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
School of Management and Economics
The Chinese University of Hong Kong, Shenzhen

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

Jianxin (Donny) Zhao
Goizueta Business School
Emory University

November 17, 2025

We appreciate helpful comments from workshop participants at the Accounting Design Project, the Southeast Summer Accounting Research Conference, the Emory Accounting Brownbag, Georgia Institute of Technology, the mEmory Accounting Conference, Southern Methodist University, the Tulane Mini Accounting Research Conference, the 2023 Terry Shevlin University of Washington conference, Texas A&M, University of Georgia, University of Michigan, University of Sao Paulo, Fucape, FGV Ebape, University of Chile, Catholic University of Chile, the Goizueta PhD program conference, and especially Andy Call, Sarah McVay, Peter Pope, Ted Christensen, Matt Lyle, Karl Schumacher, and Teri Yohn.

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

Abstract: Existing earnings forecasts are typically point estimates. Future earnings are uncertain, however, and are therefore represented by probability distributions. Thus, proper earnings forecasts are distributional. We empirically demonstrate the value of distributional forecasts for decision making. We first estimate distributional forecasts using quantile regressions employed by prior research, then introduce a machine learning approach called BoXHED. Using the continuous ranked probability score, we find clear evidence that distributional approaches provide more accurate forecasts than point estimates, with BoXHED performing best among distributional approaches. We demonstrate practical applications through stock trading around earnings announcements. Going long (short) on stocks with extreme positive (negative) expected payoffs based on distributional forecasts produces returns of about 90 basis points over three-day windows, substantially outperforming point forecast strategies. Additionally, we document that management and analyst forecasts severely underestimate earnings variability. Critically, our real-time distributional forecasts at the firm-quarter level can identify and correct such miscalibration.

Keywords: Distributional forecasts, Statistical machine learning, Analyst forecasts, Stock returns

AI disclosure: The authors have not used AI tools in producing this study

1. Introduction

This study explores how to use distributional forecasts of earnings. We focus on earnings because existing research shows that earnings is the single most important number about firm performance, e.g., Graham, Harvey, and Rajgopal (2005) and Brown, Call, Clement, and Sharp (2015). Consistent with this importance, there is a massive literature on earnings forecasting, including major streams on 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).

A review of this voluminous literature reveals that most earnings forecasts are produced as point estimates, where examples include most analyst forecasts and the output of most extant time-series and cross-sectional models of earnings forecasts, e.g., OLS regressions in Abarbanell and Bushee (1997) and Penman and Zhang (2002), see also a 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 advantages in terms of allowing non-linear relations, complex interactions between predictive variables, and nonparametric estimation, the earnings outputs are still produced as 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 comparatively easy to produce. However, by their very nature, point estimates are limited because they fail to capture the uncertainty about future earnings. We know from probability theory that uncertainty is fully characterized by the probability distribution over all possible outcomes (Dawid 1984; Gneiting 2008). Thus, since the future earnings number is uncertain ex-ante, proper forecasts of earnings should be in the form of a probability distribution.

The desirability of distributional forecasting is well-established in statistics. Stigler (1975) describes the shift from point forecasts to distributional forecasts, with early roots as far back as the 19th century. Gradually, this shift has reached a variety of disciplines (Gneiting 2008), including weather and climate prediction, election predictions, and the forecasting of a variety of economic and financial variables (see Gneiting and Katzfuss 2014 for a review).[1]

Earnings forecasting has also made advances towards a distributional approach. The use of ranges in most management forecasts of earnings and some analyst forecasts (e.g., see Joos, Piotroski, and Srinivasan 2016 for Morgan Stanley range forecasts) essentially represents a crude attempt at quantifying uncertainty.[2] In the academic literature, early efforts focus on forecasts of the variance of earnings, e.g., Baginski and Wahlen (2003), Pastor and Veronesi (2003) and Donelson and Resutek (2015). Such efforts are further developed in two studies that represent the current state of the art in earnings uncertainty characterization. Konstantinidi and Pope (2016) use linear quantile regressions to estimate 11 quantiles of future earnings, which they use to create summary measures of earnings risk such as dispersion, asymmetry, and tail risk. Chang, Monahan, Ouazad, and Vasvari (2021) also use linear quantile regressions to estimate 150 quantiles of future earnings, and use them to estimate moments of earnings such as the variance, skewness, and kurtosis. They then investigate whether these measures are reflected in equity prices and credit spreads.

Although distributional forecasts of earnings are statistically appealing, they remain uncommon in practice and underexplored in the accounting literature. While prior research has

[1] As one example, *The Survey of Professional Forecasters*, the oldest quarterly survey of macroeconomic forecasts in the U.S., now provides distributional forecasts of variables like inflation, unemployment and GDP changes. On the desirability of distributional forecasting of macroeconomic variables, see also Manski (2015) which argues that producing aggregate economic statistics like GDP and household income as point estimates is deficient because they are subject to transitory and permanent statistical uncertainties.
[2] Unlike a distribution, a range provides no information about the occurrence probability of each point in the range.

made important progress toward distributional forecasting of earnings, these studies stop short of demonstrating when and how such forecasts improve decision-making relative to traditional point estimate forecasts. Without such demonstrations, the practical value of distributional forecasts remains unclear despite their theoretical appeal. This paper addresses this gap by clarifying the conditions under which distributional forecasts are valuable, and providing concrete empirical evidence of how they enhance decisions as compared to point estimates. The key insight is based on the "flaw of averages" idea that decisions based on a single point forecast (e.g., the mean) can yield severely flawed outcomes when the underlying payoff function is nonlinear (Savage, 2003).[3] We formally develop this theoretical foundation in Section 2.2 to show why distributional forecasts are needed whenever the payoff for a decision problem is nonlinear in future earnings.

To empirically demonstrate the value of distributional forecasts, we proceed in two stages. In the first stage, we estimate the distributional forecasts using a range of methods, and assess the accuracy of distributional vs. point forecasts. We begin with linear quantile regression employed by Konstantinidi and Pope (2016) and Chang et al. (2021) to obtain multiple quantiles of the distribution. We then move beyond linear models by using recent advances in statistical machine learning to nonparametrically estimate the full cumulative distribution forecasts of earnings. Specifically, we introduce a statistical machine learning method known as BoXHED that is rigorously developed in Lee, Chen, and Ishwaran (2021). With these methods, we produce distributional forecasts of earnings right before earnings announcements, conditional on

[3] See https://hbr.org/2002/11/the-flaw-of-averages for real-world examples and also Section 2.2 for more details. In brief, Savage (2003) formalizes the flaw through the inequality $E[r(X)] \neq r(E[X])$. X is random variable (e.g., earnings, product demand, or cost), $r(X)$ is a nonlinear payoff function. Here, the payoff based on the average value of X, $r(E[X])$, is different from the expected payoff $E[r(X)]$. Hence, to maximize the expected payoff, one must use the entire distribution of X to calculate $E[r(X)]$.

observable inputs such as company fundamentals and analyst variables. Our sample period is 2001-2021, where the distributional models are trained on observations from the preceding decade to produce out-of-sample forecasts for the next year on a rolling basis.

We then assess the accuracy of distributional vs. point forecasts against actual earnings realizations. We adopt the *continuous ranked probability score* (CRPS), a statistically rigorous measure that is well established in the statistical literature. The major advantage of CRPS is that it allows direct comparisons of point estimates vs. distributional forecasts of earnings (and among distributional forecasts), which has not been done before. Specifically, we compare the accuracy of two point forecasts (OLS mean and mean of the BoXHED distribution), two distributional forecast approaches from the existing literature (Konstantinidi and Pope 2016; Chang et al. 2021), and BoXHED distributional forecasts. We find clear evidence that distributional approaches provide more accurate forecasts than point estimates. Among distributional approaches, BoXHED performs the best, with statistically significant improvements in forecast accuracy over the two other distributional approaches.

In the second stage, we show two applications of the derived distributional forecasts. We start with stock trading as a setting to concretely demonstrate how distributional forecasts can be used to make better investment decisions compared to point forecasts. Specifically we introduce a novel stock trading strategy around earnings announcements based on the distributional forecasts. As discussed above, distributional forecasts are especially valuable to investors seeking to capitalize on pronounced non-linearities in earnings-based return outcomes. The earnings announcement setting is a natural fit for this because existing research documents much higher absolute returns for negative earnings surprises than for positive earnings surprises (Koh, Matsumoto, and Rajgopal 2008). We capitalize on this asymmetric non-linearity by constructing a measure of expected investor payoffs based on the full distributional forecasts of earnings.

Controlling for existing return factors, our results indicate that a hedge strategy with a long (short) position in firms belonging to the top (bottom) decile of expected payoff yields an average return of about 90 basis points (bps) during the three-day earnings announcement window over the period 2011–2021, which corresponds to about 76% annualized abnormal returns. These findings are fairly consistent over time, and are robust to controlling for established return factors. More importantly, we find that the expected payoff measure generates almost double the abnormal returns compared to approaches based on point forecasts, where investors solely predict mean earnings surprises. Overall, these results suggest that using distributional forecasts of earnings offers substantial benefits to stock investors compared to traditional point forecasts.

To further illustrate the utility of distributional forecasts, we show how they 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 future earnings is uncertain. These management range forecasts, however, appear to be too narrow, substantially underestimating the variability of earnings. Specifically, we show that the typical management forecast range has only a 30% chance of covering actual earnings, while managers believe that they have about 80% coverage. While this magnitude of underestimation seems rather large, these findings are consistent with survey-based evidence of miscalibration in management forecasts (Call, Hribar, Skinner, and Volant, 2024).

Crucially, and in contrast to survey-based evidence of miscalibration, we are the first to provide distributional forecasts in real time at the *firm-quarter* level, providing a practical way for investors and other users to identify and correct such miscalibration. Turning to analyst forecasts, we use the availability of multiple analyst forecasts for a firm-quarter to produce an analyst-implied range forecast of earnings. We find that such analyst range forecasts are also too narrow, although less so than management range forecasts. Just like with management forecasts, the availability of distributional forecasts at the firm-quarter level allows users to correct for such analyst miscalibration.

This study makes several contributions and opens new opportunities for research. It provides the first direct empirical evidence that distributional forecasts of earnings lead to better forecasting accuracy and investment decisions relative to point forecasts. Specifically, the study offers a theoretical characterization of distributional forecasts of earnings and implements, for the first time, statistical machine learning techniques to estimate them. In addition, it introduces direct and specific tests for evaluating the accuracy of distributional forecasts, which allow their benchmarking against point estimates and each other. These tools are potentially useful for future research on distributional forecasts of other variables of importance in accounting, finance and economics. For example, the blueprint of our approach can be easily applied to the forecasting and calibration of sales, accruals, cash flows, capital expenditures, and GDP and its components.

Our study is also the first to both theoretically elucidate and empirically establish how distributional forecasts of earnings can be used to earn superior stock returns by exploiting non-linearities in the return response to earnings surprises. The upshot of this evidence is that distributional forecasts are likely to be especially useful for other settings with significant non-linearities in earnings-based outcomes, including modeling executive compensation, bankruptcy, and financial distress as a function of future earnings. More broadly, this same intuition applies to modeling and capitalizing on other non-linear relations such as recording an impairment as a function of future cash flows, option pricing as a function of future stock returns, and bond pricing as a function of future interest rates.

The remainder of the paper is organized as follows. Section 2 lays out the theory and methodology for estimating the distributional forecasts of earnings, and validates the forecasts. Section 3 shows that a stock trading strategy using the distributional forecasts of earnings earns substantial abnormal returns. Section 4 shows how the distributional forecasts can be used to assess and improve management and analyst forecasts of earnings. Section 5 provides robustness checks, discussion of the results, and implications for other research. Section 6 concludes.

2. Estimating and validating distributional forecasts of earnings

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

We start with a brief theoretical grounding on the desirability of distributional forecasting. This theory helps to organize the thinking about existing empirical efforts, and provides clear directions for our own work. The broad motivation for distributional forecasting of earnings is straightforward: future earnings is by definition an uncertain event, and we know from statistical theory that the probability distribution itself is the *complete* mathematical description of the uncertainty in a future outcome.[4] Hence, the most informative earnings forecast is a distributional one.

To more fully explain this core insight, let us formally define what is uncertainty in a future outcome. An uncertain outcome has more than one possible outcome, and attached to each outcome is the probability of its occurrence. For instance, suppose the uncertain outcome is the result of tossing a fair die. Then there are six possible outcomes, with a probability of 1/6 for each outcome. Note that this set of outcomes and their probabilities is precisely the probability

[4] As noted in Dawid (1984), "the only concept needed to express uncertainty is Probability… Consequently, the 'forecasts' I shall be considering will be probability distributions over future events".

distribution of the outcome. This mathematical formulation of uncertainty is well established in probability theory (Dawid 1984; Gneiting 2008). Formally, a probability distribution is uniquely characterized by the cumulative distribution function (CDF) $P(Y \leq y)$.

Establishing the mathematical equivalence between the uncertainty of a future outcome and its distribution allows us to now turn to the central question in this paper - given a set of predictors *X*, what can a researcher say about the future earnings *Earn* of a firm in quarter *q*? 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:

$$Earn = g(X, \varepsilon). \tag{2.1}$$

A simple example of Equation (2.1) is the linear regression framework $g(X, \varepsilon) = X'\beta + \varepsilon$, with ε being normally distributed with mean 0 and variance σ^2. Here, the uncertainty in *Earn* is characterized by the normal distribution $N(X'\beta, \sigma^2)$.

