Visualizing Earnings to Predict Post-Earnings Announcement Drift: A Deep Learning Approach

Jon A. Garfinkel[a], Paul Hribar[b], Lawrence Hsiao[c]

Abstract

We construct a deep learning model to test whether AI can learn from visual depictions of earnings and earnings quality. Quarterly earnings series are transformed into bar-chart images, with bar height representing earnings levels and shading reflecting earnings quality. To avoid look-ahead bias, the model is trained from scratch on corpora excluding references to prior return-prediction research. The model predicts post-earnings announcement returns out-of-sample, with accuracy improving substantially when shading incorporates earnings quality. Predictions based on visualized earnings also outperform trend-detectable models such as Long Short-Term Memory (LSTM) and Temporal Fusion Transformers (TFT). The predictive power of the unshaded model is largely explained by standardized unexpected earnings (SUE) and earnings acceleration, while the shaded model's ability, though partly related to SUE and gross profitability, remains largely unexplained. Overall, we show that visual depictions of earnings and earnings quality forecast returns beyond established prediction models, highlighting the value of combining accounting insights with AI-driven visualization.

[a]Henry B. Tippie Research Professor of Finance, Tippie College of Business, University of Iowa, jon-garfinkel@uiowa.edu

[b]Henry B. Tippie Excellence Chair in Accounting and Professor, Tippie College of Business, University of Iowa, paul-hribar@uiowa.edu

[c]Assistant Professor of Finance, College of Management, National Taiwan University, and Researcher, Center for Research in Econometric Theory and Applications, National Taiwan University, lawrencehsiao@ntu.edu.tw

Hsiao acknowledges the financial support from the National Science and Technology Council (112-2410-H-002-152-) and the Ministry of Education of Taiwan (113L900203). We thank Joshua Yang for excellent research assistance. The authors have no conflicts of interest to declare.

“If I can’t picture it, I can’t understand it.” –Albert Einstein

1 Introduction

Human brains process visual information more quickly than textual or numerical data, often interpreting visual stimuli instantaneously. Tools like graphs, charts, and infographics distill complex information into easily digestible formats, revealing patterns that may remain hidden in tables or raw numbers. The importance of visualization in financial reporting was highlighted by former SEC Chairman Christopher Cox, who remarked that “the visual presentation of information is such a key element of making disclosure understandable to investors” (SEC, 2007). Reflecting this advantage, financial disclosures increasingly incorporate visual elements to improve accessibility and engagement for diverse stakeholders, particularly investors. Recent research underscores this trend. For example, Christensen et al. [2024] document a significant increase in the use of visuals and infographics in 10-K filings, signaling a shift toward visual communication in financial reporting. Similarly, Nekrasov et al. [2021] demonstrate that earnings announcements enhanced with visuals attract greater investor attention, as evidenced by higher engagement metrics such as retweet volumes on platforms like Twitter.

Simultaneously, advancements in AI and machine learning have introduced powerful tools for analyzing and interpreting financial statements. Recent research illustrates the usefulness of AI in this regard: Brown et al. [2020] employ a Bayesian topic-model algorithm to link specific topics in 10-K filings to financial misreporting risks; Bao et al. [2020] develop a machine learning model that predicts fraud using raw financial data; Chen et al [2022] apply machine learning to a detailed set of financial data to predict one-year ahead earnings changes; and Kim, Muhn, and Nikolaev [2024], demonstrate that a large language model (LLM) can analyze income statements and balance sheets without accompanying text and outperform analysts in predicting earnings changes. While these studies provide compelling evidence of the usefulness of AI at interpreting textual or numerical accounting data, the potential for analyzing *visualized* accounting data remains largely unexplored.

