst earnings forecasts using the same procedure employed for management earnings forecast ranges. As is probably clear, individual analysts typically do not produce earnings forecast ranges, and so our measure of analyst forecast ranges, and the corresponding results, have a somewhat different interpretation as compared to those for management range forecasts.

The lower box plot in Figure 4 summarizes the imputed coverage probabilities for analyst forecast ranges. Compared to the upper box plot, it is clear that the coverages here are considerably better than that for management range forecasts, but are still low in absolute terms. For example, the median coverage probability is 46%, and the 75th percentile is 58%.[11] We also see substantial cross-sectional variation in coverage, ranging from approximately 15% probability of coverage for the 5th percentile to almost 74% for the 95th percentile. This variation

[11] Note that for many firms there are both management and analyst forecasts, so user expectations of earnings variability possibly reflect the combined information from these two sources of information. To accommodate this possibility, we also calculate coverage probabilities by combining management forecast ranges and analyst forecast ranges for a subsample of firm-quarters with both ranges available. That is, we use the widest possible range that contains the extreme endpoints of both management and analyst forecast ranges. We find that the median coverage probability for such widest ranges is 45%; in comparison, the median coverage probability for this same sample is 30% based on management forecast ranges only, and 37% based on analyst forecast ranges only. Thus, while the joint consideration of analyst and management forecast ranges brings modest increases in coverage, the main finding of low coverage probability in forecasts is still valid.

underscores the importance of knowing the distributional forecasts. Similar to the management forecast setting, the ex-ante availability of our firm-quarter distributional forecasts enables them to serve various interested parties as a powerful correction mechanism for the miscalibration in analyst forecasts.

In summary, Figure 4 suggests that both management and analyst forecast ranges are way too narrow as compared to the variability of the actual earnings numbers. Both also display substantial cross-sectional variation in their ability to reflect future earnings variability. By shifting the focus away from point estimates, and toward considering the full distributions of possible earnings outcomes, users can quantify the coverage associated with either management or analyst earnings forecasts, and take corrective action.

4. The application of distributional forecasts to stock trading

As another application of our distributional forecasts, we develop a stock trading strategy that exploits the information conveyed by the probability distribution of future earnings. Note that the consideration of the full distribution of future earnings opens the possibility for employing a wide variety of trading strategies, e.g., investors can capitalize on forecasts of pronounced left skews by buying put options, and forecasts of heavy tails can be used to take positions in option straddles. In this paper, we aim to keep things simple by examining returns from plain-vanilla trading strategies, basically short-window Buys and Sells in publicly traded U.S.-listed common stocks. We emphasize that maximizing abnormal returns is not a major goal of our investigation. We view the abnormal return evidence as more of an illustration of the utility of the distributional approach rather than as an exercise in seeking optimal trading rules.[12]

[12] Hence, the trading strategy we propose may or may not be the best possible. We leave the refinement of this strategy to future academic research, and possibly efforts in practice.

Prior research suggests that stock investors penalize firms if their earnings miss analyst expectations but reward them for meeting or beating analyst expectations (e.g., Barth, Elliot, and Finn 1999; Defond and Park 2000; Bartov, Givoly, and Hayn 2002; Kasznik and McNichols 2002; Skinner and Sloan 2002). Motivated by these findings, we create a measure from our distributional forecasts that reflects the differential probability of beating (missing) analyst expectations by *N* cents per share at earnings announcements. Larger values of this differential probability measure indicate that a firm's earnings have a higher probability of beating analyst expectations by *N* cents relative to the probability of missing by *N* cents.

We use the latest consensus (median) analyst forecasts as a proxy for analyst expectations. Specifically, for a given *N* and a given firm-quarter q, we use our distributional forecast to calculate the probability of beating the consensus forecast by *N* cents conditional on our set of predictive variables *X* that are publicly available before the earnings announcement:

$$\hat{P}(EPS_q \geq Consensus_q + N|X),$$

and we also calculate the probability of missing the consensus forecast by *N* cents:

$$\hat{P}(EPS_q \leq Consensus_q - N|X).$$

The former is probability mass from the right tail of our predictive distribution, and the latter is probability mass from the left tail. We then compute the *differential probability*

$$\begin{aligned} \Delta Prob = \hat{P}(EPS_q \geq Consensus_q + N|X) \\ - \hat{P}(EPS_q \leq Consensus_q - N|X). \end{aligned} \quad (4.1)$$

A larger value of $\Delta Prob$ indicates that the firm has a higher chance of beating the consensus forecast by *N* cents relative to the chance of missing by *N* cents. Note that our differential probability measure is based on the tails of our predictive distribution rather than the mean of earnings (surprises), which is the focus of prior research. This distinction is probably consequential because, as discussed above, the distributions of realized earnings tend to be ill-behaved, with pronounced skewness and heavy tails. In any case, in Section 4.2 later in the paper, we provide specific evidence that the differential probability measure captures information beyond what is conveyed by the mean of the distribution.

We sort and bin the differential probabilities into deciles, where the decile cut-offs are determined by the deciles of the differential probabilities for the prior year. Panels C and D in Appendix C present some examples of predictive distributions for stocks in the top and bottom portfolios. As one might expect, the top decile firms in Panel C tend to have heavier right tails, and the bottom decile firms in Panel D tend to have heavier left tails.

We then form a trading strategy that takes a long position in the top decile (those with the highest values of $\Delta Prob$) and a short position in the bottom decile (those with the lowest values of $\Delta Prob$). Finally, we compute the cumulative abnormal returns (market-adjusted), CAR, for the portfolios over the three trading days surrounding the earnings announcement date.

Table 3 presents the cumulative abnormal returns for the differential probability portfolios for each year in our test period. In Panel A, we focus on the probability of beating and missing consensus forecasts by 1 cent or more. Columns 1 and 2 present the *CAR* for firms in the top (bottom) group of differential probabilities. Column 3 reports the *CAR* difference in the preceding two columns, which corresponds to returns generated by a hedge portfolio that takes a long position in the top decile and a short position in the bottom decile. Column 1 reveals that the average *CAR* for firms in the top portfolio is 7 bps but statistically insignificant, whereas in Column 2 the average *CAR* for firms in the bottom portfolio is –54 bps and significant. Thus, the abnormal returns are concentrated in the bottom portfolio, where firms have the highest probability of missing the analyst expectation by one cent or more. This finding is consistent with existing research that shows that the penalty for missing the consensus is much greater than the premium for beating the consensus. For example, Koh, Matsumoto, and Rajgopal (2008) show that in years beyond 2003, the earnings announcement return premium for beating the consensus by one cent is close to zero, while the penalty for missing the consensus by one cent is about –4%.

The *CAR* difference between the top and bottom groups, reported in Column 3, is 61 bps and is highly statistically significant. Note that this hedge return is rather substantial economically, corresponding to an annualized return of 51% using the convention of 250 trading days.

The extent to which such returns are actually achievable in practice is less clear. Efforts in this direction need to consider various trading costs and implementability issues. However, such questions are not a major thrust of our investigation for the same reasons that we eschew the fine-tuning of trading strategies for the highest returns. Still, there are reasons to believe that these results likely have practical importance for traders and investors for two reasons. First, trading costs have dramatically declined over the last 20-30 years, and our results are from the period 2011-2021, so trading costs are likely on the low side (Frazzini, Israel, and Moskowitz 2018). Second, our firms are comparatively large and well-followed, as shown in Table 1, which again suggests that trading costs are on the low side.

Note also that systematic risk factors are unlikely to explain the magnitude of the abnormal returns given the short horizon of the stock holding window. Even if the firms on the two sides of the hedge portfolio are only imperfectly matched on risk, possible variations in returns due to hedge portfolio residual risk are unlikely to exceed the average risk premium for *unhedged* portfolios, about 5% annually.[13] This reasoning translates to an upper bound on possible risk effects of about 6 bps over three-day windows, an order of magnitude smaller than the documented effects.

Similar considerations apply to the much broader class of "return factor" variables that predict realized returns, either because of unmodeled risk or because of mispricing, e.g., size, book-to-market, accruals, earnings volatility, gross profitability, default risk, employee growth, etc. (Feng, Giglio, and Xiu 2020; Harvey, Liu, and Zhu 2016). The count of such factors has now reached into the hundreds, and it has become close to impossible to explicitly control for all such potential confounding variables in investigations of abnormal returns. However, the short horizon of our returns offers a resolution to this return-factors conundrum.

The point is that the typical abnormal returns on such variables are 5% to 10% annualized, implying 6-12 bps for three-day windows, which is at least 5 times smaller than our returns. Thus, even if our hedge portfolios are correlated with such factors, and even if the correlation is quite high, the magnitude of our hedge returns practically precludes explanations based on existing return factors. Summarizing, the magnitude of the documented abnormal returns and the short horizons of the earnings announcement windows make risk-based or factor-based explanations unlikely.

Column 3 of Table 3 also suggests that the hedge portfolios are able to generate positive abnormal returns in all years in the sample except 2017 and 2020. In particular, 2020 is the outlier year in which the hedge return is negative, large, and significant. While it is hard to be sure, the COVID-19 pandemic likely has something to do with the breakdown of performance in 2020. It would not be surprising for any model that is trained on data from the preceding ten

---

[13] Annual returns due to systematic risk are on average equivalent to the equity risk premium, the estimate of which typically ranges from 5% to 6%, see Fernandez, García de Santos, and Acin (2022).

years to underperform in the unprecedented environment of 2020, which often played out in ways completely at odds with how the economy functioned in the preceding decade. For example, firms that missed analyst forecasts by a larger margin might be able to receive more financial aid from the government, which results in positive stock market reactions. This pattern of underperformance in 2020 is repeated in pretty much all other specifications in Table 3, which again suggests that there was something systematically amiss in 2020. Excluding 2020, the *CAR* difference between the top and bottom decile jumps from 61 bps to 91 bps.