Returning to the general case of Equation (2.1) 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*:

$$P(Earn \leq y|X) = P(\{\varepsilon : g(X, \varepsilon) \leq y\}|X). \tag{2.2}$$

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

Summarizing, 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 object that tells the researcher everything they need to know about future earnings, such as its mean, variance, and all higher moments, probability of meeting or beating certain thresholds, probability of earnings being confined to a certain range, and so on. However, the converse is not true: the distribution is not always recoverable from these statistics, which underlines the primacy of the distribution in describing uncertainty. 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 Distributional forecasts vs. point estimates

While the predictive distribution provides a complete representation of uncertainty, practitioners and researchers have mostly relied on point forecasts when facing uncertain future events. However, this reliance on point estimates can lead to substantially flawed decisions and outcomes. Savage (2003) describes this problem as the "flaw of averages"—the principle that plans based on average values of uncertain quantities are systematically wrong, on average. Savage (2003) illustrates this concept with the cautionary tale of a statistician who drowns while crossing a river that is, *on average*, three feet deep—a vivid reminder that relying solely on average values can be dangerously misleading.[5]

The key insight is that when an outcome $r(X)$ varies non-linearly with the uncertain input variable X—as is often the case in business and financial contexts—the outcome based on the

[5] See https://hbr.org/2002/11/the-flaw-of-averages

mean value of *X* does not generally equal the mean value of the outcome, also referred to as the *expected value* of the outcome in statistics. Savage (2003) formalizes this with the inequality

$$\underbrace{\mathrm{E}[r(X)]}_{\text{expected value of outcome}} \neq r\left(\underbrace{\mathrm{E}[X]}_{\text{mean of } X}\right)$$

for a nonlinear function $r(X)$. If the outcome is of the form $r(X) = c_0 + c_1 X$ for some constants c_0, c_1, then equality holds. In fact, if the equality holds for any random variable *X*, then $r(X)$ is necessarily of this form.

Examples of the uncertain *X* can be earnings, product demand or cost, and examples of the nonlinear outcome $r(X)$ can be profit or payoff. The inequality says that the outcome based on the average input, $r(\mathrm{E}[X])$, is different from the expected outcome $\mathrm{E}[r(X)]$. Hence, to choose the business decision that maximizes the expected outcome, one needs to first calculate $\mathrm{E}[r(X)]$ using the entire distribution of *X*, because the expected outcome cannot always be determined from just the point $\mathrm{E}[X]$.

To illustrate, we present a simple example showing why investment decisions based on point estimates (means) of earnings can be suboptimal compared to those based on a distributional forecast of earnings. Assume that:

- Earnings (EPS) surprise, *SUR*, is uniformly distributed between −3 to 3 cents, i.e., $P(SUR = y) = 1/7$ for $y \in -3¢,\ -2¢, -1¢, 0¢, +1¢, +2¢, +3¢$.
- Stock return as a function of earnings surprise is asymmetric (i.e., nonlinear) in the following way:

$$r(SUR) = \begin{cases} SUR & if\ SUR > 0 \\ 0 & if\ SUR = 0 \\ 2 \cdot SUR & if\ SUR < 0 \end{cases}.$$

Clearly, the mean earnings surprise $\mathrm{E}[SUR]$ is zero in this example, and therefore the investment decision based on mean surprise is to neither buy or sell because the corresponding payoff $r(\mathrm{E}[SUR])$ is zero. However, the expected payoff $\mathrm{E}[r(SUR)]$ based on the earnings distribution and stock return function is $-0.86¢$, which is negative despite the expected/mean earnings surprise being zero. This negative expected payoff arises because the negative surprises (which occur with equal probability as positive surprises) are penalized more heavily. Thus, the optimal investment decision is to short the stock even though the mean earnings surprise is zero.

Summarizing, this example provides a clear illustration of how using information from distributional forecasts helps investors make better decisions when the payoff is a non-linear function of earnings. In Section 3, we formally apply this insight to the earnings announcement context, where stock returns react asymmetrically to positive and negative earnings surprises. Specifically, we devise a stock trading strategy that harnesses the distributional forecasts of earnings and the asymmetric return response to earn higher abnormal returns relative to strategies based on point estimates alone.

2.3. Obtaining practical distributional forecasts of earnings

As discussed above, the predictive distribution in Equation (2.2) completely characterizes the uncertainty in earnings conditional on *X*. But since $\mathrm{P}(Earn \leq y|X)$ in Equation (2.2) 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.

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, since 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. To avoid restrictive assumptions, prior studies have adopted nonparametric or semi-parametric estimation methods for distributional forecasting. Notably, Konstantinidi and Pope (2016) and Chang et al. (2021) employ a linear quantile regression approach to estimate various quantiles of the predictive distribution.

In this paper, we consolidate these existing efforts by re-estimating them in our sample, and extend them by introducing a distributional machine learning approach called BoXHED (Wang, Pakbin, Mortazavi, Zhao, and Lee 2020; Pakbin, Wang, Mortazavi, and Lee 2025) to estimate Equation (2.2) nonparametrically via gradient boosted trees.[6] BoXHED is an open-source implementation of the tree-boosting method established mathematically in Lee, Chen, and Ishwaran (2021), originally used to estimate hazard functions, which are the conditional density functions for distributions.[7]

Unlike quantile regression methods that estimate one quantile of the distribution at a time, BoXHED estimates the entire conditional distribution in one go without quantile-crossing problems. In addition, it has been shown that tree-boosted methods consistently achieve state-of-the-art performance on structured tabular data like those in our setting (Grinsztajn, Oyallon and Varoquaux 2022). Part of their success can be explained by the fact that tree-boosted methods minimize prediction error in a very effective way: they iteratively grow new tree functions to reduce residual errors from the existing tree ensemble. This approach is known as gradient descent, one of the most powerful methods in optimization theory. The resulting BoXHED estimator flexibly captures nonlinear interactions among the predictor variables to produce

[6] In Section 2.5, we evaluate the accuracy of our BoXHED distributional forecasts and compare them against distributional forecasts obtained from quantile regression methods.
[7] The BoXHED package is available from https://github.com/BoXHED, which includes detailed tutorials. The theoretical framework underlying BoXHED is described in Internet Appendix, Section IA.1.

accurate estimates for the conditional earnings distribution. We describe the details of BoXHED estimation in Internet Appendix, Section IA.1.

Our predicted variable *Earn* for quarter q is defined as EPS_q/P_{q-1} where P_{q-1} is the price per share at the beginning of quarter q. Note that once the BoXHED estimation is complete, and the distributional forecasts of (scaled) *Earn* are produced, it is straightforward to obtain distributional forecasts in terms of unscaled EPS because beginning price is known at the time of the forecast, so EPS is just *Earn* multiplied by price. In addition, in Section 5.1, we provide a robustness check re-doing the BoXHED estimation using unscaled EPS as the predicted variable.

We emphasize that the main thrust of our study is not about searching for the best estimation method for distributional forecasts of earnings. Rather, the main thrust is to convincingly establish the superiority of distributional forecasting relative to point forecasting, and to show some applications of distributional forecasting that are impossible with point estimates. Thus, while we show some results about the relative accuracy of distributional models in Section 2.5, our main interest in the empirical tests and results is between distributional and point forecasts.

2.4. Predictor variables, data, and sample

We aim to estimate distributional forecasts of 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), we employ two sets of predictor variables. First, we include firm-fundamental variables based on accounting information, as these variables form the backbone of most earnings prediction models in the literature. These variables include 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, *Gross Profit*; 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 depreciation expense divided by market capitalization, *WC Accruals*; scaled earnings from four quarters ago, $Earn_{q-4}$. All variables are measured at the beginning of the quarter.

Second, we include analyst-related variables because analysts are primary information providers in capital markets. These variables 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;* and median analyst cash flow forecast before earnings announcements divided by beginning stock price, *CPS*. Please see Appendix A for a full list of variables used in this study, and their definition and source.

Note that while there may be additional variables that can improve the model's accuracy, the pursuit of the "best" possible set of variables is not the primary goal 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.

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.[8] Our test

[8] Observations with missing values in the other variables are included in our sample because our machine learning technique is able to handle missing values. At each split, the algorithm learns a default direction (left or right branch) for missing values based on which choice minimizes loss. This way, missing values don't need to be imputed in advance—the model decides the optimal path for them during training.

period is 2011-2021. For each year in the test period, we use the preceding ten years as the training period to fit Equation (2.2).[9] The fitted distribution is then used to compute out-of-sample estimates for the test year. To remove the effects of outliers, problematic data entries, and marginal firms, we delete observations with (1) stock price < $1 or market capitalization < $5 million; (2) earnings announcement dates in IBES and Compustat that are more than one day apart; or (3) *Earn* > 0.5 in absolute value. The final sample includes 283,356 firm-quarter observations between 2001-2021.

Table 1, Panel A 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 better 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

Internet Appendix Section IA.2 presents some examples of the BoXHED-derived distributional forecasts estimated using the full model with both firm-fundamental and analyst-related variables. Panel A of Section IA.2 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,

[9] 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.

and girth of the tails. Panel B of Section IA.2 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 examples in Section IA.2 are purely for illustrative purposes, they do provide some ground-level feel that the outputs of the BoXHED estimator seem "reasonable".

More systematic evidence about the properties of the estimated distributional forecasts is presented in Table 1, Panel B. Since distributional forecasts are likely to be especially useful when they deviate from well-behaved distributions, we focus on summary statistics for skewness and kurtosis. [10] The average skewness is −1.46, and the average kurtosis is 13.40, which suggests that the distributional forecasts tend to be moderately left-skewed and heavy-tailed.[11] Notably, there is considerable cross-sectional variation in the degree of skewness and kurtosis. The 5th, 25th, 75th, and 95th percentiles of skewness are −4.69, −2.21, −0.34, and 0.42, while those for kurtosis are 2.22, 4.27, 15.53, and 42.62. The pronounced deviation of the distributional earnings forecasts from a normal distribution suggests that considering the full distribution of earnings can provide much more information than point estimates. This is especially relevant for investment and other outcomes that are non-linear functions of earnings, as discussed further in section 5.2.

---

[10] Skewness and kurtosis are calculated as follows:

$$Skewness(Earn_{i,q}) = \frac{\mathbb{E}\left[(Earn_{i,q} - \mu_{i,q})^3\right]}{\sigma_{i,q}^3}, Kurtosis(Earn_{i,q}) = \frac{\mathbb{E}\left[(Earn_{i,q} - \mu_{i,q})^4\right]}{\sigma_{i,q}^4}$$

where $\mu_{i,q}$ is the mean and $\sigma_{i,q}$ is the standard deviation of $Earn_{i,q}$.

[11] We truncate each distribution at the top and bottom 0.1% in calculating the statistics to reduce the impact of extreme outliers. For normal distributions, skewness is 0 and kurtosis is 3. While there is no universal agreement, skewness with absolute value of more than 1, and kurtosis of more than 6 are generally considered "high".

Next, we turn to a more formal evaluation of the quality of the estimated distributional forecasts, using actual realized earnings as the benchmark. Traditional forecasting accuracy metrics, which typically focus on point estimates, are insufficient for this purpose. Accordingly, we turn to the statistical literature for an approach that fits our distributional setting. A particularly intuitive and commonly employed approach for evaluating the accuracy of distributional forecasts is the *continuous ranked probability score* (CRPS). It satisfies a property called propriety that is essential for scoring probabilistic forecasts (Gneiting and Raftery 2007).

The CRPS directly extends the mean squared error (MSE) accuracy measure for point forecasts to distributional forecasts (Gneiting and Raftery 2007). As a 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 (i.e., a degenerate distribution). Its corresponding CDF $P(y \leq t)$ equals

$$I_{[y,\infty)}(t) = \begin{cases} 0 & t \text{ is less than } y \\ 1 & t \text{ is greater than or equal to } y \end{cases} \tag{2.3}$$

A graphical representation of $I_{[y,\infty)}(t)$ is provided as the (blue) solid line in Appendix B, where the CDF is strictly zero below y but jumps to 100% at y, i.e., the realized value is always y.

The integrated squared error between the CDF of the realized outcome and the CDF of the distributional forecast is

$$\int_{-\infty}^{\infty} \left\{I_{[y,\infty)}(t) - \hat{F}(t)\right\}^2 dt, \tag{2.4}$$

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 CDFs for the distributional forecasts. Just like the MSE, the CRPS is always non-negative, and a smaller value indicates a more accurate distributional forecast. Please see Appendix B for a visualization of CRPS, where the CDF of the distributional forecast is depicted by the dashed line, and CRPS is represented by the grey area defined between the CDF lines for the forecast and the actual outcome. This visualization clearly conveys the intuition that a smaller CRPS indicates more accurate distributional forecasts, i.e., a smaller CRPS indicates that the distributional forecast is strongly peaked at the actual outcome.

Note from our discussion above that a point estimate, just like a realized point outcome, is also a (degenerate) distribution. Thus, CRPS can be used to provide a direct comparison between the performance of point forecasts and distributional forecasts, which is both the lynchpin of our investigation and has never been done before in the earnings forecasting literature. Specifically, we use CRPS to compare the performance of the following five types of earnings forecasts, including two point estimates and three distributional forecasts:

i) The forecasted mean of earnings based on an OLS regression model, $\hat{y}_{OLS}$. OLS is the most common existing technology for predicting earnings, and thus it provides a natural benchmark for our distributional forecasts.

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 using nonparametric machine learning methods.

iii) A distributional forecast based on linear quantile regressions using 11 quantiles, with quantile levels in $\{0.01, 0.05, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 0.95, 0.99\}$ following Konstantinidi and Pope (2016).

iv) A distributional forecast based on linear quantile regressions using 150 quantiles, with quantiles evenly spaced between 0 and 1, as implemented by Chang et al. (2021).

v) The estimated distributional forecast using BoXHED. Note that a comparison of the results across ii)-iv) helps to distinguish whether the predictive gains between i) and v) are due to the use of BoXHED, to the use of the distributional approach in forecasting, or both.