Building on these developments, we investigate whether a trained AI model can extract features from visualized earnings data that are predictive of post-earnings announcement drift. Importantly, we assess whether the integration of accounting domain knowledge into the construction of these visualizations enhances the model's predictive performance. Specifically, we transform firms' historical quarterly earnings into bar charts and employ a convolutional neural network (CNN), which is a deep learning algorithm inspired by the human visual system, to extract predictive features.[1]

We begin by plotting earnings. For each firm announcing quarterly earnings from a 20-year in-sample period (1974Q1 to 1993Q4), we plot its most recent eight quarterly earnings in a black-and-white bar chart that visualizes the magnitude as well as the sign of the earnings.[2] Next, each earnings bar chart image is paired with one of the three labels ("sell", "hold", or "buy") based on the relative performance of the firm's 63-day post-announcement buy-and-hold abnormal returns among the cross-section of announcing firms in the same quarter. We train CNN on these in-sample earnings images for it to learn features that best distinguish between the three assigned labels.

We then create out-of-sample earnings images from 1994Q2 onward, and apply the CNN trained model to generate the *CNN buy* probability (CNNBP), which can be

[1]A unique feature of CNN is its use of two-dimensional convolutional filters to scan images, enabling the model to capture fine details and progressively learn the relationship between an image and its corresponding label.

[2]Earnings bars are in white, while the background is in black. We plot raw earnings figures rather than standardized unexpected earnings, as this approach better reflects the types of figures that firms typically showcase during their earnings calls. The plots are standardized and scaled so that the actual level of earnings is not discernible across different firms. See section 2 for details.

thought of as the CNN-predicted likelihood for an image to be a "buy" when "sell" and "hold" options are also available.[3] We assign firms announcing earnings into decile portfolios based on *CNN buy*, where the cutoffs are based on the previous quarter distribution of *CNN buy*. Each decile portfolio has an average post-earning announcement return (cumulated over t+2 through t+64), averaged across all the firms in that decile (that quarter). The key dependent variable for our analyses is then the high-minus-low hedge portfolio (decile 10 minus decile 1) return. We find that the high-minus-low hedge portfolio sorted on the *CNN buy* probability earns positive and highly significant post-announcement returns.

To ascertain the relative contribution of visualization compared to simple machine-interpretation of earnings numbers, we compare the CNN hedge portfolio returns to a pair of well-accepted machine-learning-constructed (using strictly numbers) hedge portfolios. The comparison techniques, used to construct buy probabilities for two trend-detectable AI models, are LSTM and TFT, both of which use raw earnings as inputs. We find that the hedge portfolio returns based on CNNBP are significantly larger than those based on LSTM and TFT. These results remain robust across various benchmarks for expected returns (i.e. various measures for buy-and-hold abnormal returns).[4] This highlights the advantage of using visualized earnings data in AI-based drift prediction.

Thus far our results, though applied to drift as a unique anomaly, are still consistent with the benefits of visualization in predicting stock returns generally (e.g. Jiang et al. (2023) and Murray et al. (2024)). To augment our contribution, particularly to the

[3]We apply the trained CNN to earnings images from 1994Q2 onward to guard against look-ahead bias. See section 2 for details.

[4]We use the market-adjusted buy-and-hold returns, the size-adjusted buy-and-hold returns, and the buy-and-hold returns adjusted by factor models including the Fama-French four- and six-factor models (Fama and French [1993], Carhart [1997], Fama and French [2015], Fama and French [2018]), the q^5-model (Hou et al. [2015], Hou et al. [2021]), and the risk-and-behavioral model (Daniel et al. [2020]). See Table 2 and the Appendix for more details.

accounting literature, we next use accounting domain knowledge to create earnings bar charts that also reflect the *quality* of the reported earnings. Since earnings of high (low) quality convey information more (less) accurately about a firm's underlying economic activity, we conjecture that CNN's drift-predictive performance will be improved when it is being trained on earnings bar charts enriched with earnings quality.

According to prior studies (e.g., Sloan [1996]; Dechow and Dichev [2002]; Penman and Zhang [2002]; Dechow and Schrand [2004]; Dechow et al. [2010]; Dichev et al. [2013]), quality earnings should be persistent, be supported by cash flows, and reliably predict future earnings. Hence, we employ three summary measures of the quality of earnings: CFO (cash flow from continuing operations), IBC_A (adjusted income before extraordinary items), and OE (Dechow-Dichev operating earnings, where all are scaled by lagged assets. Specifically, earnings corroborated by strong operating cash flows (high CFO) as well as earnings free from transient items (high IBC_A and OE) are regarded as higher quality. To incorporate earnings quality into the earnings bar charts, we adjust their shading (greyscale) by assigning lighter shades to bars representing higher-quality earnings and darker shades to those representing lower-quality earnings. CNN taking these shaded earnings images as inputs is hereafter referred to as CNN+.