Panels B and C in Table 3 display the *CAR* results based on the probability of beating/missing analyst expectations by 2 cents and 3 cents respectively. The hedge portfolios earn a 3-day abnormal return of approximately 57 bps in both analyses, and these returns jump up to 84-85 bps when we exclude the year 2020. Thus, the results for these more stringent earnings surprise thresholds are in line with those for the one-cent specification in Panel A.

Summing up, the results in Table 3 suggest that forming trading portfolios that exploit the information conveyed by the predictive distribution of earnings yields abnormal returns on the magnitude of 60 bps over the three-day earnings announcement window. The returns seem economically substantial, are fairly robust, and are unlikely to be fully attributable to systematic risk or other known return factors.

4.1. Does the differential probability measure predict beating/missing consensus forecasts?

The preceding evidence suggests that the differential probability measure $\Delta Prob$ can generate abnormal stock returns. We now turn to extending and verifying this evidence by closer examination of the possible drivers of these returns. Specifically, we examine whether the abnormal returns can be attributed to the power of the differential probability measure to predict the actual beats/misses. We use two approaches to do this: (1) two-way frequency tables, and (2) OLS regressions to predict actual beats/misses.

Table 4 displays the results for the two-way frequency tables, starting in Panel A for the beat/miss by 1 cent specification. The top and bottom rows in Panel A are respectively the top ("beat") and bottom ("miss") deciles based on $\Delta Prob$, while the columns represent the actual beats and misses. Thus, the top-left cell of the 2X2 table represents the percentage of observations that are in the "miss" decile and also missed the consensus by 1 cent, while the bottom-right cell represents the percentage of "beat" observations and also beat the consensus. If our differential probability measure is accurate, we expect most of the observations to fall into the two cells along the downward-sloping diagonal.

The results in Table 4 are consistent with our expectations. In Panel A, we observe that the majority of the observations in the "miss" decile missed the consensus forecast by 1 cent, while most of the observations in the "beat" decile beat the consensus forecast by 1 cent. As a summary statistic, while the expectation of the sum along both diagonals is 50% under the null of no relation between forecasts and beat/miss realizations, the actual sum of the two cells on the downward-sloping diagonal is 71%, and this difference is significant at the 0.1% level in a Chi-square test. The 2-cent and 3-cent specifications in Panels B and C tell a similar story. Overall, the two-way frequency tables show clear evidence that our differential probability measure is predictive of beating/missing the analyst consensus, enabling investors to earn the abnormal returns documented in Table 3.

We also test if our differential probability measure is predictive of the likelihood of beating/missing consensus forecasts by estimating the following OLS regression:

$$Dsurp_q = \beta_1 \Delta Prob\_decile_q + Controls_q + Fixed Effects + \varepsilon. \tag{4.2}$$

$Dsurp_q$ equals +1 (−1) if the actual earnings beat (miss) the consensus by *N* cents in quarter q, and 0 otherwise. We sort firm-quarters into deciles based on their differential probability values $\Delta Prob$, and then scale the decile ranks down to a value between 0 and 1, creating our key independent variable of interest, $\Delta Prob\_decile_q$. We estimate the equations with and without the predictors mentioned in section 2.1 as control variables. We also include industry and year fixed effects, and cluster standard errors by firm and quarter. If our differential probability measure is predictive of the probability of beating/missing consensus forecasts, we would expect $\beta_1 > 0$. We estimate the equations separately for *N* = 1, 2, and 3 cents. Table 5 reports the regression statistics for the estimations. The key finding is that the coefficient on $\Delta Prob\_decile_q$ is significantly positive across all columns, which is consistent with our expectation.

Taken together, the findings in Section 4.1 offer solid evidence that our differential probability measure is predictive of actual beats/misses. In other words, since the beats/misses are defined with respect to the analyst consensus forecasts, the differential probability measure identifies systematic errors in analyst forecasts. The implication is that the documented abnormal returns can be attributed to the power of our differential probability measure to exploit such predictable errors in analyst forecasts, which are revealed and corrected at the earnings announcements.

4.2. Does the differential probability measure capture additional information beyond what is conveyed by the mean?

Our trading strategy demonstrates that using information conveyed by the distributional earnings forecast enables investors to earn abnormal returns in the stock market. Note that by design our differential probability measure utilizes the tails of the distributional forecasts, seeking to accommodate possible skews and heavy tails in building a consensus-beating metric.

However, the differential probability measure is just one of many possible summaries of the forecast distribution. Another summary statistic is its expected value (i.e., the mean), which is the archetypal point estimate of future earnings, and is the focus of most existing research on earnings predictability. Recall that the CRPS analyses earlier in the paper suggest that our distributional forecasts perform better in predicting future earnings as compared to mean-based approaches. In this section, we perform two additional analyses to investigate whether our differential probability measure captures relevant information about future earnings beyond what is conveyed by the mean.

Our first analysis studies the returns from hedge portfolios formed based on the means of the distributional forecasts. If the returns are below those for the trading strategy based on the differential probability measure, the implication is that the distributional forecasts convey valuable information beyond the mean (while holding the BoXHED technology constant).

Specifically, for each firm-quarter, we calculate the expected value of the earnings surprise by subtracting the consensus from the mean of the distributional earnings forecast. We then follow the same return strategy as before, sorting and binning these expected earnings surprises into deciles. Our trading strategy takes a long position in the top group (those with the highest expected surprises) and a short position in the bottom one (those with the lowest expected surprises). Finally, we calculate the market-adjusted cumulative abnormal returns (CAR) for the portfolios over the three trading days around the earnings announcement date.

Table 6 presents the cumulative abnormal returns for the expected surprise portfolios for each year in our test period, using the same format as Table 3. The findings in Columns 1 and 2 suggest that the average *CAR* for firms in the top decile is –7 bps but is statistically insignificant, whereas the average *CAR* for firms in the bottom decile is –48 bps and significant. The *CAR* difference hedge portfolio, reported in Column 3, is 41 bps, and is significant. The magnitude of the abnormal return is economically substantial, confirming the intuition that predicting the mean captures a lot of the information about the distributional forecasts. The 41 bps return, however, is about one-third lower than the hedge return of 61 bps based on the differential probability measure. This result suggests that the differential probability measure is indeed more informative than the mean-based measure.

Our second analysis regresses the returns around earnings announcements (*CAR*) onto the differential probability measure and the expected surprise. Specifically, we estimate the following equation:

$$CAR_q = \beta_1 \Delta Prob\_decile_q + \beta_2 Mean\_decile_q + Controls_q + Fixed Effects + \varepsilon. \qquad (4.3)$$

The variable $\Delta Prob\_decile_q$ ($Mean\_decile_q$) represents the decile ranks of the differential probabilities (expected surprises) within the current year, scaled down to a value between 0 and 1. We include industry and year fixed effects and also cluster standard errors by firm and quarter. If the differential probability measure provides incremental information beyond the expected value, we would expect $\beta_1 > 0$. We estimate equation (4.3) separately for N = 1, 2, and 3 cents.

Table 7 presents the summary statistics for the estimation of Equation (4.3). Columns 1, 2, 4, and 5 of all panels confirm that both the differential probability measure and the expected surprise measure are positively associated with returns around earnings announcements when used independently. However, Columns 3 and 6 indicate that while the coefficient for the differential probability measure remains positive and statistically significant when both measures are included in the regression (e.g., coefficient=63.1 and 59.3, *t*-statistic=2.77 and 2.62 in Panel A), the coefficient for the expected surprise becomes statistically insignificant.

The conclusion is that the BoXHED differential probability measure provides incremental information beyond the BoXHED mean. More broadly, it suggests that the distributional forecast approach dominates mean-based forecast approaches.

5. Conclusion

Existing forecasts of earnings are typically expressed as point estimates. However, ex-ante, the future earnings number is unknown, so it is inherently a probability distribution over all possible earnings outcomes. Therefore, the most informative earnings forecast is a distributional one. Our findings suggest that modeling the predictive distribution of earnings has substantial advantages over point forecasts of earnings.

We use a statistical machine learning approach called BoXHED to nonparametrically estimate the distribution of earnings right before earnings announcements. We examine the utility of the distributional forecast approach along three dimensions. First, we verify that our distributional forecasts map well into the subsequent earnings realizations. Second, we use the distributional forecasts to show that management and financial analyst forecasts are quite narrow, vastly underestimating the variability of future earnings outcomes. Since our distributional forecasts are available ex-ante at the firm-quarter level, they provide calibrated forecast alternatives to managers, analysts, and other users. Finally, we illustrate the utility of our distributional forecasts to stock investors. We use the tails of the distributional forecasts to rank firms based on the likelihood of beating or missing the consensus analyst forecast. Hedge portfolios going long (short) on firms most likely to beat (miss) the consensus earn abnormal returns on the magnitude of 60 bps over the three-day earnings announcement window during the

2011-2021 test period. Further tests verify that these results are fairly robust, and that the superior returns are due to the ability to predict earnings surprises.

We see two broad areas for extending and enriching our approach. The first is more circumscribed and practice-oriented, essentially fine-tuning our approach to better serve various users. For example, investors interested in earning superior returns can use more variables, design different predictive summary statistics, consider trading costs explicitly, and possibly use other financial instruments such as futures, options, and credit default swaps. The second possible area of research is to more broadly seek other areas of application in accounting, finance, and economics, where distributional considerations in forecasting are likely to be fruitful. Possibilities include the prediction of Gross Domestic Product and its components, inflation, capital budgeting, accounting estimates, and others.

References

Abarbanell, J.S., Bushee, B.J., 1997. Fundamental analysis, future earnings, and stock prices. Journal of Accounting Research 35 (1), 1–24.

Barth, M.E., Elliott, J.A., Finn, M.W., 1999. Market rewards associated with patterns of increasing earnings. Journal of Accounting Research 37 (2), 387–413.