We estimate each of the five forecasting methods under two model specifications, as described in Section 2.4: one with only firm-fundamental variables, and one with both firm-fundamental and analyst-related variables. Within each specification, all five methods use the same predictor set.[12] Table 2 presents the CRPS results for each year in the test period. Panels A and B present the results for the models using only firm-fundamental variables, whereas Panels C and D report the results for the models with both firm-fundamental and analyst-related variables.

2.5.1 Results using firm-fundamental variables only (Panels A and B)

Panel A presents the raw CRPS values for each forecasting method, while Panel B reports the percentage reduction in CRPS relative to the baseline OLS mean forecast. Column (1) in both panels reports the CRPS for the OLS mean forecasts $\hat{y}_{OLS}$, which serves as the benchmark. Panel B Column (2) reports the percentage reduction in CRPS when switching over

[12] When estimating the OLS regression models and quantile 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. To ensure distributional consistency, we apply monotonic rearrangement to address any quantile crossing issues.

to the nonparametric mean forecasts $\hat{y}_{BoXHED}$. This switch results in a strict improvement in each year, with an average reduction of 25%, where the improvements are statistically significant at the 1% level for all years in the sample, as denoted by the superscript letter a. Thus, using nonparametric methods and machine learning leads to considerable gains over OLS estimation.

A comparison of Columns (3) and (4) with Columns (1) and (2) reveals that distributional forecasting leads to further gains over point forecasting. The fairly crude 11-quantile regression approach in Column (3) provides only modest improvement over the sophisticated BoXHED point estimates in Column (2). However, increasing the sophistication of the distributional approaches provides considerable gains, where the 150-quintile specification achieves 30% reduction in CRPS compared to the baseline.

Column (5) shows that BoXHED distributional forecasts reduce CRPS by approximately 42% on average relative to OLS. Thus, combining the power of statistical machine learning and the full distributional approach leads to huge gains in forecast accuracy as compared to traditional mean-based estimates. In addition, BoXHED significantly outperforms the 150-quantile regressions in every year (as indicated by the b superscripts).[13] The overall impression is that both sophisticated distributional approaches perform rather well, with BoXHED achieving the best accuracy.

[13] BoXHED is a recent advance in top outlets in statistical machine learning (Wang et al. 2020; Lee et al. 2021). It offers at least three major advantages over quantile-based approaches. First, BoXHED produces an estimator of the entire CDF, giving us all the quantiles of the distribution in one go. By contrast, a separate quantile regression needs to be run for each quantile estimate, and further post-processing is required to avoid quantile crossing. Second, BoXHED naturally provides estimates of both the probability density function as well as CDF, which is useful for settings where the density is needed. One example is in visualizing the probability distribution, and another is for calculating the expected value of a function of the random variable. In contrast, even the CDF is not directly available from quantile regression, which is only able to estimate a finite number (e.g., 150) of quantiles. Obtaining the CDF requires inverting the finite number of quantiles just to get the values of the CDF at those 150 locations. To further obtain the density from the CDF requires applying a smoothing procedure that can be sensitive to the tuning parameters employed for smoothing. Third, BoXHED works even on censored data that appear in survival analysis, for which the standard quantile regression is not valid.

2.5.2 Results using both firm-fundamental and analyst-related variables (Panels C and D)

Panels C and D present the CRPS results for models estimated using both firm-fundamental and analyst-related variables. The broad pattern of results mirrors that in Panels A and B. Specifically, Panel D, Column (2) shows that using machine learning for mean forecasting leads to considerable gains over OLS estimation, with an average reduction of 18%. Most importantly, a comparison of Columns (3) through (5) with Columns (1) and (2) reveals that all distributional forecasting approaches lead to clear gains over point forecasting. Note that even the fairly crude 11-quantile regression approach in Column (3) provides material improvement over the BoXHED point forecasts of the mean in Column (2). Increasing the sophistication of the distributional approaches provides further gains, where the 11-quintile specification achieves 27% reduction in CRPS compared to the baseline, rising to 34% for the 150-quintile specification and 36% for BoXHED distributions.[14]

Summarizing, the empirical evidence in Table 2 is clear: distributional forecasts significantly dominate point forecasts. The magnitude of the improvements from the best distributional forecasts is about 40% reduction in prediction errors relative to the OLS benchmark, pointing to the massive economic importance of using distributional forecasts to enhance accuracy in earnings forecasting.

[14] While the reductions from BoXHED in Column (5) exceed those from the 150-quantile regressions in Column (4) in each year of the test period, the incremental gain from BoXHED relative to the 150-quantile regressions is economically small, on the magnitude of about 2%. This is likely because analysts provide high quality and valuable information for distributional forecasting, which narrows the performance gap between machine learning and linear quantile regression methods. We also estimated 150 quantiles using gradient-boosted quantile regression. This approach delivers a CRPS reduction of 35% relative to the OLS mean forecast (with a raw CRPS of 0.0083), nearly matching the performance of BoXHED distributions. These results suggest that the BoXHED approach is already close to the limit of how well any distributional forecasting can perform.

3. The application of distributional forecasts to stock trading

As discussed in Section 2, distributional forecasts should be especially valuable to stock investors if they allow them to capitalize on non-linearities in earnings-based return outcomes. To harness this insight, we develop a stock trading strategy around earnings announcements. Because successful trading in this setting requires the most timely information, we use distributional forecasts from the model incorporating both firm-fundamental and analyst-related variables. Our distributional forecasts are available in real time, and are out-of-sample, which naturally allows for a trading strategy implementation.

We emphasize that maximizing abnormal returns is not a major goal of our investigation. We view the abnormal return evidence as illustration of the utility of the distributional forecast approach rather than as an exercise in seeking optimal trading rules.[15] Since the BoXHED distributional forecasts perform best in the validation tests above, we adopt them as the main specification going forward. We emphasize that our main interest is in the performance of the distributional forecasts approach against point estimates, rather than optimizing performance within the available options for distributional forecasts.[16]

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). In addition, Koh, Matsumoto, and Rajgopal (2008) document a pronounced asymmetry in return responses to earnings surprises, where returns are much higher in magnitude for negative earnings surprises. Given this nonlinearity, the expected payoff from

[15] Hence, the trading strategy we propose may or may not be the best possible. We leave the refinement of this strategy to future research.
[16] Untabulated statistics reveal that our results are qualitatively similar when using alternative distributional forecast methods based on quantile regressions.

buying or shorting a stock around earnings announcement can only be properly calculated from knowledge of the whole earnings distribution, as explained in Section 2.2.

Recall that the underlying intuition is illustrated in the numerical example in Section 2.2, where earnings (EPS) surprise, SUR, is uniformly distributed between −3 and 3 cents and the stock response to positive vs. negative earnings surprises is asymmetric. The expected earnings surprise is zero, which implies neither buy or sell if investors make stock trading decisions based on point estimates only. However, the optimal trading decision requires computing the expected payoff of buying the stock using the full distribution of SUR and the non-linear return function $r(\cdot)$. This calculation yields a negative expected payoff of $-0.86¢$, so the optimal investment decision is to sell.

To model this investment decision around earnings announcements in practice, let $P_{EPS}(y|X)$ be the conditional probability distribution of the EPS, c the consensus forecast, and $r(SUR)$ is the expected stock response function conditional on earnings surprise $SUR = y - c$. Then the distribution for earnings surprise is simply $P(SUR|X) = P_{EPS}(c + SUR|X)$.[17] It follows that the expected payoff from buying the stock is

$$\Pi(X) = \mathbb{E}\{r(SUR)|X\} = \sum_{SUR=-\infty}^{\infty} r(SUR) \cdot P(SUR|X). \tag{3.1}$$

A trading strategy can then be formed by taking a long position in the top decile (those with the most positive values of expected payoff) and a short position in the bottom decile.

It is important to note that the expected payoff $\Pi(X)$ equals a multiple of the mean earnings (surprises) if and only if the stock response function $r(SUR)$ is linear in earnings surprise, in which case it is not necessary to know the distributional forecast. However, because

[17] For each EPS distribution, because c, the consensus forecast, is known at the time of our forecast, the probability distribution of earnings surprise, $SUR = y - c$, is simply a shift of the EPS distribution by a constant c.

stock returns are in fact asymmetric in earnings surprise (hence r is nonlinear), we need the whole distributional forecast in order to calculate the expected payoff.

We also need the stock response function $r(\cdot)$, which we approximate by partitioning the range of possible earnings surprises into seven regions, and use the average historical returns within each region as a proxy for the expected return reaction for that region. The specific regions are $SUR \leq -3$, $SUR = -2$, $SUR = -1$, $SUR = 0$, $SUR = 1$, $SUR = 2$, and $SUR \geq 3$, where the numbers refer to EPS in cents. The corresponding approximation of the expected payoff (3.1) is

$$\begin{aligned}\Pi = {} & P(SUR \leq -3) \cdot \underline{r}(-3) \\ & +P(SUR = -2) \cdot r(-2) + P(SUR = -1) \cdot r(-1) \\ & +P(SUR = 0) \cdot r(0) \\ & +P(SUR = 1) \cdot r(1) + P(SUR = 2) \cdot r(2) \\ & +P(SUR \geq 3) \cdot \overline{r}(3)\end{aligned} \tag{3.2}$$

Here, for example $P(SUR = -1)$ is the estimated probability that earnings surprises is −1 cent per share, i.e., miss consensus c by 1 cent.[18] The expected return under this outcome can be approximated by the prior-year's average stock return when earnings miss consensus by one cent, which is denoted as $r(-1)$. Likewise, the prior-year's average stock return for missing consensus by three cents or more is denoted as $\underline{r}(-3)$, and the average return for beating by three cents or more is denoted as $\overline{r}(3)$.[19]

We create a measure, *Payoff*, based on Equation (3.2) for all earnings announcements in the test period. Larger values of this expected payoff measure indicate that the firm's abnormal

---

[18] Since *SUR* is continuous in our estimation, we define *SUR*= −1 if *SUR* is in the range of (−1.5, 0.5]. Please also note that we assume $r(0) = 0$, i.e. the stock return is zero when $SUR = 0$, consistent with findings from extant research.

[19] It is important to reiterate that knowledge of the earnings distribution is required for calculating even these approximations. While one could attempt to separately model the probability of missing by 3 cents or more, the probability of missing by 2 cents, and so on, the probabilities from these separate models may not sum to one. Furthermore, since such approaches are essentially indirect attempts at approximating the earnings distribution, one should simply model the exact distribution directly.

returns will be higher during the earnings announcement window.[20] We sort and bin the measure into deciles, where the decile cut-offs are determined by the deciles of the payoff measure for the prior year. We then form a trading strategy that takes a long position in the top decile (those with the highest values of *Payoff*) and a short position in the bottom decile (those with the lowest values of *Payoff*). Finally, we compute the cumulative abnormal returns (CAR), specifically market-adjusted CAR (denoted as *Mkt-adj CAR*), for the portfolios over the three trading days surrounding the earnings announcement date.

Note that while our return windows are short, it is still important to control for the effect of systematic risk and other return factors as an explanation for returns, e.g., Engelberg, McLean, and Pontiff (2018) show that abnormal returns are especially concentrated at earnings announcements, generating returns six times those realized during non-earnings announcement windows. We use the Fama and French (2016) five-factor model to control for known return factors and calculate a second measure of CARs after controlling for these factors, denoted as *FF5 CARs*.

The results in Table 3, Panel A reveal that the average *Mkt-adj CAR* (*FF5 CAR*) is 22 (16) bps for firms in the top portfolio, and –69 (–68) bps for the bottom portfolio, all statistically significant. Thus, the abnormal returns are concentrated in the bottom portfolio, consistent with existing research that shows that the penalty for missing the consensus is much greater than the premium for beating the consensus, e.g., Koh, Matsumoto, and Rajgopal (2008). The return on the hedge portfolio is 91 (83) bps and is highly statistically significant. Note that this hedge

[20] In addition, we have tested a simpler measure from our distributional forecasts. This measure reflects the differential probability of beating (missing) analyst expectations by *N* cents per share at earnings announcements. Using this alternative measure earns significant abnormal returns on the magnitudes 60 bps over the three day earnings announcement window (or about 50% annualized). For further details, please refer to Internet Appendix, Section IA.3.

return is rather substantial economically, corresponding to an annualized return of 76% (69%) using the convention of 250 trading days.[21]

Figure 1 presents an expanded view of the results in Panel A of Table 3 by plotting the cumulative abnormal returns and the 95% confidence intervals for the expected payoff portfolios for each year in our test period 2011-2021. The findings suggest that the hedge portfolios are able to generate positive returns in most years in the sample except 2015 and 2017 (both close to zero), 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 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.[22]

Summing up, the results in Table 3, Panel A and Figure 1 suggest that forming trading portfolios that exploit the information conveyed by the distributional forecasts of earnings yields abnormal returns on the magnitude of 90 bps over the three-day earnings announcement window.

---

[21] 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 goal 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; Lyle and Yohn 2024). 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. Finally, our hedge returns compare favorably to results from existing research for comparable settings. For example, Johnson, Kim, and So (2020) find that firms with high incentives to manage market expectations exhibit a V-shaped pattern of abnormal returns leading to earnings announcements, with negative abnormal returns on the magnitude of 50 bps in the *month* before the earnings announcement, and 60 to 80 bps in the month of the earnings announcement.

[22] For example, firms that missed analyst forecasts by a larger margin in 2020 might be able to receive more financial aid from the government, which leads to muted or even positive stock market reactions. To provide a more quantitative estimate on the year 2020 effect, we rerun the analyses excluding 2020, and find that the *Mkt-adj CAR* (*FF5 CAR*) hedge return jumps from 91(83) bps to 120 (105) bps.

These returns seem economically substantial, are fairly robust, and remain largely unchanged after controlling for known return factors.