We separately train each of the three CNN+ using a 20-year in-sample period starting from 1990Q1.[5] Then, we create out-of-sample earnings images for firms announcing earnings starting from 2010Q2 onward and whose 10-K/10-Q filing is released no later than one day after earnings announcement date. The additional filter is introduced to ensure that all information required to plot new earnings bar charts is available prior to making out-of-sample predictions, and thus there is no look-ahead

[5] We require statement of cash flows data to differentiate earnings quality, which provides the accrual and cash components of earnings. This data broadly became available from 1988 onward. Because of the requirement of 8 quarters for our charts, the analysis using the quality shading begins in 1990Q1.

bias. Finally, we apply the trained CNN+ to these new shaded earnings images, to generate the *CNN+ buy* probability (CNNBP+).

To assess whether the integration of additional accounting domain knowledge into earnings bar charts enhances the model's ability to capture drift-predictive features, we compare the post-earnings announcement drift forecasting performance of CNNBP+ with that of CNNBP. To have a fair comparison, we train a CNN model using the same setup as the CNN+ model, so that the only difference between the CNN model and the CNN+ model lies in the construction of earnings bar charts. We find that the hedge portfolios based on the *CNN buy* probability yield quarterly returns ranging from 2.1% to 2.5% on average, depending on different risk-adjustments (i.e. benchmark expected returns). In contrast, the average hedge portfolio returns based on the *CNN+ buy* probability are significantly larger, ranging from 3.9% to 4.4%, 2.9% to 3.6%, and 3.0% to 3.6% in a quarter, when CFO, IBC_A, and OE are used as the earnings quality measure, respectively. These findings provide strong support for the outperformance of the CNN+ model relative to the CNN model.

We then turn to exploration of the drivers of the return predictability of the CNN and *CNN+ buy* probabilities. To begin, we ask if predictability of them persists after controlling for known determinants of post-earnings announcement drift. Specifically, we estimate Fama and Macbeth [1973] regressions of post-announcement Fama-French five-factor and momentum-adjusted buy-and-hold returns on either the CNN or *CNN+ buy* probability, along with 13 variables.[6] Our results show that the *CNN+ buy*

[6] These "controls" include: standardized unexpected earnings (Ball and Brown [1968], Bernard and Thomas [1989], Foster et al. [1984]); earnings acceleration (He and Narayanamoorthy [2020]); trend in gross profitability (Akbas et al. [2017]); market capitalization (Fama and French [1992], [1993]); book-to-market ratio (Fama and French [1992], [1993]); pre-announcement return (Carhart [1997]); earnings persistence (Francis et al. [2004]); earnings volatility (Cao and Narayanamoorthy [2012]); gross profitability (Novy-Marx [2013]); operating profitability (Ball et al. [2016]); operating accruals (Sloan [1996], Hribar and Collins [2002]); total accruals (Richardson et al. [2005]); and asset growth (Cooper et al. [2008]).

probability that uses accounting domain knowledge to indicate earnings quality, provides incremental predictability for post-earnings announcement drift beyond the 13 variables, whereas the baseline *CNN buy* probability does not. In short, accounting domain expertise is particularly valuable within the context of machine learning visualization tools to predict post-earnings announcement drift.

What does this imply about CNN(+)'s visual learning, and how much of it is related to these factors vs independent of them? To explore which of these extant drift-predictive features are utilized by CNN and CNN+, we employ the decomposition methodology of Hou and Loh [2016]. Applying their 3-step process to our setting allows to estimate the fraction of CNNBP's (CNNBP+'s) drift predictability that is attributable to either an extant-explainer of drift, or a residual. In short, each of the 13 typical explainers of drift is assessed (independently) as a potential driver of the CNNBP(+) relationship with drift in our sample. The residual in each assessment captures the fraction that remains unexplained. We find that standardized unexpected earnings (SUE) explains 63.8% of CNNBP's return predictability, followed by earnings acceleration (EA) at 37.1%. SUE and EA also contribute to the drift predictability of three versions of CNNBP+, ranging from 31.4% to 37.3% and from 14.1% to 20.2%, respectively.