Bartov, E., Givoly, D., Hayn, C., 2002. The rewards to meeting or beating earnings expectations. Journal of Accounting and Economics 33 (2), 173–204.

Ben-David, I., Graham, J.R., Harvey, C.R., 2013. Managerial miscalibration. The Quarterly Journal of Economics 128, 1547–1584.

Bradshaw, M., 2011. Analysts' forecasts: what do we know after decades of work?. Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1880339.

Brown, L.D., Call, A.C., Clement, M.B., Sharp, N.Y., 2015. Inside the "black box" of sell-side financial analysts. Journal of Accounting Research 53 (1), 1–47.

Burgstahler, D., Dichev, I., 1997. Earnings management to avoid earnings decreases and losses. Journal of Accounting and Economics 24 (1), 99-126.

Cao, K., You, H., 2020. Fundamental analysis via machine learning. Working paper. Hong Kong University of Science and Technology.

Call, A. C., Hewitt, M., Shevlin, T., Yohn, T. L,. 2016. Firm-specific estimates of differential persistence and their incremental usefulness for forecasting and valuation. The Accounting Review 91 (3), 811-833

Call, A.C., Hribar, P., Skinner, D.J., Volant, D., 2023. Corporate managers' perspectives on forward-looking guidance. Working paper. Arizona State University and ssrn.com.

Chen, X., Cho, Y.H., Dou, Y., Lev, B., 2022. Predicting future earnings changes using machine learning and detailed financial data. Journal of Accounting Research 60 (2), 467-515.

Ciconte, W., Kirk, M., Tucker, J.W., 2014. Does the midpoint of range earnings forecasts represent managers' expectations?. Review of Accounting Studies 19, 628-660.

Cunningham, L.A., 2005. Finance theory and accounting fraud: Fantastic futures versus conservative histories. Buffalo Law Review 53, 789-814.

DeFond, M., Park, C., 2000. Earnings surprises expressed in cents per share: stock price effects and implications for accruals management. Working paper. Leventhal School of Accounting, University of Southern California.

Feng, G., Giglio, S., Xiu, D., 2020. Taming the factor zoo: A test of new factors. The Journal of Finance 75 (3), 1327-1370.

Fernandez, P., García de Santos, T., Acin, J.F., 2022. Survey: Market risk premium and risk-free rate used for 95 countries in 2022. Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3803990

Frazzini, A., Israel, R., Moskowitz, T.J., 2018. Trading costs. Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3229719

Givoly, D., Hayn, C., 2000. The changing time-series properties of earnings, cash flows and accruals: Has financial reporting become more conservative? Journal of Accounting and Economics 29 (3), 287-320.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102 (477), 359-378.

Graham, J.R., Harvey, C.R., Rajgopal, S., 2005. The economic implications of corporate financial reporting. Journal of Accounting and Economics 40 (1-3), 3-73.

Gu, Z., Wu, J.S., 2003. Earnings skewness and analyst forecast bias. Journal of Accounting and Economics 35 (1), 5-29.

Harvey, C.R., Liu, Y., Zhu, H., 2016. … and the cross-section of stock returns. Review of Financial Studies 29, 5-68.

Hou, K., van Dijk, M.A., Zhang, Y., 2012. The implied cost of capital: A new approach. Journal of Accounting and Economics 53 (3), 504-526.

Kasznik, R., McNichols, M.F., 2002. Does meeting earnings expectations matter? Evidence from analyst forecast revisions and share prices. Journal of Accounting Research 40 (3), 727-759.

Koh, K., Matsumoto, D.A., Rajgopal, S., 2008. Meeting or beating analyst expectations in the post-scandals world: Changes in stock market rewards and managerial actions. Contemporary Accounting Research 25 (4), 1067-1098.

Kothari, S.P., So, E., Verdi, R., 2016. Analysts' forecasts and asset pricing: A survey. Annual Review of Financial Economics 8, 197-219.

Lee, D. K. K., Chen, N., Ishwaran, H., 2021. Boosted nonparametric hazards with time-dependent covariates. Annals of Statistics 49 (4), 2101-2128.

Manski, C.F., 2015. Communicating uncertainty in official economic statistics: An appraisal fifty years after Morgenstern. Journal of Economic Literature 53 (3), 631-653.

Monahan, S.J., 2018. Financial statement analysis and earnings forecasting. Foundations and Trends in Accounting 12 (2), 105-215.

Pakbin, A., Wang, X., Mortazavi, B. J., Lee, D. K. K., 2021. BoXHED2.0: Scalable boosting of dynamic survival analysis. arXiv preprint arXiv:2103.12591.

Penman, S.H., Zhang, X.-J., 2002. Accounting conservatism, the quality of earnings, and stock returns. The Accounting Review 77 (2), 237-264.

Rice, J.A., 2006. Mathematical statistics and data analysis. Cengage Learning.

Skinner, D.J., Sloan, R.G., 2002. Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. Review of Accounting Studies 7 (2-3), 289-312.

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.

Wang, X., Pakbin, A., Mortazavi, B. J., Zhao, H., Lee, D. K. K., 2020. BoXHED: Boosted eXact Hazard Estimator with Dynamic covariates. International Conference on Machine Learning 9973-9982.

Appendix A: Estimating distributions with BoXHED

Given a dependent variable Y and a set of predictors X, the conditional probability of Y being larger than y is given by

$$S(y|X) = P(Y > y|X), \tag{A.1}$$

from which we obtain the conditional CDF

$$P(Y \leq y|X) = 1 - S(y|X). \tag{A.2}$$

Thus, knowledge of $S(y|X)$ yields the conditional distribution of Y. For example, in the case where Y represents EPS_q, the probability of beating/missing analyst consensus by N cents per share can be expressed in terms of $S(y|X)$:[14]

$$\begin{aligned} P(EPS_q \geq Consensus_q + N|X) &= S(Consensus_q + N|X), \\ P(EPS_q \leq Consensus_q - N|X) &= 1 - S(Consensus_q - N|X). \end{aligned} \tag{A.3}$$

It is a basic result in statistics that $S(y|X)$ can be derived from the conditional hazard function $\lambda(t|X)$ via[15]

$$S(y|X) = \exp\left\{ -\int_0^y \lambda(t|X)dt \right\}. \tag{A.4}$$

In other words, having an estimate of the hazard function allows us to obtain a plug-in estimate of $S(y|X)$, and hence (A.3). More generally, this framework can also estimate distributions from censored observations. While the earnings forecasting problem is not subject to censoring, it is nonetheless a special case of this framework.

BoXHED is an open-source machine learning package for estimating the conditional hazard function nonparametrically (Wang et al. 2020 and Pakbin et al. 2021). It is a scalable implementation of a gradient boosting procedure proposed in Lee, Chen, and Ishwaran (2021),

[14] Under the assumption that the distribution of Y has a probability density, $P(Y \leq y|X) = P(Y < y|X)$.

[15] While (B.2) only applies when Y is non-negative, it can be easily extended to the case where Y can be negative but is bounded below by some Y_{min}: Apply (B.2) to the shifted variable $Y' = Y - Y_{min} \geq 0$, and the distribution for Y can be recovered from the one for Y' by a simple translation.

and BoXHED inherits the mathematical consistency guarantees from that paper. We use the current version BoXHED2.0 to estimate the conditional hazard function for earnings per share divided by stock price at the beginning of the quarter, i.e.

$$Y = Earn = EPS_q/P_{q-1}. \tag{A.5}$$

The estimated hazard function is then used to obtain $S(y|X)$ via (A.4). To calculate (A.3) from this, note that

$$\begin{aligned} P(EPS_q \geq Consensus_q + N|X) &= P\left(\frac{EPS_q}{P_{q-1}} \geq \frac{Consensus_q + N}{P_{q-1}}\middle| X\right) \\ &= S\left(\frac{Consensus_q + N}{P_{q-1}}\middle| X\right). \end{aligned}$$

The calculation of $P(EPS_q \leq Consensus_q - N|X)$ proceeds in a similar manner.