However, an important question remains: do trading portfolios based on distributional forecasts of earnings yield greater abnormal returns than those based on the mean, which is the archetypal point estimate of future earnings? We perform two analyses to investigate this question. Our first analysis studies the returns from hedge portfolios formed based on mean forecasts. If the returns are below those for the trading strategy based on the payoff measure, the implication is that leveraging distributional forecasts allows investors to make better investment decisions.

Specifically, for each firm-quarter, we calculate the expected value of the earnings surprise by subtracting the consensus from the mean 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 decile (those with the highest expected surprises) and a short position in the bottom decile (those with the lowest expected surprises). Finally, we calculate CAR for the portfolios over the three trading days around the earnings announcement date. We use two versions of mean forecasts, OLS forecasted mean and BoXHED-derived mean. Note that the means from the BoXHED distributions allow us to hold the machine-learning approach constant for an apples-to-apples comparison of mean-based vs. full distributional forecast-based trading strategies.

Table 3, Panels B and C show that the hedge returns, *Mkt-adj CAR* (*FF5 CAR*), based on mean forecasts are statistically and economically significant at 48 (50) bps for the BoXHED mean and 47 (29) bps for the OLS mean, confirming the intuition that predicting the mean captures a lot of information about future earnings. These returns, however, are only about half the hedge returns of 91 (83) bps based on the expected payoff measure. This result suggests that using the full distributional forecast of earnings is indeed substantially more useful than using the mean of the earnings surprise.

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

$$CAR_q = \beta_1 Payoff\_decile_q + \beta_2 Mean\_decile_q + FixedEffects + \varepsilon. \tag{3.3}$$

The variable $Payoff\_decile_q$ ($Mean\_decile_q$) represents the decile ranks of the payoff (expected surprise) measure within the current year, scaled down to a value between 0 and 1. As before, we use two versions of mean forecasts—OLS mean and BoXHED mean—to calculate expected surprises. We include industry and year fixed effects and cluster standard errors by firm and quarter. If the distributional forecasts provide incremental information beyond the expected value and enable investors to earn higher abnormal returns, we would expect $\beta_1 > 0$.

Table 4 presents the results for the estimation of Equation (3.3). Columns (1) – (3) confirm that the payoff measure and the two expected surprise measures are positively and statistically significantly associated with returns around earnings announcements when used independently, where the expected payoff measure earns the highest returns. Columns (4) and (5) indicate that while the coefficient for the payoff measure remains positive and statistically significant when we control for the effect of mean surprises (coefficients=47.166 and 48.777, *t*-statistics=3.42 and 3.80), the coefficient for the two versions of mean-based surprise becomes statistically insignificant. The conclusion is that the expected payoff measure based on full distributional forecasts subsumes the information conveyed by the mean-based measure, and provides information incremental to it. More broadly, the findings in Tables 3 and 4 suggest that there are important applications for which the distributional forecast approach dominates mean-based forecast approaches.

4. Evaluating the quality of management and analyst forecast ranges

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

4.1. Prediction ranges for earnings forecasts

Distributional forecasts can be used to produce a *prediction range* within which an upcoming earnings number would fall. Critically, we can attach a degree of certainty to this interval. For example, for a given firm-quarter we can provide a prediction range that has an 80% or 95% chance of covering earnings.[23] 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 point estimate, 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 managers 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

[23] 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 lie.

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[24]

$$\hat{P}\left(l \leq EPS_q \leq u \middle| X\right) = \hat{P}\left(\frac{l}{P_{q-1}} \leq Earn_q \leq \frac{u}{P_{q-1}} \middle| X\right) = \alpha, \tag{4.1}$$

we need to find the value l corresponding to $\hat{P}\left(EPS_q \leq l \middle| X\right) = \alpha/2$, and the value u corresponding to $\hat{P}\left(EPS_q \geq u \middle| X\right) = 1 - \alpha/2$. 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 (4.1) 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 calibration plot* for our prediction ranges, which is an intuitive graphical way to reflect the accuracy of our prediction ranges. The plot is constructed as follows. The horizontal axis represents the desired coverage levels α for our prediction ranges.[25] 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 in Figure 2, we compute 80% coverage prediction ranges for every firm-quarter in our sample. Then, we compute the percentage of actual earnings realizations that fell within their corresponding 80% prediction

[24] 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.

[25] 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.

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 in 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.

4.2. Imputing the coverage probabilities of management forecasts

We now use our distributional forecasts 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:

$$\hat{P}(EPS_q \leq upper\_bound|X) - \hat{P}(EPS_q \leq lowest\_bound|X). \tag{4.2}$$

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.[26]

[26] 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 zero mean and a standard deviation of 0.25, the range $[-0.5, +0.5]$ will have a 95% coverage.

4.2.1. 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 a unimodal distribution where most of the probability mass is bound between $0.50 and $1.00 in EPS. Panel A also includes the Delta management range forecast for that quarter, and its corresponding coverage probability. Notice that the management forecast 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.

4.2.2.Results for the full management forecast sample

The box plot in Figure 4 Panel A 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 5th and the 95th percentiles of the distribution of coverage probabilities across firm-quarters. An inspection of Figure 4 Panel A reveals a stark message: management forecast ranges appear quite narrow, where the *median* value of coverage probabilities is only 29%, 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 on average 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. Note also that 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 and external validation, in the Call et al. (2024) 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. Managers are likely motivated by a number of incentives beyond accurate forecasting (Aghamolla and Smith 2023). 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 unobservable intent are often difficult to answer, in this case we actually have some suggestive evidence. The Call et al. (2024) 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 Panel A - and perhaps most importantly - note that the degree of miscalibration by managers exhibits great cross-sectional (and likely over-time) variation. The 5/25/75/95 percentile of coverage probability is 6%/17%/44%/67%, respectively. Since our distributional forecasts are available in real time, and can be tailored to the firm-quarter level, they can serve as a powerful 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 indicates that managers have much stronger incentives to avoid negative earnings surprises as compared to positive earnings surprises, see 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. 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. For example, future research can investigate the role of manager fixed effects in contributing to miscalibration, and how such fixed effects drive differential manager propensities for making major decisions such as capital investments and pursuing mergers and acquisitions.

4.3. 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 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 box plot in Figure 4 Panel B summarizes the imputed coverage probabilities for analyst forecast ranges. Compared to the box plot for management range forecasts in Panel A, the coverages here are considerably better but are still low in absolute terms. For example, the median coverage probability is 46%, and the 75th percentile is 59%.[27] 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 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.

The availability of probability coverage for analyst forecast ranges also allows us to extend the results for earnings announcement returns in Sections 3. Notice that probability coverage is essentially a (reverse) measure of underestimation of uncertainty, where lower probability coverage likely indicates that the market underestimates the uncertainty in the forthcoming earnings announcement. Since at least some of this underestimation is resolved at the announcement, the absolute value of announcement returns is likely increasing in the underestimation of uncertainty. We confirm this conjecture in untabulated results using regression of |FF5 CARs| on uncertainty underestimation. This finding further illustrates the utility of the distributional forecast approach, and confirms earlier impressions that our distributional metrics help to identify underestimation of information uncertainty that will be resolved at the earnings announcement.

[27] 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 44%; 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 still holds.

4.4 Summary of the evidence for management and analyst forecasts

The evidence in Figure 4 suggests that both management and analyst forecast ranges are too narrow as compared to the variability of actual earnings realizations. Both types of ranges 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.

5. Robustness checks, discussion of the results, and implications for other research

In this section, we provide evidence from alternative test specifications and robustness checks. We also discuss the implications of the results, and offer suggestions for future research. A major theme of our discussion is that the distributional forecast approach has broad utility but seems especially promising for settings with non-linearities in earnings-based outcomes.

5.1 Robustness checks and alternative specifications

Our first robustness check is whether it makes a difference that the predicted earnings variable is scaled or unscaled, EPS_q/P_{q-1} or EPS_q. As is probably clear, once the BoXHED model is estimated, these two variables are equivalent because lagged price is known at the time of the forecast, so one variable is a simple transformation of the other. We re-estimate the BoXHED model on unscaled EPS and find that the main results remain qualitatively the same.

We also compare the distributional forecast properties between samples with and without analyst coverage. Recall that our primary sample consists of firms with analyst coverage, which tend to be larger and have a better information environment. We use analyst forecast information as input into our BoXHED estimation because analyst forecasts themselves are a major source of information about future earnings. A natural question is whether BoXHED distributional forecasts are well calibrated for firms without analyst coverage.

To investigate this conjecture, we re-estimate the BoXHED model using the same full set of variables but using the full Compustat/CRSP sample, including firms with and without analyst coverage. For firms without analyst coverage, we set the analyst variables as missing values. This is feasible because one of the strengths of BoXHED is that it can incorporate missing values into its estimation without discarding observations. We then conduct validation tests similar to those in Section 2, and find that the distributional forecasts map quite well into actual earnings realizations.[28] Overall, this evidence suggests that our distributional forecast approach is useful for both firms with and without analyst coverage, highlighting its broad applicability.

5.2 Discussion of the results and implications for other research

We identify at least three broad areas for extending and enriching our approach. The first is more circumscribed and practice-oriented, essentially fine-tuning our approach to better meet the needs of different users. For example, investors interested in earning superior returns can use more variables in estimating the BoXHED model, develop different predictive summary statistics, consider trading costs explicitly, and possibly use more complex trading strategies.

Second, we believe that the distributional forecasting of earnings opens up great new research possibilities, especially for settings with pronounced non-linearities in earnings-based outcomes. As discussed in Section 3, our distributional forecasts enable investors to earn abnormal returns at earnings announcements by capitalizing on the pronounced asymmetry of returns for positive vs. negative earnings surprises. This same intuition can be extended to many

[28] We restrict our attention to the validation tests of this paper because the rest of the tests rely on having analyst forecasts, so they are not applicable (e.g., the expected payoff depends on earnings surprise, which is defined with respect to the analyst consensus forecast).

other settings in accounting and finance which have pronounced non-linearities in earnings-based outcomes. For example, bankruptcy risk is strongly dependent on earnings but the relationship is highly non-linear, which implies that distributional forecasting of earnings is likely to be important for such settings. Similar considerations apply to management bonus compensation around earnings-based thresholds, and to the pricing of securities that display option-like characteristics with respect to earnings outcomes.

Moving beyond the domain of earnings forecasting, it is also possible to 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. For example, accounting accruals, such as Accounts Receivable, can be thought of as current point-estimates of future events like cash collections from customers. The technology in this study can be used to produce distributional forecasts of such future events, and the characteristics of these distributions can be used as proxies for the characteristics of the underlying accruals, e.g., the standard deviation of the distributional forecast of cash collections from customers can be used as a (inverse) proxy for the quality of the Accounts Receivable accrual.

6. Conclusion

Existing earnings forecasts are typically expressed as point estimates. However, the future earnings number is unknown ex-ante, so it is inherently a probability distribution over all possible earnings outcomes. Therefore, the most informative earnings forecast is a distributional one. We implement this intuition by using linear quantile regressions and a statistical machine learning approach called BoXHED to nonparametrically produce out-of-sample distributional forecasts of earnings right before earnings announcements.

We examine the utility of the distributional forecast approach along three dimensions. First, we establish that distributional forecasts clearly dominate point estimates in forecasting earnings. Specifically, BoXHED distributional forecasts and two distributional forecast alternatives based on the existing literature (Konstantinidi and Pope 2016 and Chang et al. 2021) all deliver lower forecasting errors than point estimates of the mean based on OLS and BoXHED. Among distributional forecasts, BoXHED forecasts are statistically significantly better than the two alternatives.

Second, we illustrate the utility of distributional forecasts for making trading decisions by leveraging the non-linear response of stock returns to earnings. We use distributional forecasts to rank firms based on the expected payoff from trading stocks around earnings announcements. Hedge portfolios going long (short) on this expected payoff measure earn abnormal returns on the magnitude of 90 bps over the three-day earnings announcement window during the 2011-2021 test period. In addition, the abnormal returns from using distributional forecasts are about double the returns from using point estimates, specifically means from OLS and BoXHED. Third, we use the distributional forecasts to show that management and financial analyst forecast ranges are too narrow, vastly underestimating the variability of future earnings outcomes. Since our distributional forecasts are available in real time at the firm-quarter level, they provide calibrated forecast alternatives to managers, analysts, and investors.

The combined impression from the totality of this evidence is that there are important applications for which distributional forecasts of earnings dominate point forecasts. In addition, the use of distributional forecasts opens up great new opportunities for future research, especially for settings with pronounced non-linearities in earnings-based outcomes, e.g., bankruptcy assessment, threshold-based compensation, option-style security pricing, 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.

Aghamolla, C., and K. Smith., 2023. Strategic complexity in disclosure. Journal of Accounting and Economics 76(2-3), 101635.

Baginski, S., and J. Wahlen, 2003. Residual income risk, intrinsic values, and share prices, *The Accounting Review* 78 (1), 327-351.

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.

Billingsley, P. 1995. Probability and Measure. 3rd edn. New York: Wiley.

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., 2024. Corporate managers' perspectives on forward-looking guidance. Journal of Accounting and Economics 78 (2024). 101731.

Chang, W.J., Monahan, S.J., Ouazad, A. and Vasvari, F.P., 2021. The higher moments of future earnings. The Accounting Review, 96 (1), 91-116.

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.

Dawid, A.P., 1984. Present position and potential developments: Some personal views; Statistical theory/The prequential approach. Journal of the Royal Statistical Society: Series A (General) 147(2), 278-290.

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.

Donelson, D., and R. Resutek, 2015, The predictive qualities of earnings volatility and earnings uncertainty, Review of Accounting Studies 20, 470-500.

Durrett, R., 2019. Probability: Theory and examples. Cambridge University Press.

Engelberg, J., McLean, R.D., Pontiff, J., 2018. Anomalies and News. The Journal of Finance 73, 1971–2001.

Fama, E.F., French, K.R., 2016. Dissecting Anomalies with a Five-Factor Model. Review of Financial Studies 29, 69–103.