More importantly, the Hou and Lo [2016] methodology can be applied in a multivariate setting. When we include all 13 (extant drift-explaining) variables in their framework, we can evaluate the marginal contribution of each candidate variable *and* the total fraction of the CNNBP's (or CNNBP+'s) drift predictability by these variables collectively. We find that SUE and EA continue to be the main contributors to the drift predictability of CNNBP, and only 14.6% of the return predictability of CNNBP is left unexplained by the 13 variables. On the other hand, SUE and gross profitability (GP) are the primary explanatory variables for the CNNBP+'s return predictability, but the fraction of the drift predictability of CNNBP+ left unexplained is large (ranging from 41.8% to 60.6%) and statistically significant at the 1% level. In other words, when we incorporate accounting domain expertise to adjust the earnings bar charts to reflect earnings quality, we uncover substantial predictive information that has not been previously shown to explain post-earnings announcement drift.

Our study makes several contributions to the literature. First, we add to a growing literature studying the applications of machine learning techniques in financial statement analysis (e.g., Brown et al. [2020], Bao et al. [2020] Chen et al. [2022], Kim et al. [2024]) and returns prediction (Rapach et al. [2013], Kelly et al. [2019], Feng et al. [2020], Freyberger et al. [2020], Kozak et al. [2020], Gu et al. [2020], Gu et al. [2021], Leippold et al. [2022], Cao et al. [2024], Chen et al. [2024], Murray et al. [2024]). The key differentiating feature of our approach from the above studies is that our input focuses on how machine learning can use *visual* representations of earnings, as well as accounting knowledge, to predict post-earnings announcement drift.

Second, we contribute to two strands of emerging literature studying visualized accounting data. The first strand of literature investigates why firms present visualized data in their disclosures more often, and how such visuals are interpreted by investors. For example, Christensen et al. [2024] document a significant increase in the disclosure of infographics in 10-K filings over time and investigate the relation between the use of infographics and uncertainty in capital markets. Nekrasov et al. [2022] find that visuals in firms' Twitter earnings announcements are associated with more retweets, representing increased attention to the earnings news. Moss [2022] finds that retail investors use their visual perception of earnings surprise displayed on Robinhood rather than the unexpected earnings scaled by stock price in their investment decisions. The second strand of literature examines whether one can extract useful information from these visualized data. For example, Hu and Ma [2024] quantify persuasion in visual, vocal, and verbal dimensions in start-up pitch videos, and find that passionate and warm pitches significantly increase funding probability. Cao et al. [2024] examine the value of visual information provided in corporate executive presentations and use AI to categorize the types of charts presented as forward looking or summarizing and examine how market participants respond to such information. Gu et al. [2023] find that a daily firm-level investor sentiment measure based on graphics interchange format images (GIFs) in postings about firms on Stocktwits.com is positively correlated with same-day stock returns while predicting stock return reversals in the following two weeks.

Our paper differs from the existing literature in two key aspects. First, instead of analyzing pre-existing visualized accounting data, we propose a universal, ground-up approach to visualize a firm's time-series of quarterly earnings into a bar-chart image, and train AI on these earnings images to examine whether it can extract relevant information for predicting post-earnings announcement drift. Second, while most studies show how AI outperforms humans in extracting information to make predictions (one notable exception is Cao et al. [2024], who show that the integration of analyst and machine learning intelligence results in improved stock return predictions), we highlight humans' relative advantages by leveraging accounting domain knowledge to create "informative" earnings images, aiming to enhance the AI's learning outcomes. Specifically, we embed earnings quality (EQ) information through variation in bar shading based on established EQ measures in the literature. Notably, we train AI on these shaded earnings bar-chart images to automatically learn associations between the visuals and post-announcement return labels, without telling the AI that height corresponds to earnings and that shading reflects EQ. Then, we show that AI trained on earnings images shaded with EQ significantly outperforms AI trained on unshaded earnings images in predicting post-earnings announcement drift. We believe that the integration of domain expertise—specifically, the understanding that earnings quality influences investor reactions and how one can reliably measure it—with a visual technique that allows AI to capture this information, adds a novel contribution to the literature.