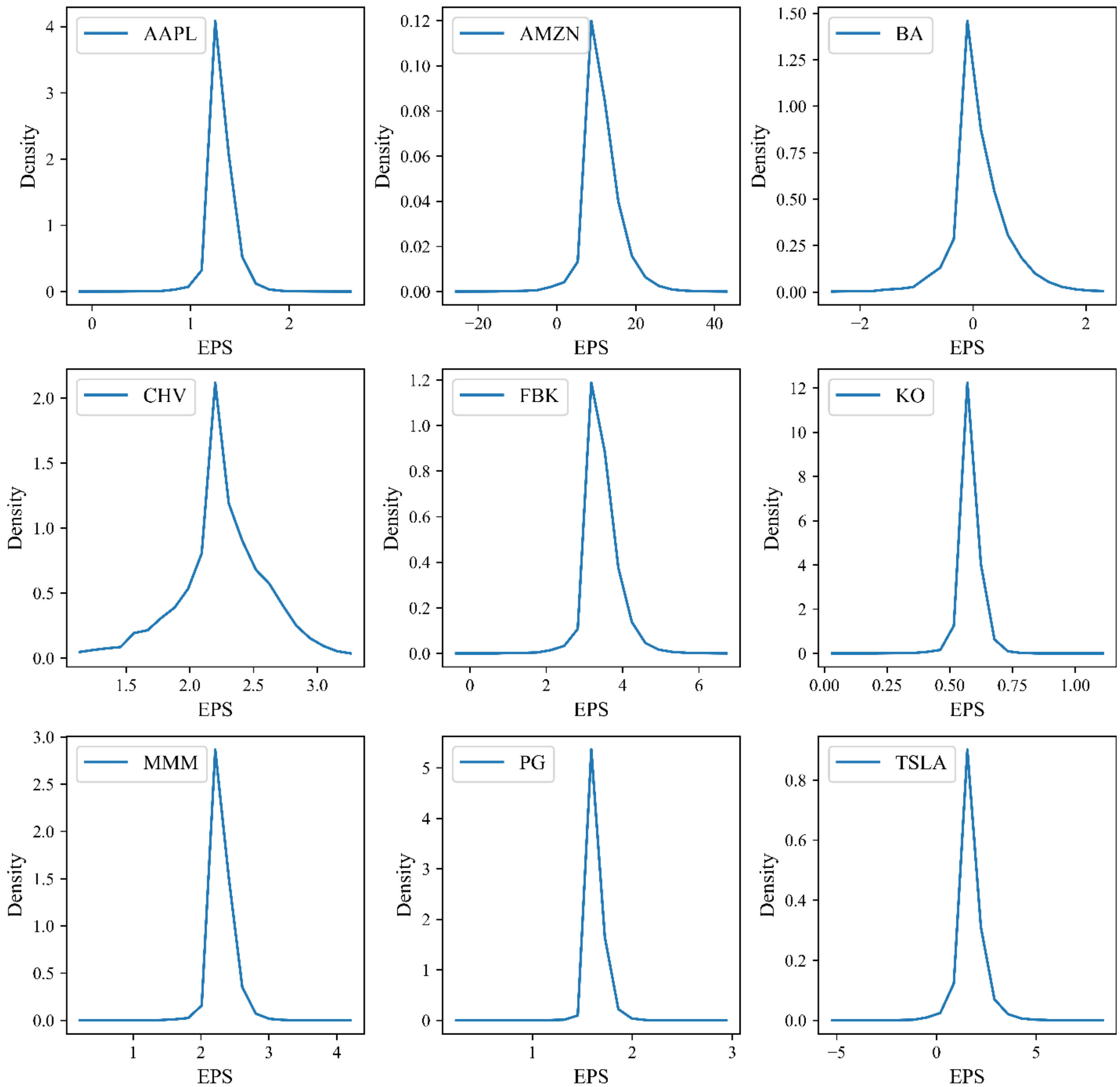
Appendix B: Variable definitions

Variable	Definition	Source
Analyst	The number of analysts following the firm.	IBES
BTM	Book-to-market ratio.	Compustat
CAR	Cumulative market-adjusted abnormal return over the three-day window starting from one trading day before earnings announcements.	CRSP
CFO	Beginning quarterly net operating cash flow divided by beginning market capitalization.	Compustat, CRSP
Consensus	The median consensus analyst forecast before earnings announcements divided by beginning price.	IBES, CRSP
CPS	The median cash flow forecast before earnings announcements divided by beginning price.	IBES, CRSP
ΔProb	The differential probability equals the probability of beating the consensus forecast by *N* cents (*Probability of beating*) minus the probability of missing the consensus forecast by *N* cents (*Probability of missing*).	N/A
Dispersion	The standard deviation of earnings forecasts before earnings announcements divided by beginning price.	IBES, CRSP
Dsurp	A category variable that equals 1 if earnings beat analyst expectation by at least *N* cents, −1 if earnings miss analyst expectation by at least *N* cents, and 0 otherwise.	IBES, Compustat
EPS	Actual earnings per share divided by stock price at the beginning of the quarter.	IBES, CRSP
EPS_{q-4}	*EPS* four quarters ago.	IBES, CRSP
Gross Profit	Revenue minus cost of goods sold divided by beginning market capitalization.	Compustat, CRSP
M_ Dispersion	An indicator variable that equals 1 if *Dispersion* is missing, and 0 otherwise.	N/A
M_EPS_{q-4}	An indicator variable that equals 1 if EPS_{q-4} is missing, and 0 otherwise.	N/A
M_ WC Accruals	An indicator variable that equals 1 if *WC Accruals* is missing, and 0 otherwise.	N/A
M_CPS	An indicator variable that equals 1 if *CPS* is missing, and 0 otherwise.	N/A
M_SG&A	An indicator variable that equals 1 if *SG&A* is missing, and 0 otherwise.	N/A
Mean_decile	Decile ranks of the expected surprise. This variable is scaled down to a range from 0 to 1.	N/A
ΔProb_decile	Decile ranks of the differential probability measure. This variable is scaled down to a range from 0 to 1.	N/A

Probability of missing	Probability of missing the consensus forecast by *N* cents conditional on a vector of covariates *X*.	N/A
Probability of beating	Probability of beating the consensus forecast by *N* cents conditional on a vector of covariates *X*.	N/A
R&D	Quarterly research and development expenses divided by beginning market capitalization.	Compustat, CRSP
Revenue	Quarterly revenue divided by beginning market capitalization.	Compustat, CRSP
SG&A	Quarterly selling, general and administrative expenses scaled by beginning market capitalization.	Compustat, CRSP
Size	Natural log of beginning market capitalization.	CRSP
WC Accruals	Quarterly changes in non-cash working capital accounts plus depreciation expense divided by beginning market capitalization.	Compustat, CRSP

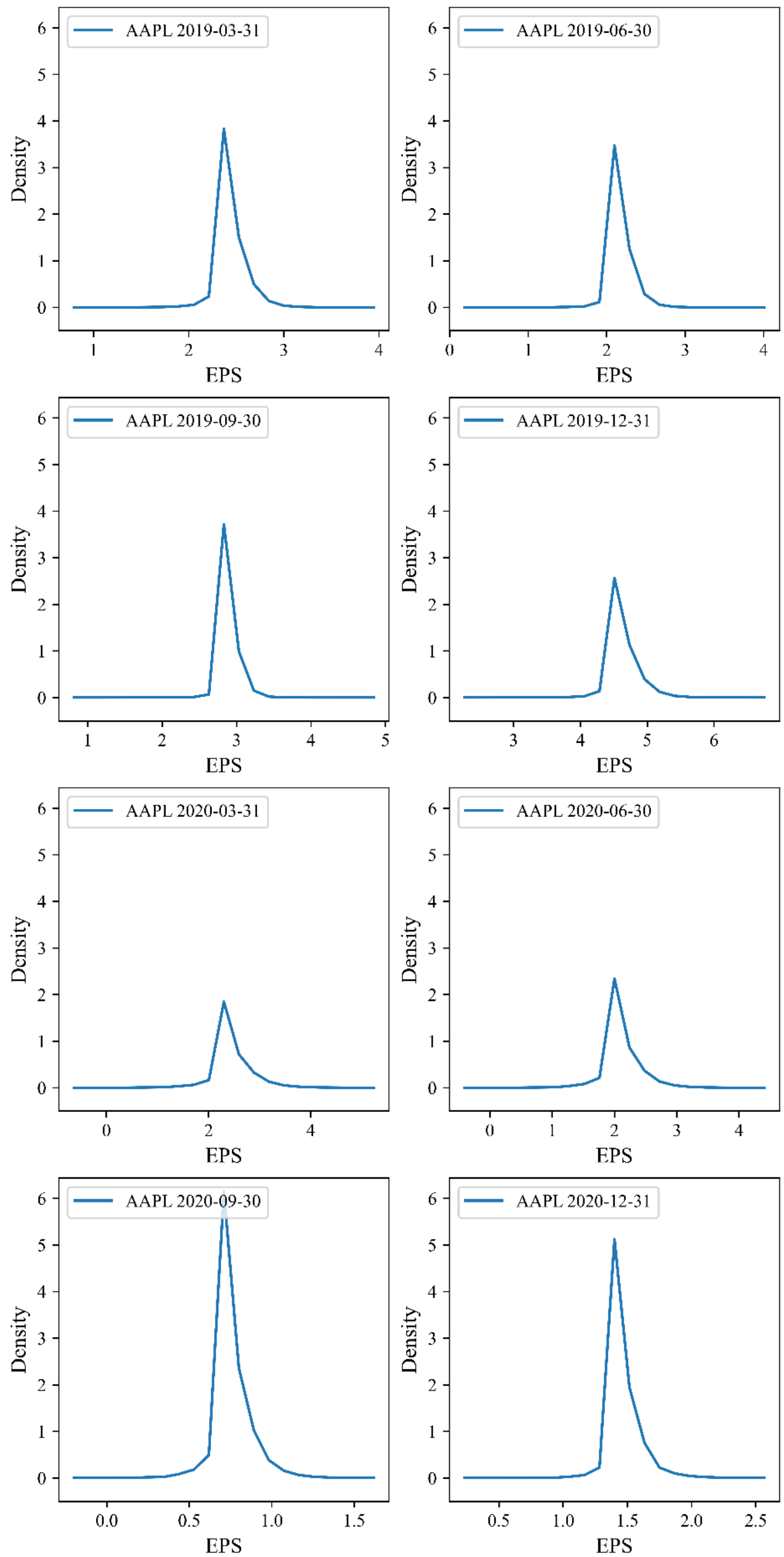
Appendix C: Examples of Earnings Distribution