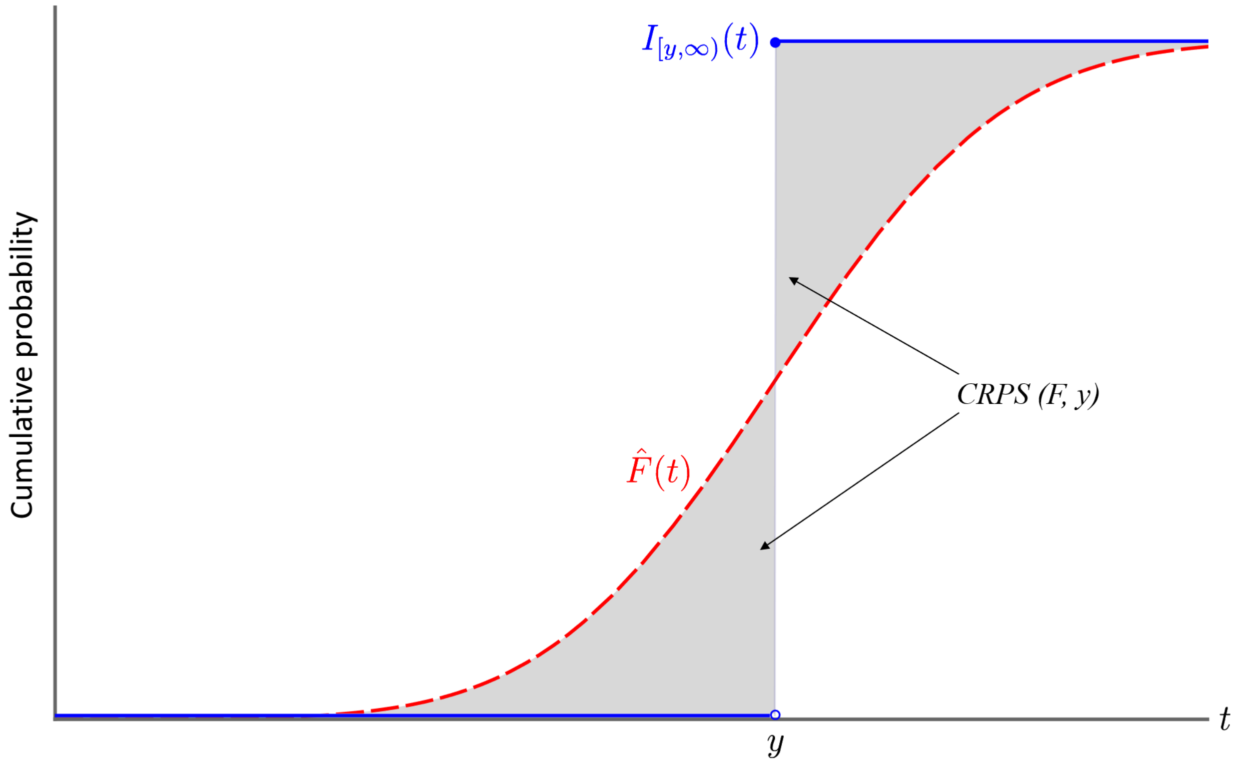
Feller, W. 1971. An Introduction to Probability Theory and Its Applications. Vol. II, 2nd edn. New York: Wiley.
Feller, W., 1991. An introduction to probability theory and its applications, Volume 2. John Wiley & Sons.
Frazzini, A., Israel, R., Moskowitz, T.J., 2018. Trading costs. Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3229719
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., 2008. Probabilistic forecasting. Journal of the Royal Statistical Society Series A: Statistics in Society 171 (2), 319-321.
Gneiting, T., Katzfuss, M., 2014. Probabilistic forecasting. Annual Review of Statistics and Its Application 1 (1),125-151.
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.
Grinsztajn, L., Oyallon, E., Varoquaux, G., 2022. Why do tree-based models still outperform deep learning on typical tabular data? Advances in neural information processing systems, 35, 507-520.
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.
Heyde, C. C. 1963. On a property of the lognormal distribution. Journal of the Royal Statistical Society Series B: Statistical Methodology 25 (2), 392-393.
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.
Johnson, T.L., Kim, J., So, E.C., 2020. Expectations Management and Stock Returns. The Review of Financial Studies 33, 4580–4626.
Joos, P., Piotroski, J.D., and Srinivasan, S., 2016. Can analysts assess fundamental risk and valuation uncertainty? An empirical analysis of scenario-based value estimates. Journal of Financial Economics 121 (3), 645-663.
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.
Konstantinidi, T., Pope, P.F., 2016. Forecasting risk in earnings. Contemporary Accounting Research 33(2), 487-525.
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.
Lyle, M.R., Yohn, T., 2024. Measuring portfolio gains: The case of earnings announcement trading signals. The Accounting Review 99 (4), 315-338.

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.
Pastor, L., and P. Veronesi, 2003, Stock valuation and learning about profitability, The Journal of Finance 58 (5), 1749-1790.
Pakbin, A., Wang, X., Mortazavi, B. J., Lee, D. K. K., 2025. BoXHED2.0: Scalable boosting of dynamic survival analysis. Journal of Statistical Software 113, 1-26.
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.
Savage, S.L., 2003. Decision Making with Insight. Belmont, CA: Duxbury Press.
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.
Stigler, S.M., 1975. The transition from point to distribution estimation. Bull Int Stat Inst 46, 332-340.
Villani, C., 2021. *Topics in optimal transportation* (Vol. 58). American Mathematical Soc..
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: Variable definitions

Variable	Definition
Analyst	The number of analysts following the firm.
BTM	Book-to-market ratio.
CFO	Beginning quarterly net operating cash flow divided by beginning market capitalization.
Consensus	The median consensus analyst forecast before earnings announcements divided by beginning price.
CPS	The median cash flow forecast before earnings announcements divided by beginning price.
Dispersion	The standard deviation of earnings forecasts before earnings announcements divided by beginning price.
Earn	Actual earnings per share divided by stock price at the beginning of the quarter.
$Earn_{q-4}$	*Earn* four quarters ago.
FF5 CAR	Cumulative abnormal returns adjusted for Fama-French five factors over the three-day window starting from one trading day before earnings announcements.
Gross Profit	Revenue minus cost of goods sold divided by beginning market capitalization.
Mean_decile	Decile ranks of the expected surprise predicted by BoXHED. This variable is scaled down to a range from 0 to 1.
Mkt-adj CAR	Cumulative market-adjusted abnormal returns over the three-day window starting from one trading day before earnings announcements.
Payoff	$\prod = P(SUR \leq -3) \cdot \underline{r}(-3) + P(SUR = -2) \cdot r(-2) + P(SUR = -1) \cdot r(-1) + P(SUR = 0) \cdot r(0) + P(SUR = 1) \cdot r(1) + P(SUR = 2) \cdot r(2) + P(SUR \geq 3) \cdot \overline{r}(3)$. Here, for example, $P(SUR = -1)$ is the estimated probablity that earnings surprises is -1 cent per share. $r(-1)$ is the expected return under this outcome, approximated by the prior three year's average stock return for missing consensus by one cent.
Payoff_decile	Decile ranks of the expected payoff measure (*Payoff*). This variable is scaled down to a range from 0 to 1.
R&D	Quarterly research and development expenses divided by beginning market capitalization.
Revenue	Quarterly revenue divided by beginning market capitalization.
SG&A	Quarterly selling, general and administrative expenses scaled by beginning market capitalization.
Size	Natural log of beginning market capitalization.
WC Accruals	Quarterly changes in non-cash working capital accounts plus depreciation expense divided by beginning market capitalization.

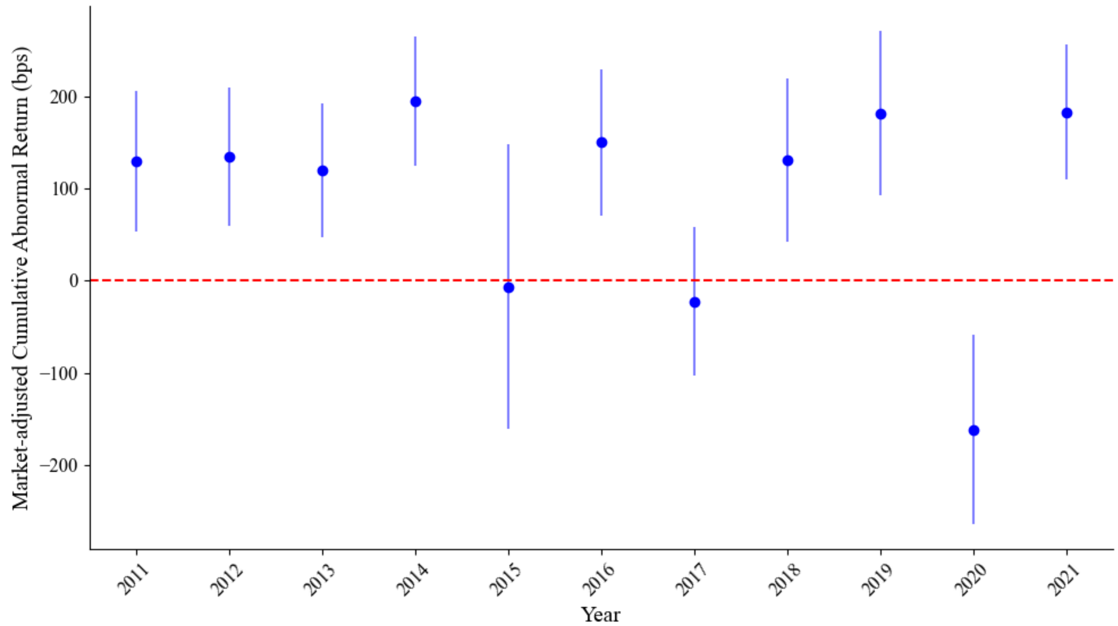
Appendix B: Visualization of the Continuous Ranked Probability Score (CRPS)



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 $\hat{F}(t)$, and the red dashed line is the forecasted CDF $I_{[y,\infty)}(t) - \hat{F}(t)$. In loose terms, CRPS represents the (squared) gray area. Technically, it is calculated by squaring the difference between the two curves and integrating over *t*.

Figure 1: Cumulative Abnormal Returns (CAR) for the expected payoff portfolios by year

Panel A: Market-adjusted CAR for the difference between top and bottom deciles



Panel B: FF5 CAR for the difference between top and bottom deciles

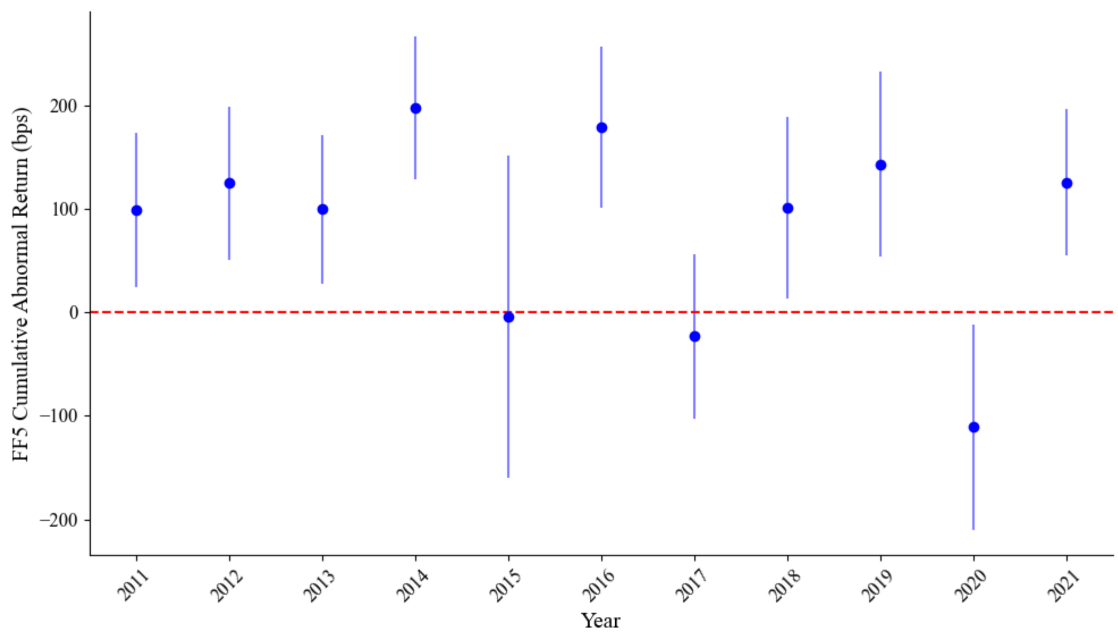


Fig. 1. This figure presents the cumulative abnormal returns for the expected payoff portfolios for each year in our test period 2011-2021. Panels A and B report the *Mkt-adj CAR* and *FF5 CAR* difference between the top and bottom decile of the expected payoff measure, which corresponds to returns generated by a hedge portfolio that takes a long position in the top decile and a short position in the bottom decile. The 95% confidence interval surrounding each cumulative abnormal return difference is also presented.