Last, we contribute to a burgeoning literature employing CNN to make predictions. For example, Obaid and Pukthuanthong [2022] extract information from a large sample of news media images and translate that information into a daily investor sentiment index. Jiang et al. [2023] extract return-predicting information from stock-level charts depicting daily open, close, high, and low prices, as well as trading volume and average prices over a past period, to forecast future returns. Murray et al. [2024] show that a one-dimensional CNN trained on historical returns can strongly predict the cross-section of future stock returns. Our work differs in two ways. First, we focus on predicting the post-earnings announcement drift as opposed to daily or monthly returns. Second, although previous studies explore whether the return predictability of CNN-based signals is subsumed by certain variables, they do not quantify the contributions of these variables. In contrast, we apply an econometric framework that allows us to assess both the magnitude and statistical significance of the extent to which the drift predictability based on CNN predictions can be attributed to known drift predictors and return anomalies in the PEAD literature. In addition, this methodology allows us to shed light on how incorporating human knowledge enhances the ability of CNN models to explain post-earnings announcement drift.

2. Visualizing Earnings, Training, and Prediction

In this section, we introduce the CNN training procedure, which can be thought of as an image classification task. First, for firms announcing earnings in the in-sample period (1974Q1 to 1993Q4), we transform their time series of earnings into bar charts, and assign one of the three labels (i.e., “sell”, “hold”, or “buy”) to each earnings bar-chart image based on the relative performance of its post-earnings announcement returns among the cross-section of firms in the same quarter. Next, we train CNN to “learn” the relationship between these earnings bar-chart images and their assigned labels. Lastly, we create earnings bar-chart images for firms announcing earnings from 1994Q2 onward and employ the trained CNN to generate the probability for these images to be classified as a “buy”.

2.1 Plotting earnings images and assigning labels

2.1. Plotting earning images

We begin by plotting the most recent eight quarterly earnings in bar charts. Our intent was to create a simple chart that would be roughly analogous to what might be presented in earnings conference call. For example, Figure 1, panel A provides a slide from Meta’s 2024 Q3 earnings call where they display nine quarters of past earnings. Rather than have the CNN attempt to classify the variety of earnings images generated by firms, we provide a set of standardized charts for the CNN to train on and process. Following Jiang et al. [2023], we generate black-and-white rather than colored images for simplicity and uniformity. Each black-and-white image is of size 24 × 24 pixels, which is recognized by the machine as a 24 × 24 matrix of 0 (black pixel) and 255 (white pixel). We use black as the background color and white as the color for earnings, and the constant image size setup is for better comparison of earnings patterns across different firms in different quarters. Figure 1, Panel B provides an example of Meta’s 2024 Q3 earnings in our standardized format.

Each quarter occupies 24 × 3 pixels in the image, and quarterly earnings are plotted as “white bars” in the middle column of each quarter. Based on the signs of the most recent eight quarterly earnings, we categorize the earnings images into three types: Type 1, where all quarterly earnings are non-negative; Type 2, where the maximum quarterly earnings is positive and the minimum quarterly earnings is negative; and Type 3, where all quarterly earnings are non-positive. Figure 2 shows Images 1, 2, and 3 as representative examples of Types 1, 2, and 3 earnings images, respectively. In Appendix A, we describe the details of plotting earnings images. For each firm announcing earnings in a given quarter q during the in-sample period (1974Q1-1993Q4), we create its earnings image using the firm's earnings in the most recent eight quarters (quarters q-7 to q).