Panel A: 2021 Q3 Distributions for some prominent firms



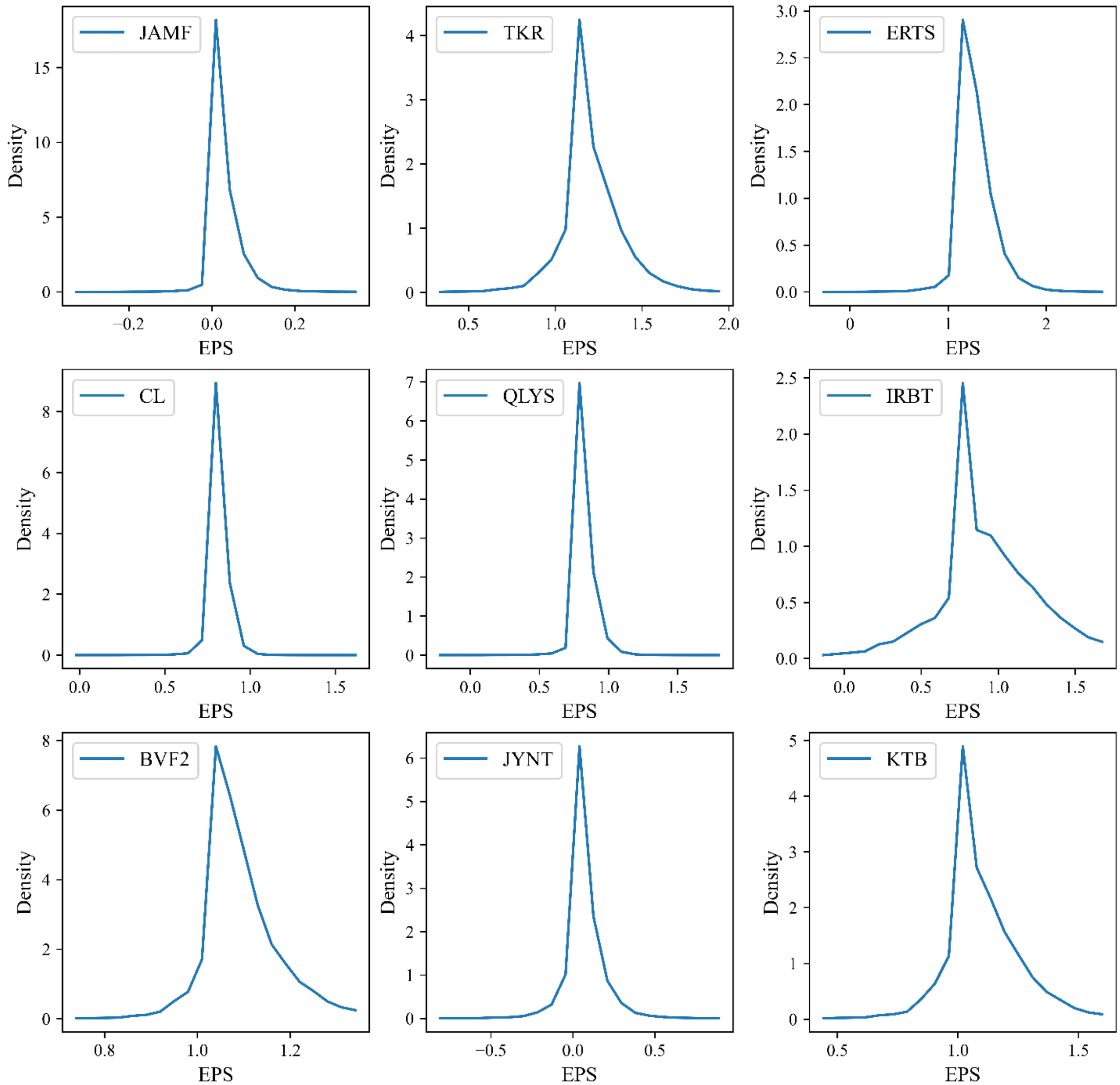
Appendix C: Examples of Earnings Distribution (continued)

Panel B: Distribution for Apple Inc. from 2019 to 2020



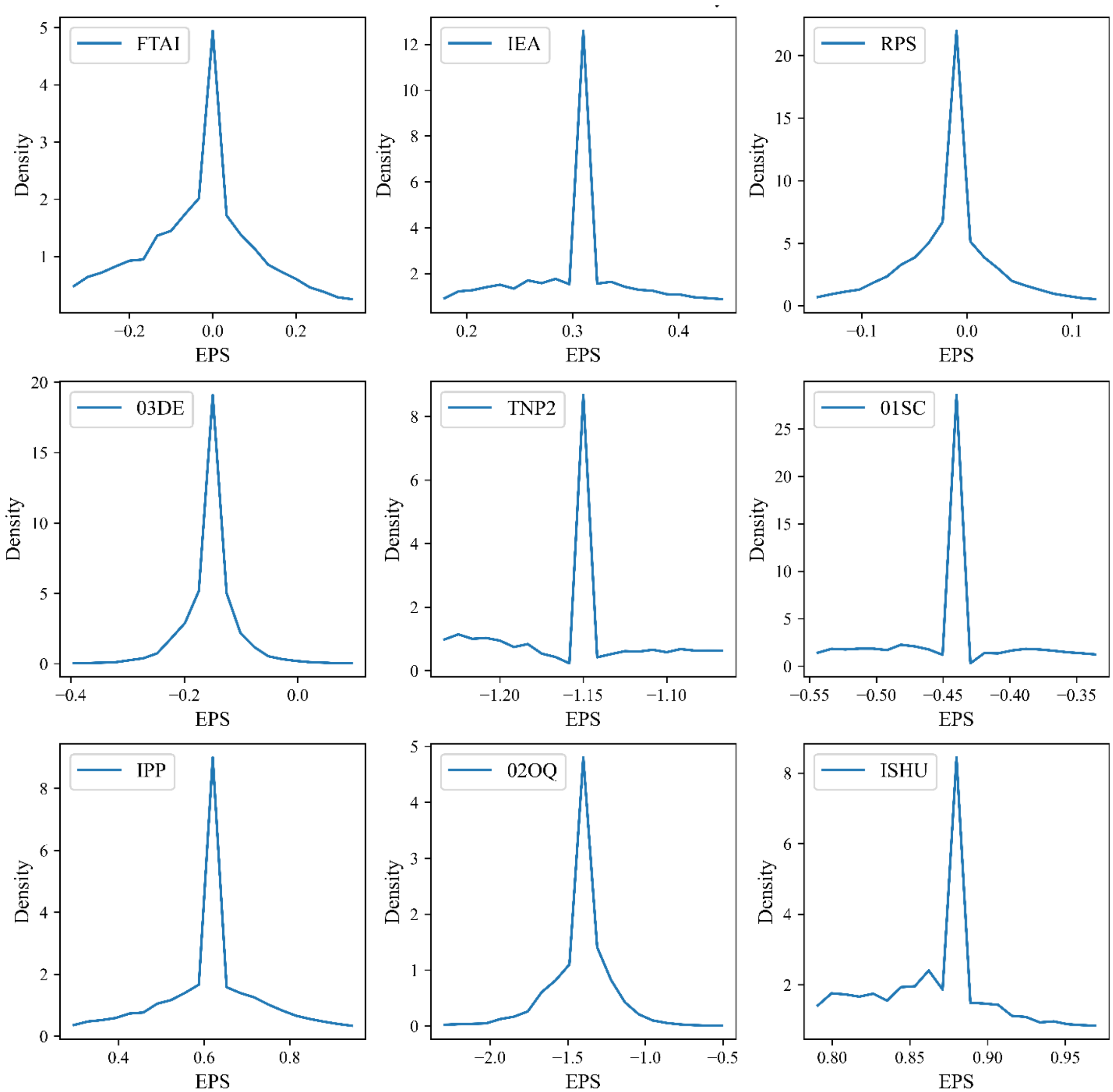
Appendix C: Examples of Earnings Distribution (continued)

Panel C: 2021 Q3 Distribution for Firms in the Top Differential Probability Decile



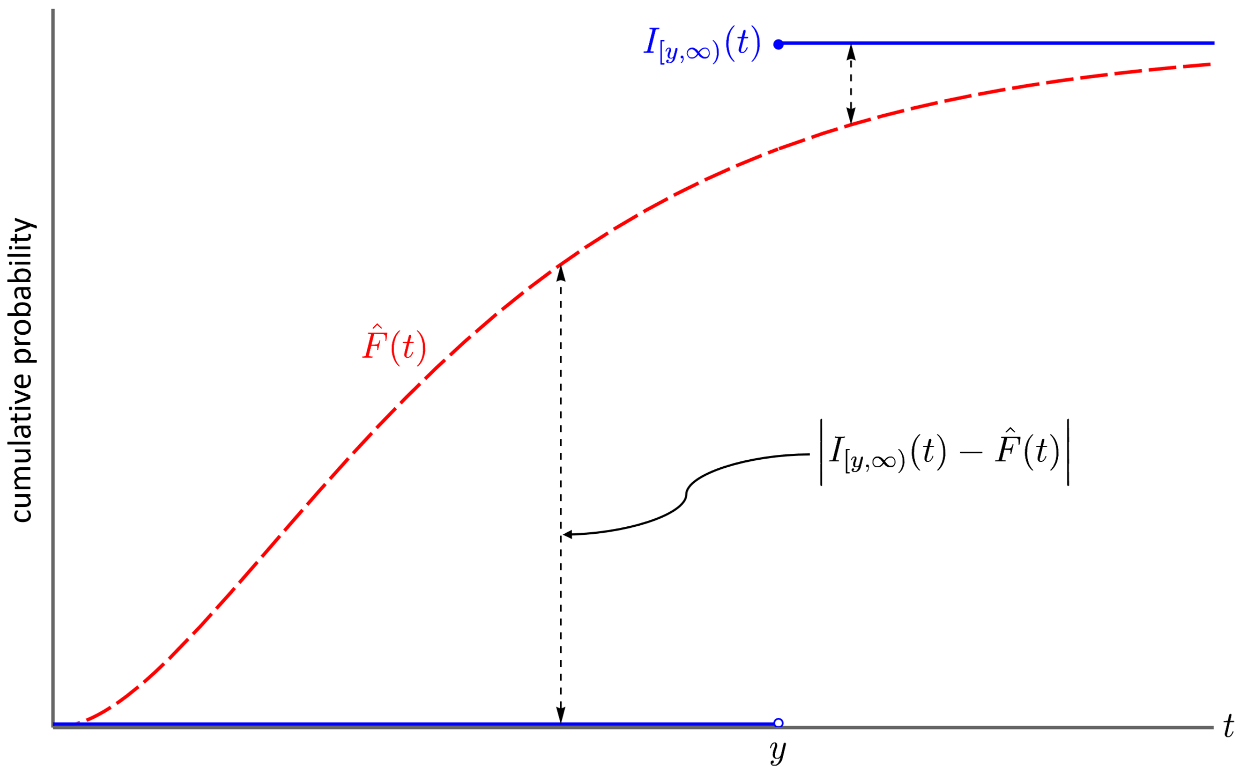
Appendix C: Examples of Earnings Distribution (continued)

Panel D: 2021 Q3 Distribution for Firms in the Bottom Differential Probability Decile



This appendix presents earnings distribution examples. Panel A exhibits the 2021 Q3 earnings distribution for big firms (identified by IBES tickers). Panel B plots the earnings distributions for Apple Inc. from 2019 to 2020. Panels C and D show 2021 Q3 distribution plots for firms randomly selected from the top and bottom differential probability deciles (identified by IBES tickers), respectively. The *x*-axis is earnings per share (EPS) and the *y*-axis is the density.