Figure 2: Actual vs. predicted coverage for prediction ranges

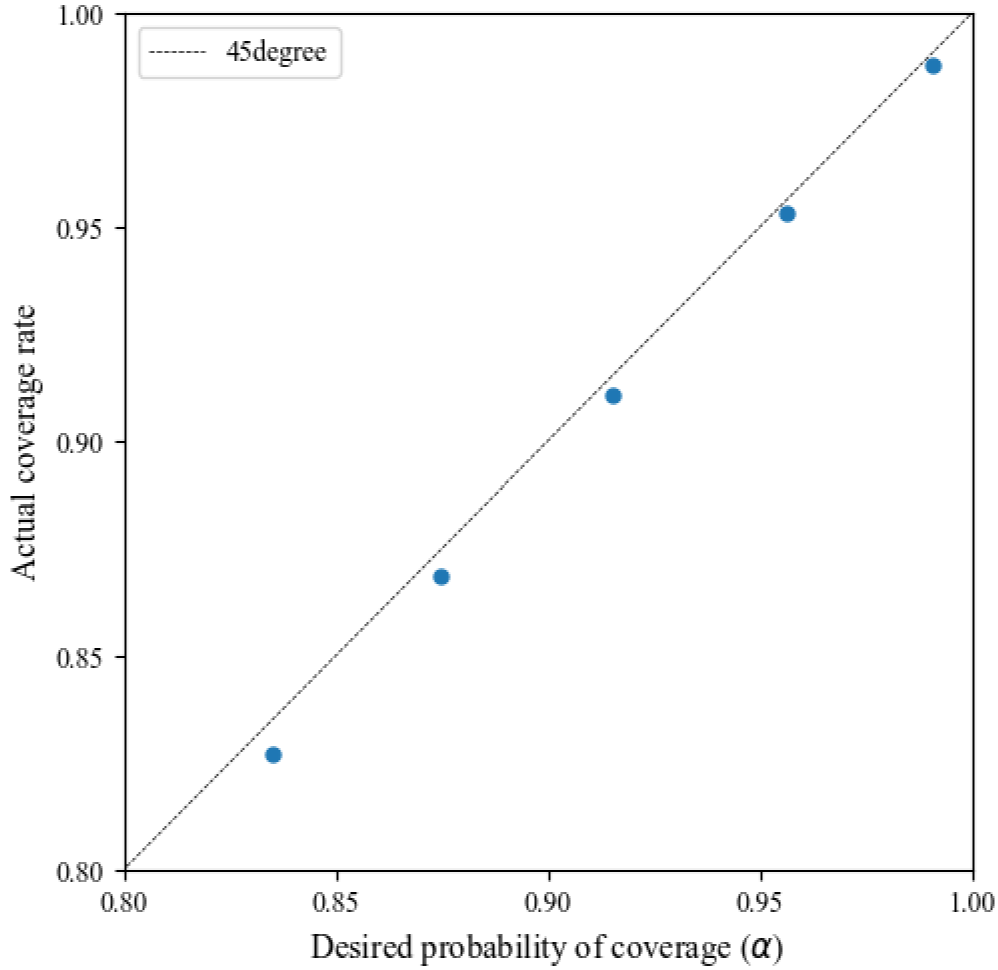
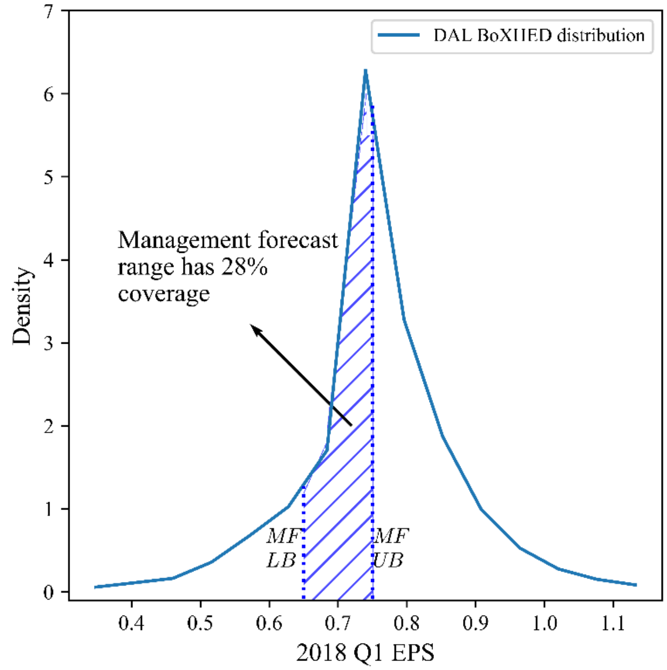


Fig.2. Fraction of prediction ranges that contain the actual earnings number, plotted against predicted coverage. For test period 2011-2021.

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

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



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

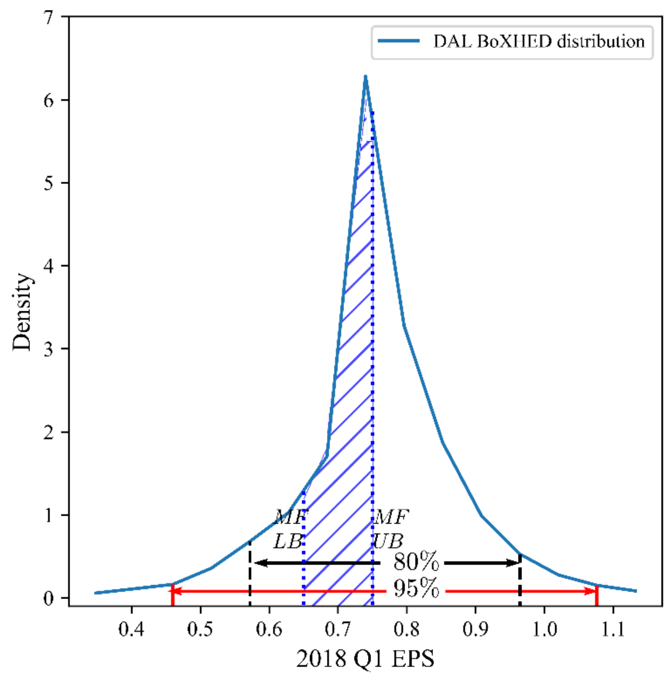


Fig. 3. Predicted coverage for Delta Air Lines 2018 Q1 earnings per share (EPS). Panel A plots the predicted coverage probability of the management forecast range. Panel B plots the management forecast range as well as the prediction ranges that have an 80% or a 95% chance of covering the actual earnings.

Figure 4: Coverage probabilities of management forecasts and analyst forecasts

Panel A: Coverage probabilities of management forecasts

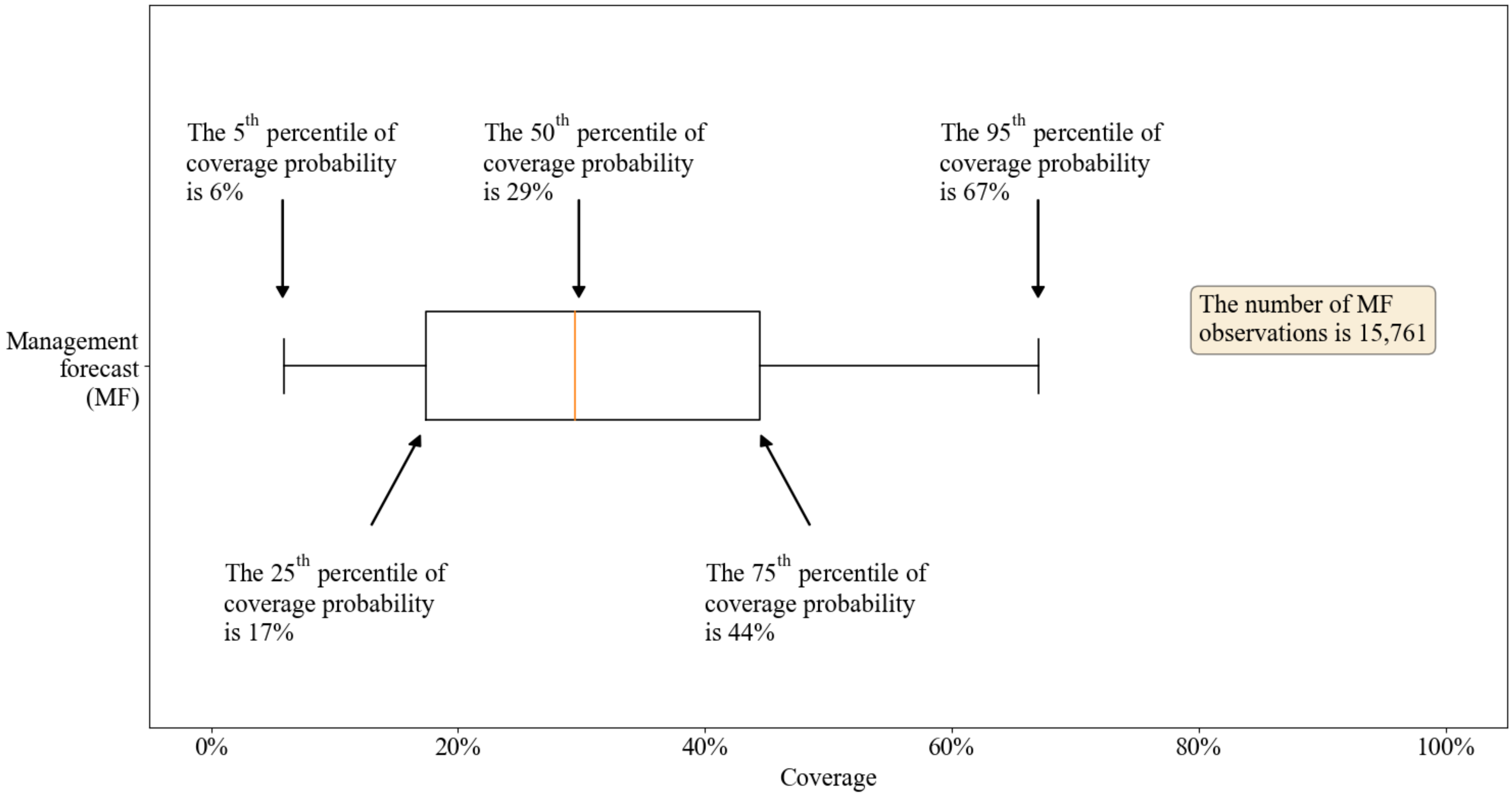


Figure 4 (continued)

Panel B: Coverage probabilities of analyst forecasts

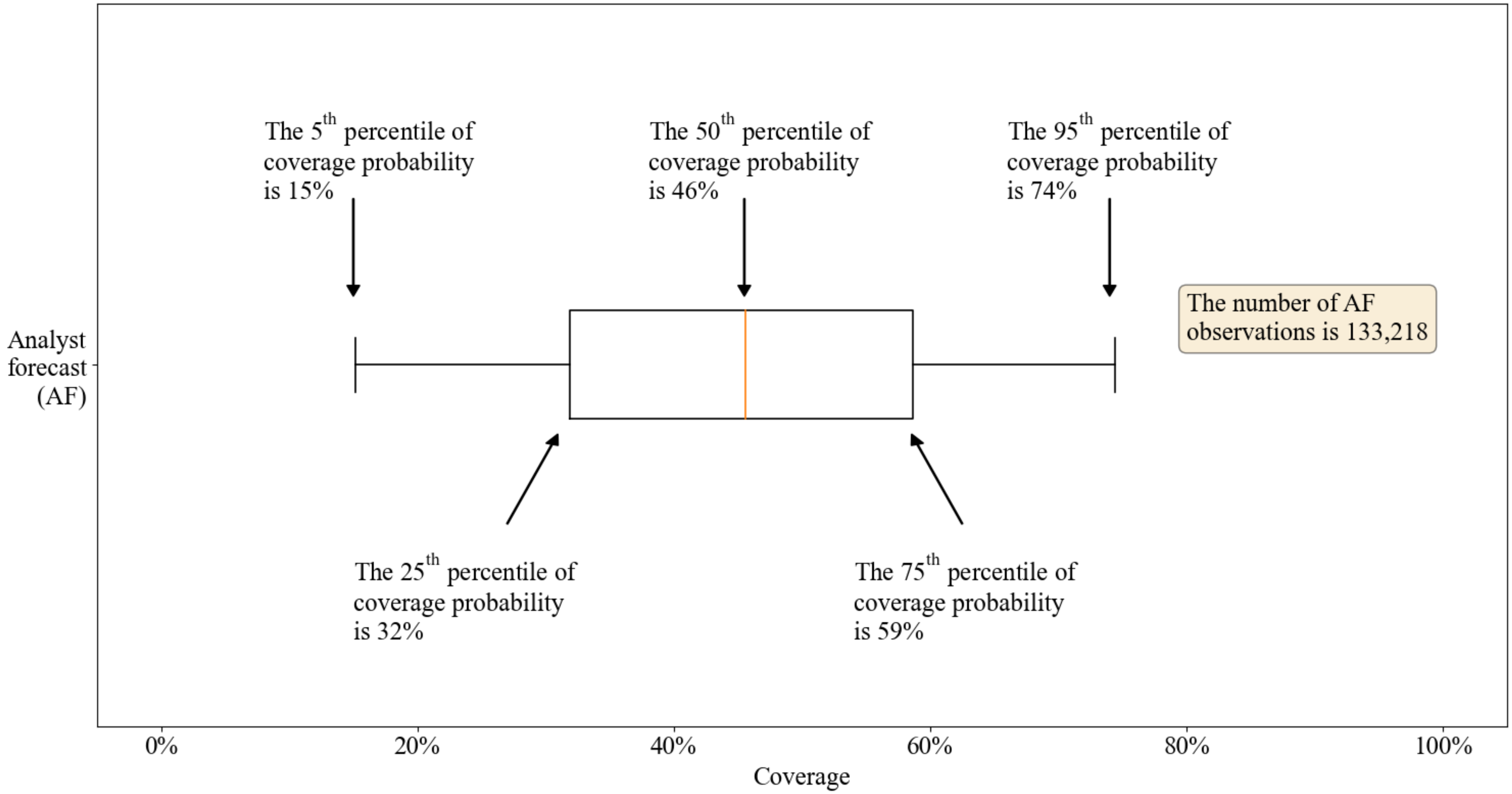


Fig. 4. Box plots summarizing the probabilities of covering actual earnings for management earnings forecast ranges (Panel A) and analyst forecast ranges (Panel B).