2.1.2 Assigning labels to earnings images

Next, we assign each earnings image a label indicative of the firm's post-announcement return performance. Specifically, we sort firms announcing earnings in the same quarter into terciles based on their 63-day post-announcement market-adjusted buy-and-hold returns (MAR).[7] An earnings image is labeled as "sell," "hold," or "buy" if its MAR falls into the bottom, middle, or top tercile, respectively. Since the number of training images for each label is about the same, we mitigate the class imbalance issue in CNN training that arises with a disproportionate ratio of labels.

In particular, $\mathrm{MAR}_{i,q+1}$ is defined as the difference between the buy-and-hold return of firm i and that of the CRSP value-weighted market portfolio over the windows [+2, +64] in trading days relative to firm i's earnings announcement date t in quarter q:

[7] In untabulated tests, we find that the final CNN out-of-sample performance is robust to using alternative definitions of abnormal returns such as size-adjusted or factor-adjusted returns in the label-assigning process.

$$\mathrm{MAR}_{i,q+1} = \prod_{k=t+2}^{t+64} \left(1 + R_{i,k}\right) - \prod_{k=t+2}^{t+64} \left(1 + R_{M,k}\right) \,, \qquad (1)$$

where R_i is the delisting-adjusted return of firm i, R_M is the return of the CRSP value-weighted market return, and t is quarter q's announcement date of firm i.[8] The 63-day holding window corresponds to the total number of trading days in three months. We follow previous studies (Vega [2006], Engelberg et al. [2012], Frank and Sanati [2018]) to compute MAR from day 2 to mitigate the impact of bid-ask bounce and other market microstructure effects, and our results are robust to MAR defined using the trading window of [+1, +63].

2.2 Training the CNN

Next, we train CNN on these earnings images (along with their assigned labels). We use three CNN building blocks, with the first block consisting of 64 convolutional filters of 7 × 7 pixels, the second block consisting of 128 convolutional filters of 3 × 3 pixels, and the third block consisting of 256 convolutional filters of 3 ×3 pixels. During training, we follow the CNN literature to use the cross-entropy loss function as the loss function for minimization, randomly select 70% earnings images for training and the other 30% for validation, and adopt similar regularization procedures in Gu et al. [2020] and Jiang et al. [2023] to prevent overfitting.[9] To conserve space, we report the detailed model architecture and training process in the Internet Appendix.

Note that the CNN is trained on historical earnings images with assigned return labels between 1974Q1 to 1993Q4, which requires return information between 1974Q1

[8] We replace missing delisting-adjusted returns with market returns, which is equivalent to reinvesting any remaining proceeds in the market portfolio until the end of the holding period.

[9] We describe the modeling choices in the Internet Appendix. Interested readers may refer to Gu et al. [2020] for detailed explanations on those modeling choices.

to 1994Q1.[10] In addition, since CNN training can yield different outcomes even when using the same architecture and dataset due to the stochastic nature of optimization algorithms and the application of dropout, we independently train the same CNN ten times (ensemble size = 10) and store their parameters for subsequent use.

2.3 Applying the Trained CNN for Prediction

Having trained the CNN on earnings images from the in-sample period, we apply the stored parameters that encapsulate the model's learned knowledge of the relationship between an earnings image and its corresponding label to out-of-sample earnings images. The trained CNN can generate for each out-of-sample earnings image the probability of being classified as "buy", our label of interest. We refer to this predicted likelihood as the *CNN buy* probability (CNNBP).

Next, for firm announcing earnings between 1994Q2 to 2023Q2, we create their earnings images using their most recent eight quarters' earnings (quarters q-7 to q) and employ the trained CNN to generate CNNBP for these earnings images. Since the trained CNN is based on information through 1994Q1, we generate CNNBP for earnings images from 1994 Q2 onward to ensure that all predictions occur strictly after the training process. As a result, there is no forward-looking bias.[11] In addition, for each out-of-sample earnings image, we average the CNNBP generated by the ten independently trained CNN, which helps achieve better accuracy and robustness.

[10]Forming post-announcement 63-day return labels for earnings images in 1993Q4 requires return information in 1993Q4 and 1994Q1.

[11]To avoid hindsight bias, when forming out-of-sample predictions on the 63-day ([+2, +64]) post-announcement returns for a given firm announcing earnings in quarter q, one should only use information available up to one day after its earnings announcement date in quarter q.