Appendix D: Visualization of the Continuous Ranked Probability Score (CRPS)



This appendix presents a visualization of the Continuous Ranked Probability Score (CRPS). The blue solid line is the cumulative distribution function (CDF) of the realized outcome y, and the red dashed line is the forecasted CDF $\hat{F}(t)$. Squaring the difference $I_{[y,\infty)}(t) - \hat{F}(t)$ between the two curves and integrating over t yields the CRPS.

Figure 1: Predicted vs. actual probabilities of earnings realizations

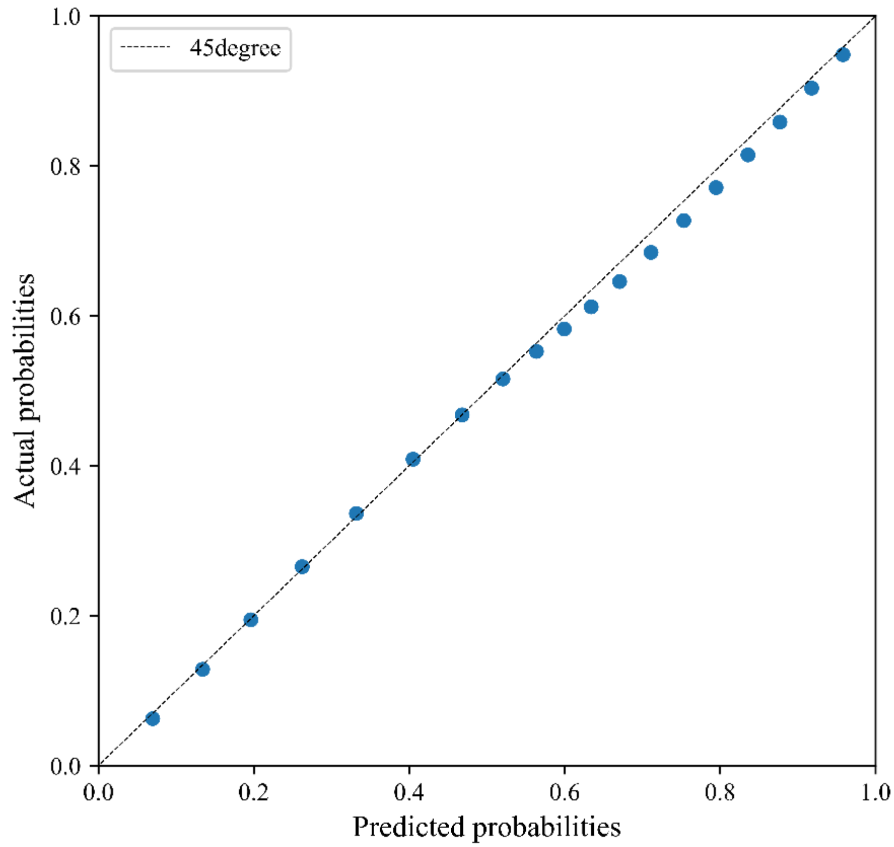


Fig. 1. This figure plots the empirical cumulative probabilities of earnings against predicted probabilities estimated by BoXHED for all observations for test period 2011-2021.

Figure 2: Actual vs. predicted coverage for prediction ranges

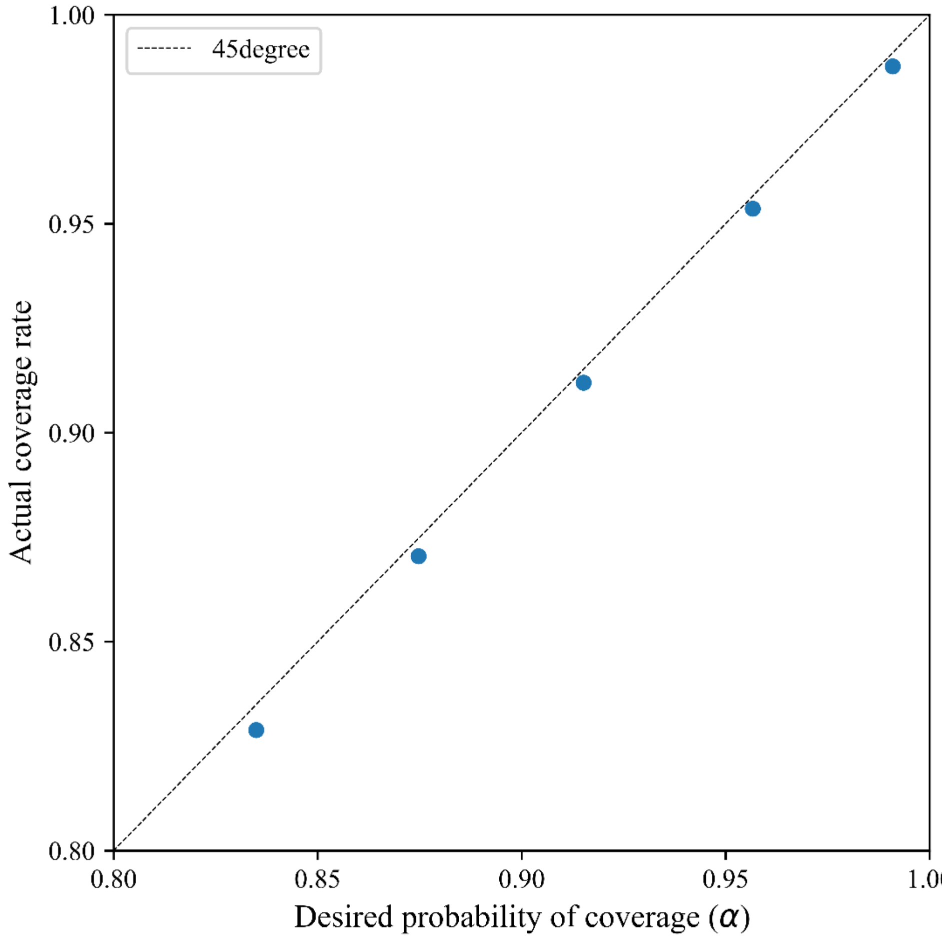


Fig. 2. Fraction of prediction ranges that contain the actual earnings number, plotted against predicted coverage. For test period 2011-2021.

Figure 3: Predicted coverage for Delta Air Lines 2018 Q1 earnings per share (EPS)

Panel A: Coverage probability of the management forecast (MF) range

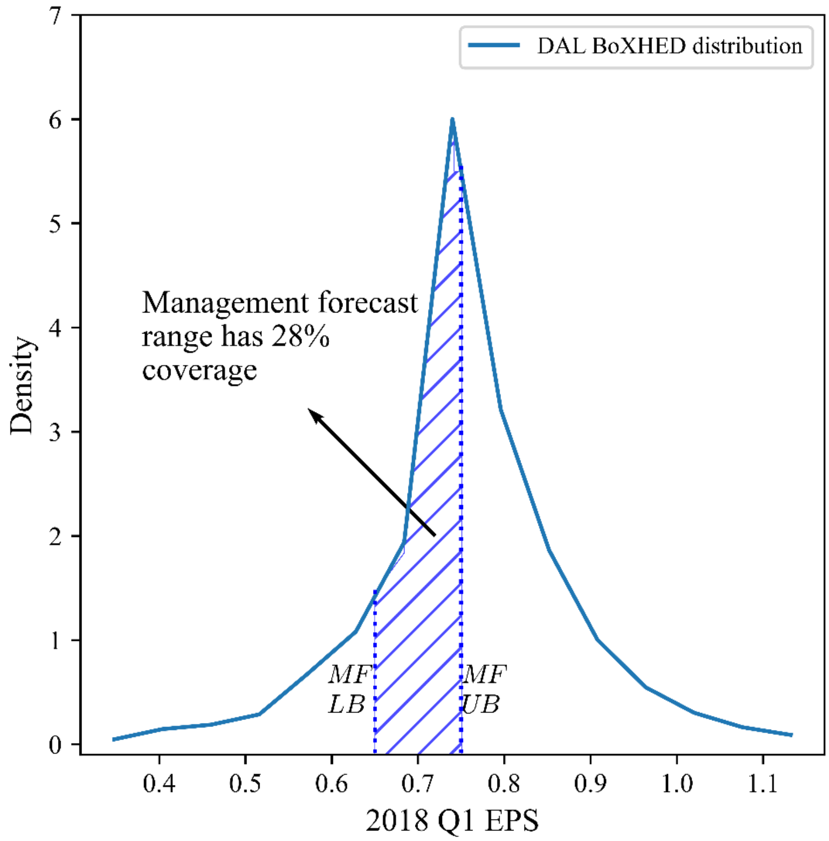


Figure 3 (continued)