Table 1

Descriptive statistics.

Panel A: Descriptive statistics for the sample

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
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
Earn	283,356	0.001	0.173	0.000	0.011	0.018
$Earn_{q-4}$	247,580	0.005	0.033	0.002	0.012	0.019
FF5 CAR	151,661	−0.001	0.098	−0.041	−0.001	0.039
Gross Profit	282,261	0.090	0.131	0.031	0.059	0.103
Mkt-adj CAR	151,661	−0.001	0.099	−0.042	−0.001	0.039
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
WC Accruals	259,330	0.016	0.067	−0.003	0.007	0.024
Payoff	145,903	0.000	0.007	−0.003	0.001	0.005

Panel B: Descriptive statistics for the skewness and kurtosis of earnings distributions

	N	mean	p5	p25	p50	p75	p95
Skewness	151677	−1.458	−4.691	−2.206	−1.045	−0.339	0.424
Kurtosis	151677	13.399	2.222	4.271	7.800	15.525	42.615

Panel A of this table presents summary statistics for the sample of 283,356 firm-quarter observations between January 2001 and December 2021. Panel B presents the descriptive statistics for the skewness and kurtosis of earnings distributions derived from BoXHED. Appendix A presents the definitions of all variables.

Table 2

Continuous ranked probability score (CRPS) for earnings forecasts

Panel A: The CRPS values for different types of earnings forecasts - accounting variables only

year	(1) OLS mean	(2) BoXHED mean	(3) Distributional forecasts using 11 quantiles	(4) Distributional forecasts using 150 quantiles	(5) BoXHED distributions
2011	0.0193	0.0143	0.0137	0.0124	0.0106
2012	0.0200	0.0143	0.0139	0.0125	0.0106
2013	0.0163	0.0113	0.0117	0.0103	0.0085
2014	0.0146	0.0097	0.0105	0.0092	0.0074
2015	0.0177	0.0131	0.0134	0.0123	0.0102
2016	0.0200	0.0149	0.0152	0.0141	0.0114
2017	0.0181	0.0131	0.0137	0.0125	0.0100
2018	0.0198	0.0151	0.0152	0.0141	0.0117
2019	0.0238	0.0188	0.0186	0.0179	0.0149
2020	0.0321	0.0281	0.0252	0.0253	0.0228
2021	0.0228	0.0169	0.0173	0.0171	0.0129
All	0.0204	0.0154	0.0153	0.0143	0.0119

Panel B: The reduction in CRPS for different earnings forecasts – accounting variables only

year	(1) OLS mean (Baseline)	(2) CRPS reduction for BoXHED mean	(3) CRPS reduction for forecasts using 11 quantiles	(4) CRPS reduction for forecasts using 150 quantiles	(5) CRPS reduction for BoXHED distributions
2011	0.0193	26.18%[a]	29.04%[a]	35.78%[a]	45.3%[ab]
2012	0.0200	28.62%[a]	30.4%[a]	37.77%[a]	47.23%[ab]
2013	0.0163	30.67%[a]	27.88%[a]	36.57%[a]	47.86%[ab]
2014	0.0146	33.31%[a]	28.19%[a]	36.64%[a]	49.5%[ab]
2015	0.0177	26.05%[a]	24.63%[a]	30.83%[a]	42.27%[ab]
2016	0.0200	25.68%[a]	23.76%[a]	29.25%[a]	43.15%[ab]
2017	0.0181	27.51%[a]	24.25%[a]	31.11%[a]	44.42%[ab]
2018	0.0198	23.71%[a]	23.41%[a]	28.79%[a]	40.86%[ab]
2019	0.0238	21.19%[a]	22.07%[a]	25.01%[a]	37.61%[ab]
2020	0.0321	12.59%[a]	21.61%[a]	21.16%[a]	29.15%[ab]
2021	0.0228	25.72%[a]	24.14%[a]	24.92%[a]	43.29%[ab]
All	0.0204	24.51%[a]	24.98%[a]	29.68%[a]	41.71%[ab]

Table 2 (Continued)

Panel C: The CRPS values for different types of earnings forecasts – full models

year	(1) OLS mean	(2) BoXHED mean	(3) Distributional forecasts using 11 quantiles	(4) Distributional forecasts using 150 quantiles	(5) BoXHED distributions
2011	0.0131	0.0092	0.0081	0.0071	0.0069
2012	0.0134	0.0090	0.0079	0.0069	0.0067
2013	0.0100	0.0075	0.0068	0.0059	0.0057
2014	0.0088	0.0065	0.0060	0.0051	0.0049
2015	0.0120	0.0092	0.0085	0.0075	0.0073
2016	0.0130	0.0102	0.0093	0.0082	0.0079
2017	0.0114	0.0089	0.0082	0.0071	0.0068
2018	0.0124	0.0105	0.0096	0.0086	0.0084
2019	0.0144	0.0130	0.0115	0.0107	0.0105
2020	0.0213	0.0207	0.0175	0.0168	0.0165
2021	0.0119	0.0114	0.0097	0.0091	0.0088
All	0.0129	0.0106	0.0094	0.0085	0.0082

Panel D: The reduction in CRPS for different earnings forecasts – full models

year	(1) OLS mean (Baseline)	(2) CRPS reduction for BoXHED mean	(3) CRPS reduction for forecasts using 11 quantiles	(4) CRPS reduction for forecasts using 150 quantiles	(5) CRPS reduction for BoXHED distributions
2011	0.0131	30.13%[a]	38.32%[a]	45.80%[a]	47.45%[ab]
2012	0.0134	32.42%[a]	40.57%[a]	48.52%[a]	49.57%[ab]
2013	0.0100	24.77%[a]	32.02%[a]	41.38%[a]	42.96%[ab]
2014	0.0088	25.66%[a]	31.60%[a]	41.71%[a]	43.73%[ab]
2015	0.0120	22.82%[a]	29.21%[a]	37.38%[a]	39.40%[ab]
2016	0.0130	21.23%[a]	28.39%[a]	37.22%[a]	38.90%[ab]
2017	0.0114	21.86%[a]	28.06%[a]	37.74%[a]	40.08%[ab]
2018	0.0124	15.24%[a]	23.17%[a]	30.68%[a]	32.43%[ab]
2019	0.0144	9.62%[a]	19.71%[a]	25.41%[a]	26.90%[ab]
2020	0.0213	2.63%[a]	17.89%[a]	21.22%[a]	22.52%[ab]
2021	0.0119	4.87%[a]	18.50%[a]	23.40%[a]	26.29%[ab]
All	0.0129	17.77%[a]	27.05%[a]	34.20%[a]	35.96%[ab]

This table presents the continuous ranked probability score (CRPS) for five types of earnings forecasts. Panels A and B present the CRPS results for models estimated using firm-fundamental variables only, while Panels C and D present the CRPS results for models estimated using both firm-fundamental and analyst-related variables. Panel A (C) reports the values of CRPS, while Panel B (D) presents the percentage reduction in CRPS relative to the baseline – OLS mean forecast in Column (1). Specifically, Columns (1) to (5) show the results for OLS-predicted mean, BoXHED-predicted mean, quantile regression forecasts using 11 quantiles, quantile regression forecasts using 150 quantiles, and BoXHED distributional forecasts, respectively. [a] denotes statistical significance at the 1% level. [b] indicates that the reduction in Column (5) is also significantly greater than that in Column (4) at the 1% level.

Table 3

Cumulative abnormal returns (bps) for hedge portfolios.

Panel A: Payoff

Payoff deciles	Cumulative abnormal returns (in bps)	
	Mkt-adj CAR	*FF5 CAR*
1 (Low payoff)	−69***	−68***
	(−5.92)	(−5.84)
10 (High payoff)	22***	16**
	(2.9)	(2.08)
Hedge return: High − low	91***	83***
	(6.47)	(5.97)

Panel B: BoXHED-predicted expected surprise

BoXHED-predicted surprise	Cumulative abnormal returns (in bps)	
	Mkt-adj CAR	*FF5 CAR*
1 (Low expected surprise)	−50***	−46***
	(−4.91)	(−4.51)
10 (High expected surprise)	−2	4
	(−0.21)	(0.43)
Hedge return: High − low	48***	50***
	(3.33)	(3.49)

Panel C: OLS-predicted expected surprise

OLS-predicted surprise	Cumulative abnormal returns (in bps)	
	Mkt-adj CAR	*FF5 CAR*
1 (Low expected surprise)	−27**	−18
	(−2.20)	(−1.47)
10 (High expected surprise)	20**	11
	(2.45)	(1.37)
Hedge return: High − low	47***	29*
	(3.15)	(1.96)

This table reports the cumulative abnormal returns for hedge portfolios. Panels A, B, and C report portfolio returns corresponding to expected payoff portfolios, expected surprise portfolios predicted by BoXHED, and expected surprise portfolios predicted by OLS. All the returns are reported in basis points. Appendix A presents the definitions of all variables. *t*-statistics are presented in parentheses. ***, **, * denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table 4

BoXHED distributions vs. mean forecasts in predicting cumulative abnormal returns.

	FF5 CAR				
VARIABLES	(1)	(2)	(3)	(4)	(5)
Payoff_decile	54.119***			47.166***	48.777***
	(3.89)			(3.42)	(3.80)
BoXHED_mean_decile		41.166***		11.204	
		(3.18)		(0.93)	
OLS_mean_decile			34.669**		12.308
			(2.53)		(0.98)
Observations	142,833	142,833	142,833	142,833	142,833
R-squared	0.002	0.002	0.002	0.002	0.002
Industry FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

This table reports the results from OLS regressions based on Equation (3.3), which investigates the association between cumulative abnormal returns around earnings announcements and the expected payoff measure and the expected surprise.

$$CAR_q = \beta_1 Payoff\_decile_q + \beta_2 Mean\_decile_q + FixedEffects + \varepsilon. \qquad (3.3)$$

Each observation in the analysis corresponds to one quarterly earnings announcement. The dependent variable is *FF5 CAR*. The independent variables of interest are the decile ranks of expected payoffs (*Payoff_decile*), the decile ranks of BoXHED-predicted surprises (*BoXHED_mean_decile*), and the decile ranks of OLS-predicted surprises (*OLS_mean_decile*). Columns (1) - (3) examine the effect of *Payoff_decile, BoXHED_mean_decile* and *OLS_mean_decile* separately. Column (4) examines the impact of *Payoff_decile* after controlling for the effect of BoXHED-predicted surprises. Column (5) examines the impact of *Payoff_decile* after controlling for the effect of OLS-predicted surprises. Appendix A presents the definitions of all variables. Industry and year fixed effects are included. Standard errors are clustered by firm and quarter. Robust *t*-statistics are presented in parentheses. ***, **, * denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Internet Appendix

Section IA.1 Distributional forecasting with `BoXHED`

IA.1.1 Review of distributional quantities in statistics

Recall that the probability distribution of an uncertain outome Y is uniquely characterized by the cumulative distribution function (CDF) $F(y) \equiv \Pr(Y \leq y)$. If the distribution has a probability density $f(y)$ then the CDF can be obtained from the density via $F(y) = \int_{-\infty}^{y} f(u)du$. Below we explain an equivalent way to uniquely characterize distributions using the survivor function and the hazard function.

As a basic result in probability theory, the distribution of Y can also be uniquely characterized by the survivor function $S(y)$, which is simply one minus the CDF:

$$S(y) \equiv \Pr(Y > y) = 1 - F(y). \tag{IA.1.1}$$

The analogue to the probability density $f(x)$ for $S(y)$ is the hazard function $\lambda(y)$. This represents the conditional probability density *given that Y is at least as large as y*:

$$\lambda(y) = \frac{f(y)}{S(y)}. \tag{IA.1.2}$$

The survivor function can be obtained from the hazard via $S(y) = e^{-\int_{-\infty}^{y} \lambda(u)du}$. In our context, the hazard $\lambda(y)$ for the earnings distribution is the probability density for the earnings number being y, conditional on it being at least as large as y.

Typically, $S(y)$ and $\lambda(y)$ appear in the context of survival analysis, but as seen above, these quantities are also defined for general distributions. For example, the standard normal distribution Z is not directly related to survival times, yet it has a hazard function that is commonly referred to as the inverse Mills ratio, which is defined even for negative values of z:

$$\lambda(z) = \frac{\phi(z)}{1 - \Phi(z)} \text{ for } -\infty < z < \infty.$$

The takeaway is that if we can estimate the hazard function for Y, then we can estimate its probability distribution and density via the following identities:

$$
\begin{aligned}
F(y) &= 1 - S(y) = 1 - e^{-\int_{-\infty}^{y} \lambda(u)du}, \\
f(y) &= \lambda(y)S(y) = \lambda(y)e^{-\int_{-\infty}^{y} \lambda(u)du}.
\end{aligned}
\tag{IA.1.3}
$$

There are advantages to estimating distributions via the survivor and hazard functions in lieu of the CDF and density function. For example, if the observed values of Y are subject to censoring, then it is very natural to estimate its distribution via $S(y)$ and $\lambda(y)$ rather than $F(y)$ and $f(y)$. This is precisely why the hazard function is the fundamental quantity of interest in survival analysis rather than the density. In our current context, the realized earnings numbers are not subject to censoring, so the distinction between the two different approaches is immaterial for our purpose.