3. Data and Variables

We focus on U.S. common stocks traded on NYSE, AMEX, and NASDAQ, and obtain data from Compustat and CRSP. First, we collect Compustat firm-quarters whose earnings announcement date (Compustat item RDQ) is between January 1974 and June 2023 and delete observations with missing RDQ in the most recent eight quarters (quarters $q-7$ to q). Next, we apply filters in He and Narayanamoorthy [2020] to eliminate announcements that are potentially subject to data errors. In particular, we delete observations if in the most recent eight quarters, a firm has (i) more than one earnings announcement on any date (ii) earnings announcement date within 30 days of a previous earnings announcement date, or (iii) earnings announcement either prior to or more than 180 days after the corresponding fiscal period-end.

We require a firm to have non-missing earnings in the most recent eight quarters and a CRSP daily price higher than one dollar at the most recent earnings announcement date (quarter q). We use income before extraordinary items (Compustat item IBQ) as earnings. Financial and utility firms with SIC codes from 6000 to 6999 and from 4900 to 4949 are excluded. In addition, firms are required to have non-missing market capitalization (SIZE) and non-negative book-to-market ratio (BM) and have at least 90 non-missing daily return observations in the $[-150, -31]$ window relative to the current quarter earnings announcement date. We are left with 403,880 firm-quarter observations after applying all the above filters.

Next, we describe the in-sample dataset and the out-of-sample dataset. The in-sample dataset is used for model training, and the out-of-sample dataset is used for testing the out-of-sample CNN performance. For Table 2, the in-sample dataset consists of 124,341 firm-quarter observations between 1974Q1 to 1993Q4, while the out-of-sample dataset consists of 239,012 firm-quarter observations between 1994Q3 to

2023Q2 with non-missing firm characteristics.[12] For Tables 3 to 6, the in-sample dataset consists of 191,118 firm-quarter observations between 1990Q4 to 2009Q4, while the out-of-sample dataset consists of 43,734 firm-quarter observations between 2010Q3 to 2023Q2 with non-missing firm characteristics and whose current quarter's 10-K/10-Q filing is released no later than one day after earnings announcement date to ensure the availability of other accounting data.

Note that we impose a three-quarter lag between the end of the in-sample period and the start of the out-of-sample period. This is because for all empirical analyses throughout the paper we focus on the decile ranks of CNNBP, and we use the distribution of CNNBP in the previous quarter to determine the cutoff points. In other words, when the final quarter of the in-sample period is quarter q, we can generate CNNBP (CNNBP decile ranks) free from look-ahead bias for earnings images beginning in quarter $q+2$ $(q+3)$.

We summarize the definitions of all the variables used in this study, in Appendix B. To mitigate the impact of outliers, we transform most variables into decile ranks (numbered 0 to 9, from low to high) following prior research (e.g., Rangan and Sloan [1998], Livnat and Mendenhall [2006], Garfinkel and Sokobin [2006]).[13] The cutoff points for quarterly variables are based on the distribution of these variables in the previous quarter, and the cutoff points for annual variables from July in year t to June in year $t + 1$ are based on the distribution of these variables at the end of June in year t. Then, we convert all the decile ranks to scaled ranks by dividing by 9 and subtracting 0.5. The resulting scaled ranks vary from −0.5 to 0.5 with a mean of zero and a range

[12] The firm characteristics include SUE, EA, TREND, PASTRET, PERSIST, VOL, GP, OP, OA, TA, and AG. Along with BM and SIZE, these firm characteristics are used as comparing/control variables throughout the paper. See Appendix B for variable definitions.

[13] Variables that are not transformed into decile ranks are the six measure of the 63-day post-announcement buy-and-hold abnormal return (BHAR), including market-adjusted return (MAR), size-adjusted return (SAR), and four factor-adjusted returns (FF4, FF6, HMXZ5, and DHS3).

of one. This variable transformation approach is to facilitate comparison of the economic magnitudes of firm characteristics. For example, the coefficient on a variable of interest (in scaled rank) in a return regression represents the return from a zero-investment strategy of going long on the highest variable decile and short on the lowest variable decile.