Panel B: Delta management forecast range vs. BoXHED-calibrated predicted ranges

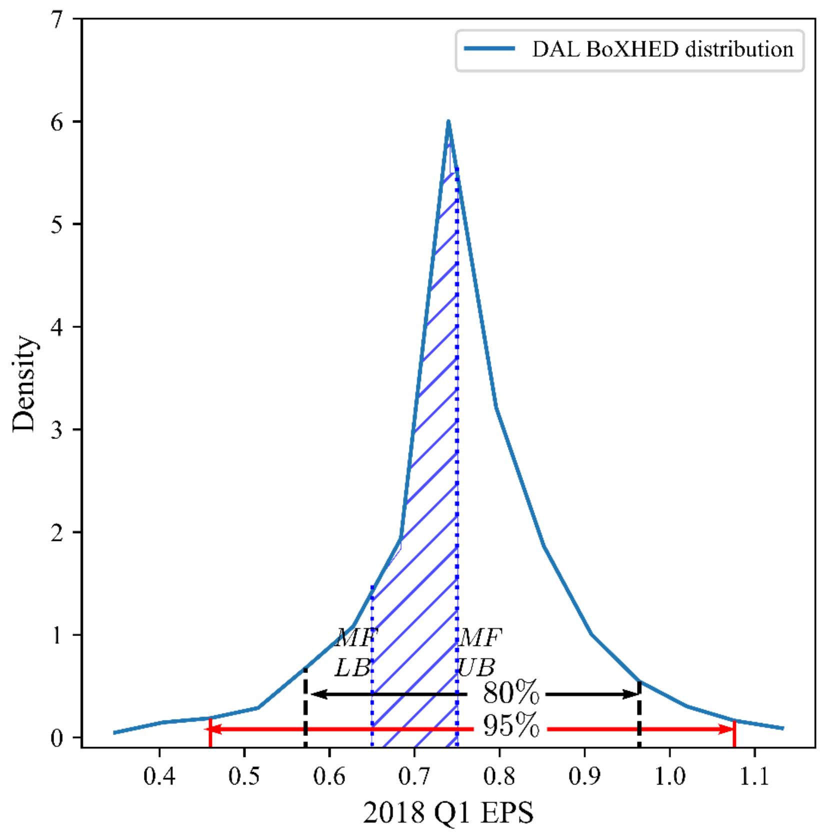


Fig. 3. Predicted coverage for Delta Air Lines 2018 Q1 earnings per share (EPS). Panel A plots the predicted coverage probability of the management forecast range. Panel B plots the management forecast range as well as the prediction ranges that have an 80% or a 95% chance of covering the actual earnings.

Figure 4: Coverage probabilities of management forecasts and analyst forecasts

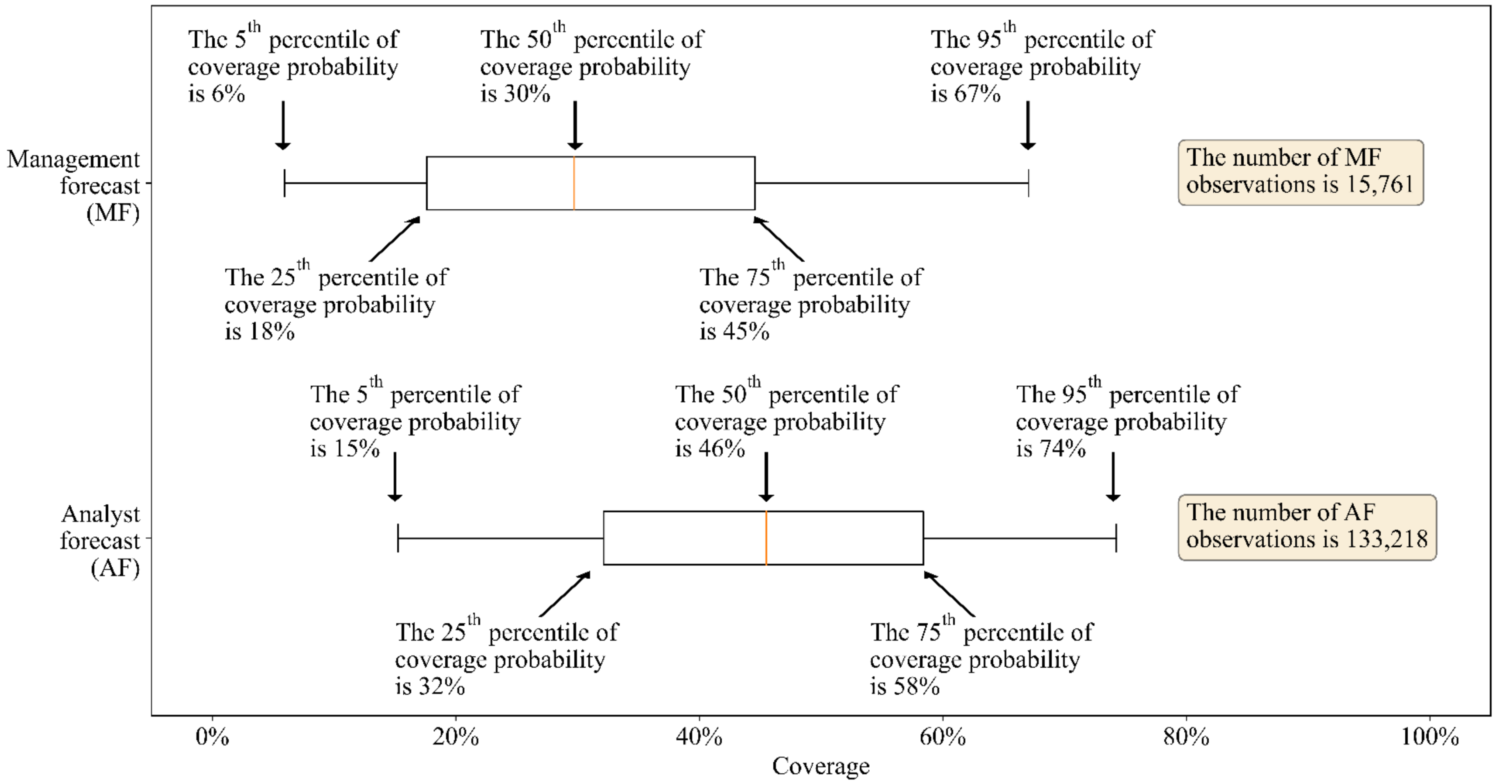


Fig. 4. Box plots summarizing the probabilities of covering actual earnings for management earnings forecast ranges (upper plot) and analyst forecast ranges (lower plot).

Table 1

Descriptive statistics.

Variables	N	mean	std	p25	p50	p75
Analyst	283,356	7.304	6.196	3.000	5.000	10.000
BTM	280,626	0.557	0.487	0.249	0.458	0.750
CAR	283,306	0.000	0.098	−0.041	−0.001	0.041
CFO	283,300	0.023	0.080	0.000	0.017	0.039
Consensus	283,356	0.003	0.039	0.001	0.011	0.017
CPS	100,117	0.026	0.035	0.011	0.020	0.035
Dispersion	248,068	0.003	0.007	0.000	0.001	0.003
Dsurp (N=1)	151,661	0.221	0.915	−1.000	1.000	1.000
Dsurp (N=2)	151,661	0.195	0.847	−1.000	0.000	1.000
Dsurp (N=3)	151,661	0.171	0.801	0.000	0.000	1.000
EPS	283,356	0.001	0.173	0.000	0.011	0.018
EPS_{q-4}	247,580	0.005	0.033	0.002	0.012	0.019
Gross Profit	282,261	0.090	0.131	0.031	0.059	0.103
M_CPS	283,356	0.647	0.478	0.000	1.000	1.000
M_Dispersion	283,356	0.125	0.330	0.000	0.000	0.000
M_EPS_{q-4}	283,356	0.126	0.332	0.000	0.000	0.000
M_SG&A	283,356	0.179	0.383	0.000	0.000	0.000
M_WC Accruals	283,356	0.085	0.279	0.000	0.000	0.000
Mean_decile	151,661	0.500	0.319	0.222	0.444	0.778
ΔProb_decile (N=1)	151,661	0.500	0.319	0.222	0.444	0.778
ΔProb_decile (N=2)	151,661	0.500	0.319	0.222	0.444	0.778
ΔProb_decile (N=3)	151,661	0.500	0.319	0.222	0.444	0.778
R&D	283,300	0.008	0.020	0.000	0.000	0.008
Revenue	282,913	0.309	0.506	0.068	0.147	0.323
SG&A	232,591	0.072	0.102	0.020	0.039	0.078
Size	283,356	6.829	1.768	5.537	6.715	7.994
Surprise	283,356	−0.002	0.168	−0.001	0.000	0.003
WC Accruals	259,330	0.016	0.067	−0.003	0.007	0.024

This table presents summary statistics for the sample of 283,356 firm-quarter observations between January 2001 and December 2021. Definitions of all variables are shown in Appendix B.

Table 2

Continuous ranked probability score (CRPS) for earnings forecasts.

YEAR	(1) CRPS for OLS mean (baseline)	(2) Reduction in CRPS for BoXHED mean	(3) Reduction in CRPS for BoXHED distribution
2011	0.013	30%***	47%***
2012	0.013	32%***	50%***
2013	0.010	24%***	43%***
2014	0.009	26%***	44%***
2015	0.012	22%***	39%***
2016	0.013	21%***	39%***
2017	0.011	22%***	40%***
2018	0.012	15%***	32%***
2019	0.014	9.1%***	27%***
2020	0.021	2.6%***	22%***
2021	0.012	4.4%***	26%***
All	0.013	18%***	36%***

This table presents the CRPS for three types of forecasts. Column 1 reports the CRPS for the forecasted mean of earnings based on OLS regression models. This is the baseline. Column 2 reports the percentage reduction in CRPS (relative to baseline) from using the mean of the predictive earnings distribution estimated using BoXHED. Column 3 reports the percentage reduction in CRPS (relative to baseline) from using the whole predictive earnings distribution estimated using BoXHED. ***, **, * denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table 3

Cumulative abnormal returns (bps) for differential probability portfolios.