IA.1.2 Earnings distributional forecasting

We established above that if the hazard function of the earnings distribution can be estimated, then we can create distributional forecasts for future earnings. Specifically, let Y be the earnings per share divided by stock price at the beginning of the quarter, i.e.

$$
Y = Earn = EPS_q/P_{q-1}. \tag{IA.1.4}
$$

Given a set of predictor variables X, we use a rigorous machine learning method (described below) to estimate the conditional hazard function $\lambda(y|X)$ for Y. The conditional survivor function $S(y|X)$ is thus obtained from (IA.1.3):

$$
S(y|X) = \exp\left(-\int_{-\infty}^{y} \lambda(u|X)du\right). \tag{IA.1.5}
$$

From (IA.1.5) we can calculate the probability of beating analyst consensus by N cents per share as

$$
\begin{aligned}
\Pr(EPS_q \geq Consensus_q + N|X) &= \Pr\left(\frac{EPS_q}{P_{q-1}} \geq \frac{Consensus_q + N}{P_{q-1}} \middle| X\right) \\
&= \Pr\left(Y \geq \frac{Consensus_q + N}{P_{q-1}} \middle| X\right) \\
&= S\left(\frac{Consensus_q + N}{P_{q-1}} \middle| X\right),
\end{aligned} \tag{IA.1.6}
$$

and the probability of missing consensus by N cents per share can be calculated in a similar manner:

$$
\Pr(EPS_q \leq Consensus_q - N|X) = 1 - S\left(\frac{Consensus_q - N}{P_{q-1}} \middle| X\right). \tag{IA.1.7}
$$

IA.1.3 `BoXHED` as a nonparametric distributional estimator

The nonparametric machine learning method we use to estimate the conditional hazard function $\lambda(y|X)$ is called `BoXHED` (Wang et al. 2020; Pakbin et al. 2025). It is a scalable implementation of a gradient boosting procedure proposed in Lee, Chen, and Ishwaran (2021), and `BoXHED` inherits the mathematical performance guarantees from that paper. We use the current version `BoXHED2.0` which is available as an open source package from `https://github.com/BoXHED/BoXHED2.0`.In the `BoXHED` framework, the "time" variable can be any random variable. Since our application does not involve traditional survival analysis, we treat earnings as the "time" variable, and an explicit event indicator is unnecessary. Each observation in our data is treated as an event whose "time" corresponds to the realized earnings. Conceptually, this means that we model the distribution as if it evolves from time zero up to the observed earnings outcome.[1] The table below gives a simple example of our data. t_{end} is the realized earnings.

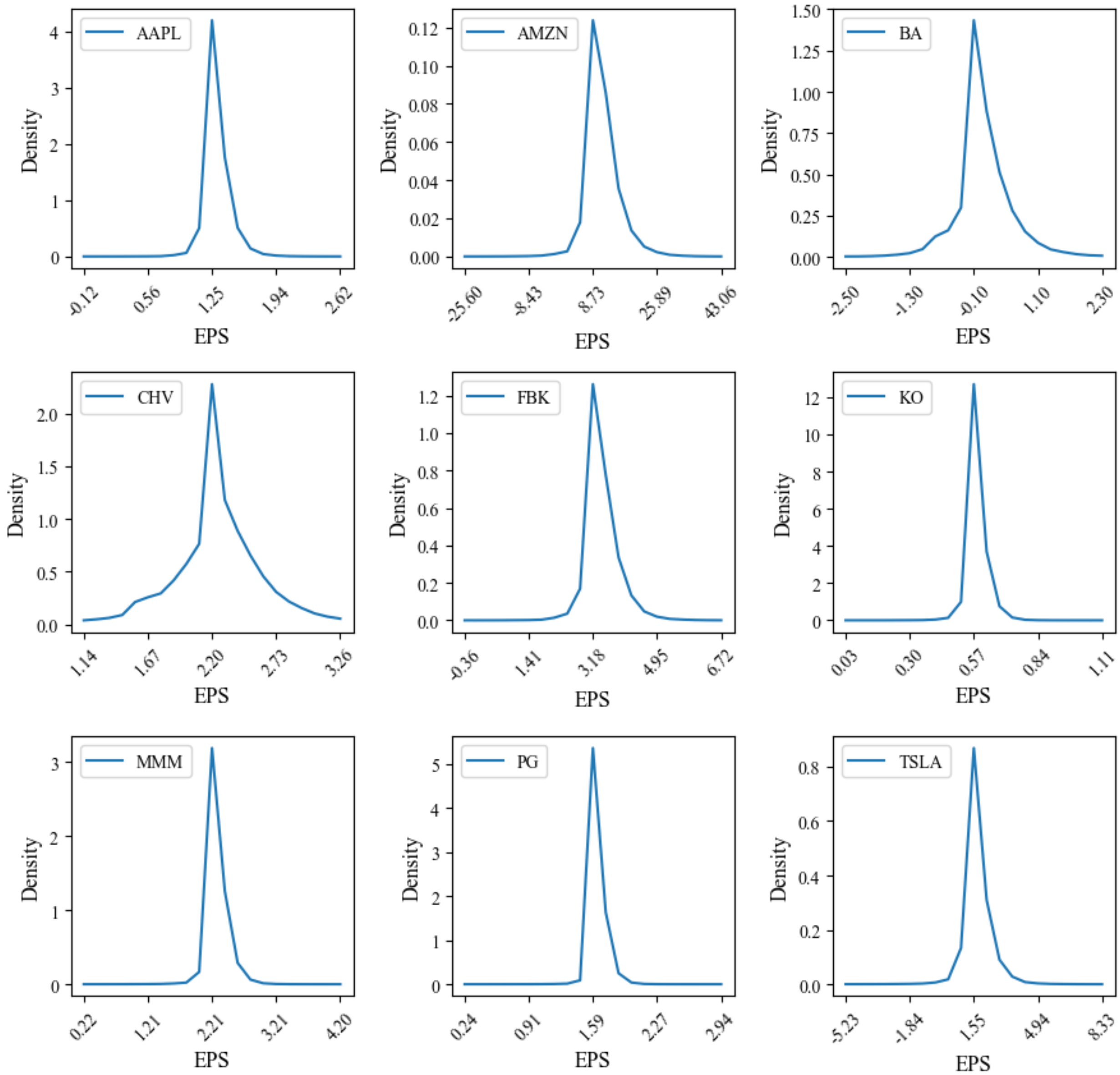
Table A1: Example Data Table

Firm-Q	t_{start}	t_{end}	X_0	$\cdots$	X_{10}	δ
1	0	0.0747	0.2655		0.2059	1
2	0	0.1072	0.7829		0.4380	1
3	0	0.1526	0.7570		0.7789	1
4	0	0.2105	0.9618		0.0859	1

[1] Because earnings can take negative values, we simply add a constant shift so that all observations are nonnegative, which does not affect the relative shape of the distribution.

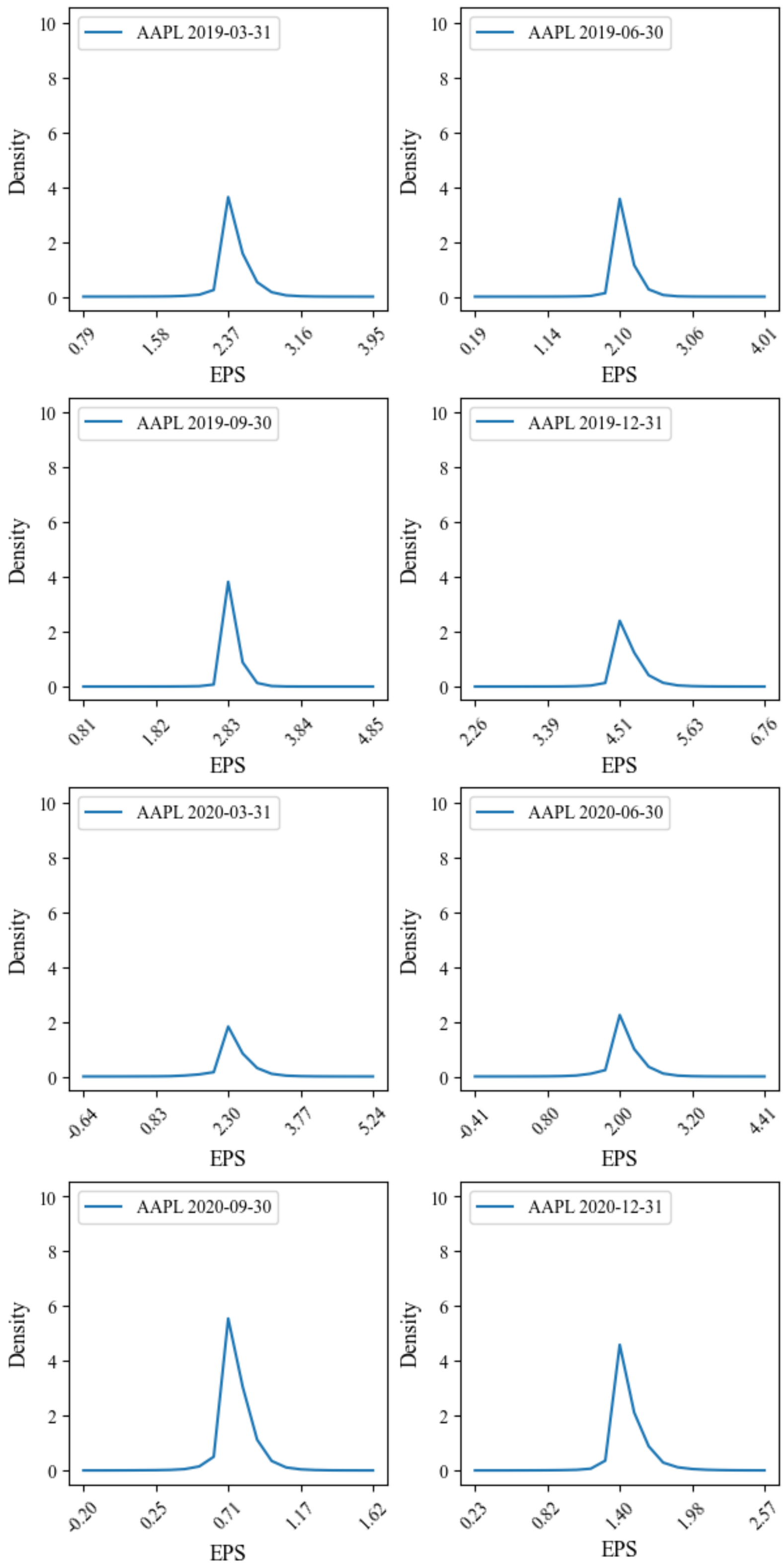
Section IA.2: Examples of BoXHED-derived distributional forecasts of earnings

Panel A: BoXHED-derived distributional forecasts for prominent firms in Q3 2021



Section IA.2: Examples of BoXHED-derived distributional forecasts of earnings (continued)

Panel B: Distributional forecasts for Apple Inc. from 2019 to 2020



Section IA.3: Alternative measure from distributional forecasts: *Differential probability*

In addition to our primary specification in Section 3, we create an alternative measure from our distributional forecasts to illustrate the utility of distributional forecasts for stock trading. Specifically, we create a measure 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:

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

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*

$$\Delta Prob = \hat{P}(EPS_q \leq Consensus_q + N|X) - \hat{P}(EPS_q \leq Consensus_q - N|X) \quad \text{(IA.1)}$$

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 the distributional forecasts rather than the mean of earnings (surprises), which is the focus of prior research. [1] 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.

[2] This distinction is probably consequential because, as discussed above, the distributions of earnings tend to be ill-behaved, with pronounced skewness and heavy tails. Untabulated statistics reveal that the differential probability measure captures information beyond what is conveyed by the mean of the distribution.

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 (CAR), specifically market-adjusted CAR (denoted as *Mkt-adj CAR*), for the portfolios over the three trading days surrounding the earnings announcement date.

Table IA.1 presents CARs for the differential probability portfolios, where in Panel A we focus on the probability of beating and missing consensus forecasts by one cent or more. The findings reveal that the *Mkt-adj CAR* (*FF5 CAR*) difference between the top and bottom deciles is 74 (69) bps and is highly statistically significant, corresponding to an annualized return of 62% (58%) using the convention of 250 trading days. Panels B and C in Table IA.1 display the CAR results based on the probability of beating/missing analyst expectations by two cents and three cents, respectively. The hedge portfolios earn a 3-day *Mkt-adj CAR* (*FF5 CAR*) of 64 bps and 61bps (59 bps and 57 bps) in Panels B and C, respectively. Thus, the results for these more stringent earnings surprise thresholds are very much in line with those for the one-cent specification in Panel A.

Summing up, the results in Table IA.1 suggest that forming trading portfolios that exploit the information conveyed by the distributional forecasts of earnings yields abnormal returns on the magnitude of 60 bps over the three-day earnings announcement window. These returns seem economically substantial, are fairly robust, and remain largely unchanged after control for known return factors.

Table IA.1

Cumulative abnormal returns (bps) for differential probability portfolios.

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

Differential probability deciles	Cumulative abnormal returns (in bps)	
	Mkt-adj CAR	*FF5 CAR*
1 (Low differential probability)	−55***	−56***
	(−5.98)	(−6.2)
10 (High differential probability)	19***	13*
	(2.68)	(1.9)
Hedge return: High − low	74***	69***
	(6.38)	(6.09)

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

Differential probability deciles	Cumulative abnormal returns (in bps)	
	Mkt-adj CAR	*FF5 CAR*
1 (Low differential probability)	−50***	−51***
	(−5.44)	(−5.6)
10 (High differential probability)	13**	8
	(2.01)	(1.16)
Hedge return: High − low	64***	59***
	(5.6)	(5.24)

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

Differential probability deciles	Cumulative abnormal returns (in bps)	
	Mkt-adj CAR	*FF5 CAR*
1 (Low differential probability)	−48***	−49***
	(−5.17)	(−5.37)
10 (High differential probability)	13**	8
	(2.02)	(1.28)
Hedge return: High − low	61***	57***
	(5.41)	(5.15)

This table reports the cumulative abnormal returns 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 (*Differential probabilities*). All the returns are reported in basis points. Variables are defined in Appendix A and above. *t*-statistics are presented in parentheses. ***, **, * denote significance at the 1 percent, 5 percent, and 10 percent levels, respectively.