Panel A: Probability of beating/missing by one cent or more

	Differential Probability		
	(1)	(2)	(3)
YEAR	High	Low	Diff (High−Low)
2011	23	−153***	175***
2012	6	−75***	81**
2013	42**	−48*	90***
2014	37*	−113***	150***
2015	−13	−54	41
2016	83**	−37	120***
2017	−50**	−34	−16
2018	30	−111***	142***
2019	−47**	−154***	107***
2020	−23	178***	−202***
2021	−11	−49*	38
All (mean)	7	−54***	61***
All (s.e.)	7	9	6
Exclude 2020 (mean)	9	−82***	91***
Exclude 2020 (s.e.)	7	9	6

Panel B: Probability of beating/missing by two cents or more

	Differential Probability		
	(1)	(2)	(3)
YEAR	High	Low	Diff (High−Low)
2011	22	−150***	171***
2012	13	−69**	82**
2013	44**	−47*	91***
2014	37**	−106***	142***
2015	−11	−48	37
2016	72**	−34	107**
2017	−44**	−18	−26
2018	8	−108***	116***
2019	−28	−144***	116***
2020	3	178***	−175***
2021	−18	−46	28
All (mean)	9	−48***	57***

All (s.e.)	7	9	6
Exclude 2020 (mean)	9	−76***	85***
Exclude 2020 (s.e.)	7	9	6

Panel C: Probability of beating/missing by three cents or more

	Differential Probability		
	(1)	(2)	(3)
YEAR	High	Low	Diff (High−Low)
2011	23	−148***	171***
2012	4	−54**	59*
2013	38**	−52*	90***
2014	21	−106***	128***
2015	8	−45	54
2016	56*	−33	89**
2017	−25	−13	−13
2018	1	−94***	96***
2019	−6	−138***	132***
2020	−1	169***	−170***
2021	−7	−49*	42
All (mean)	10	−46***	57***
All (s.e.)	7	9	6
Exclude 2020 (mean)	11	−73***	84***
Exclude 2020 (s.e.)	7	9	6

Table 3 reports the market-adjusted cumulative abnormal returns (*CAR*) for differential probabilities portfolios. Panels A, B, and C report portfolio returns corresponding to the probabilities of beating and missing analyst expectations by at least one cent, two cents, and three cents. All the returns are reported in basis points. ***, **, * denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table 4

Two-way frequency tables for the performance of differential probability measures

Panel A: Probabilities of beating/missing by one cent or more

		Actual: Miss	Actual: Beat
Predicted	"Miss"	29%	21%
	"Beat"	8%	42%

p-value<0.001

Panel B: Probabilities of beating/missing by two cents or more

		Actual: Miss	Actual: Beat
Predicted	"Miss"	29%	20%
	"Beat"	8%	43%

p-value<0.001

Panel C: Probabilities of beating/missing by three cents or more

		Actual: Miss	Actual: Beat
Predicted	"Miss"	29%	20%
	"Beat"	9%	42%

p-value<0.001

Table 4 reports the performance of the top and bottom deciles of differential probability measures in classifying firms into beating/missing consensus forecasts by *N* cents. Panels A, B, and C report the results when *N* equals one cent, two cents, and three cents, respectively. The predicted differential probability measures are calculated as the probability of beating minus the probability of missing. *p*-values for row and column independence are presented.

Table 5

The prediction of beating/missing consensus forecasts using differential probabilities.

Variables	(1) *Dsurp* *N*=1	(2) *Dsurp* *N* =1	(3) *Dsurp* *N* =2	(4) *Dsurp* *N* =2	(5) *Dsurp* *N* =3	(6) *Dsurp* *N* =3
ΔProb_decile	0.679***	0.564***	0.610***	0.500***	0.549***	0.445***
	(0.016)	(0.018)	(0.016)	(0.017)	(0.017)	(0.017)
Consensus		−0.232*		−0.252*		−0.283**
		(0.138)		(0.137)		(0.135)
M_Dispersion		−0.006		−0.004		−0.010
		(0.015)		(0.013)		(0.013)
Dispersion		−1.384**		−1.636***		−1.825***
		(0.598)		(0.574)		(0.577)
Analyst		0.002**		0.002**		0.002**
		(0.001)		(0.001)		(0.001)
M_CPS		−0.000		0.001		0.001
		(0.009)		(0.008)		(0.008)
CPS		0.128		0.194		0.190
		(0.155)		(0.146)		(0.131)
Size		0.014***		0.012***		0.011***
		(0.005)		(0.004)		(0.004)
Revenue		−0.028*		−0.025*		−0.027*
		(0.014)		(0.014)		(0.014)
CFO		0.172*		0.210**		0.223***
		(0.086)		(0.082)		(0.079)
Gross Profit		0.136**		0.145**		0.156**
		(0.063)		(0.061)		(0.061)
R&D		0.290		0.364		0.405*
		(0.244)		(0.225)		(0.208)
M_SG&A		−0.024*		−0.025*		−0.023*
		(0.014)		(0.013)		(0.012)
SG&A		−0.228***		−0.246***		−0.234***
		(0.083)		(0.078)		(0.078)
M_ WC Accruals		−0.010		−0.004		−0.004
		(0.017)		(0.016)		(0.016)
WC Accruals		−0.211*		−0.242**		−0.248**
		(0.116)		(0.111)		(0.106)
BTM		0.017*		0.015		0.014
		(0.010)		(0.010)		(0.009)
M_EPS_{q-4}		−0.031**		−0.038***		−0.035***

	(1)	(2)	(3)	(4)	(5)	(6)
		(0.014)		(0.013)		(0.012)
EPS_{q-4}		0.440**		0.503***		0.548***
		(0.167)		(0.178)		(0.170)
Observations	147,119	146,046	147,119	146,046	147,119	146,046
R-squared	0.072	0.074	0.069	0.072	0.065	0.068
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

This table reports the finding from estimation of equation (4.2). Each observation in the analysis corresponds to a firm quarter. The dependent variable in Columns (1) through (6) is *Dsurp*, a category variable that equals 1 if earnings beat analyst expectation by at least *N* cents, −1 if earnings miss analyst expectation by at least *N* cents, and 0 otherwise. The variable of interest in Columns (1) through (6) is *ΔProb_decile*, the decile of the differential probability of beating/missing consensus by *N* cents. Columns (1) and (2), Columns (3) and (4), and Columns (5) and (6) present the results regarding the likelihood of beating/missing consensus forecasts by one, two, and three cents, respectively. Appendix B presents a description of the variables. Standard errors clustered by firm and quarter are presented in parentheses. ***, **, * denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table 6

Cumulative abnormal returns (bps) for expected surprise portfolios.

	Expected surprise		
	(1)	(2)	(3)
YEAR	High	Low	Diff (High−Low)
2011	0	−145***	145***
2012	−1	−20	18
2013	−17	−56*	39
2014	−43	−131***	88**
2015	1	−15	17
2016	59*	−51*	110**
2017	−2	−29	27
2018	22	−120***	142***
2019	−143***	−175***	32
2020	67*	214***	−147***
2021	−28	−117***	89
All (mean)	**−7**	**−48*** **	**41*** **
All (s.e.)	**10**	**10**	**7**
Exclude 2020 (mean)	−15	−85***	69***
Exclude 2020 (s.e.)	10	10	7

This table presents the market-adjusted cumulative abnormal returns (*CAR*) for expected surprise portfolios. All the returns are reported in basis points. ***, **, * denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table 7

Differential probabilities versus expected surprise in predicting cumulative abnormal returns.

Panel A: Probabilities of beating/missing by one cent or more, and expected surprise

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*CAR*	*CAR*	*CAR*	*CAR*	*CAR*	*CAR*
ΔProb_decile	58.4***		63.1***	63.4***		59.3**
	(16.7)		(22.8)	(17.4)		(22.6)
Mean_decile		42.6***	−5.7		40.0**	4.8
		(14.8)	(19.1)		(15.5)	(20.1)
Observations	147,119	147,119	147,119	146,046	146,046	146,046
R-squared	0.003	0.003	0.003	0.004	0.004	0.004
Controls	No	No	No	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Probabilities of beating/missing by two cents or more, and expected surprise

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	*CAR*	*CAR*	*CAR*	*CAR*	*CAR*	*CAR*
ΔProb_decile	55.1***		55.4**	57.4***		48.8**
	(16.3)		(22.3)	(16.3)		(21.0)
Mean_decile		42.6***	−0.361		40.0**	10.3
		(14.8)	(19.345)		(15.5)	(20.1)
Observations	147,119	147,119	147,119	146,046	146,046	146,046
R-squared	0.003	0.003	0.003	0.004	0.004	0.004
Controls	No	No	No	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 7 (continued)
Panel C: Probabilities of beating/missing by three cents or more, and expected surprise

Variables	(1) *CAR*	(2) *CAR*	(3) *CAR*	(4) *CAR*	(5) *CAR*	(6) *CAR*
ΔProb_decile	52.7***		49.4**	52.5***		40.5**
	(15.5)		(20.2)	(15.2)		(19.0)
Mean_decile		42.6***	4.1		40.0**	14.8
		(14.8)	(18.8)		(15.5)	(19.9)
Observations	147,119	147,119	147,119	146,046	146,046	146,046
R-squared	0.003	0.003	0.003	0.004	0.004	0.004
Controls	No	No	No	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

This table reports the findings from estimation of equation (4.3). Panel A, B, and C reports the prediction of cumulative abnormal returns using decile ranks of the differential probability of beating/missing consensus by one cent, two, and three cents, respectively, and the decile ranks of expected surprise. Each observation in the analysis corresponds to one quarterly earnings announcement. The dependent variable in Columns (1) through (6) is the market-adjusted cumulative abnormal returns (*CAR*). The independent variables of interest are *ΔProb_decile*, the decile ranks of the differential probability of beating/missing consensus by *N* cents (*N*=1,2, and 3), and *Mean_decile*, the decile ranks of the expected surprises. Appendix B presents a description of the variables. Standard errors clustered by firm and quarter are presented in parentheses. ***, **, * denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.