4. CNN Performance

4.1 Portfolio Analysis

If CNN is capable of detecting the features of earnings images that are indicative of post-earnings announcement performance, firms with higher *CNN buy* probability should outperform firms with lower *CNN buy* probability in post-earnings announcement returns. To test this hypothesis, we assign firms announcing earnings into decile portfolios based on their *CNN buy* probability, where the cutoffs are based on the distribution of the previous quarter's *CNN buy* probability. This approach prevents hindsight bias from classifying firms into portfolios based on information not available at the time the strategy is implemented (Foster et al. [1984], Bernard and Thomas [1989]). Hence, while the *CNN buy* probability can be generated without hindsight bias starting from 1994Q2, our empirical analysis focuses on the period from 1994Q3 onward, during which the *CNN buy* probability can be reliably transformed into decile ranks.

Next, we compute the average difference in 63-day post-announcement MAR between firms in the highest *CNN buy* probability decile and firms in the lowest *CNN buy* probability decile and report the results in Table 2. Panel A indicates that the average difference in MAR is 3.5% (t-statistic = 6.990) in a quarter, which corresponds to an annualized return of 14%. To ensure that the results are robust to alternative risk adjustments, we proceed to examine the average difference in size-adjusted (SAR) and factor-adjusted buy-and-hold returns between the highest and lowest *CNN buy* probability deciles.

SAR is defined as the difference between the buy-and-hold return of an announcing firm and that of a size-matched portfolio over the 63-day window ([+2, +64]) following its earnings announcement date. We use the monthly NYSE size decile breakpoints at the end of June in year t to determine the size-matched portfolio for a firm whose earnings announcement date is between July of year t to June of year $t+1$. Monthly size breakpoints and daily size portfolio returns are obtained from Kenneth French's website.

To compute the 63-day factor-adjusted buy-and-hold returns, we replace $R_{M,k}$ in equation (1) with daily return $\widehat{R_{F,k}}$ predicted by factor models. To compute $\widehat{R_{F,k}}$, we first estimate individual stock factor loadings by regressing returns on the factors on a 120-day rolling window from $t-150$ to $t-31$ for each stock:

$$r_{i,t} = \alpha_i + \beta_i' F_t + \epsilon_{i,t}, \tag{2}$$

where $r_{i,t}$ is the excess return on stock i and F_t is a vector of factors. The predicted return $\widehat{R_{F,k}}$ is then computed as $\widehat{\beta'_\iota} F_k$.[14] In particular, we consider the factors in the Fama-French four- and six-factor models (Fama and French [1993], Carhart [1997], Fama and French [2015], Fama and French [2018]), the q^5-model (Hou et al. [2015], Hou et al. [2021]), and the risk-and-behavioral model (Daniel et al. [2020]). The 63-day factor-adjusted buy-and-hold returns following an earnings announcement of these models are denoted FF4, FF6, HMXZ5, and DHS3, respectively.[15]

[14]See, for example, Savor [2012] and Kapadia and Zekhnini [2019].

[15]Fama and French [2015] extends the Fama-French three-factor model (Fama and French [1993]) to control for operating profitability (RMW) and investment (CMA). After the inclusion of a momentum factor (Carhart [1997]), we have Fama-French four-factor and six-factor models (Fama and French

Returning to Table 2, the rest of columns in Panel A present qualitatively similar results: the average return differential between the highest and lowest *CNN buy* probability deciles in SAR or factor-adjusted buy-and-hold returns (FF4, FF6, HMXZ5, and DHS3) range from 3.1% to 3.5%, with *t*-statistics all statistically significant at the 1% level. Overall, Panel A of Table 2 shows a significantly positive relation between *CNN buy* probability and post-announcement buy-and hold abnormal returns that are robust to various risk adjustments.[16]

4.2 Comparison with Alternative Deep Learning Models

While Panel A of Table 2 indicates that CNN exhibits decent return-predicting performance in the out-of-sample period, a natural question arises: what value is added by manually converting raw earnings into two-dimensional images? For example, one may consider using trend-detectable models such as Long Short-Term Memory (LSTM) or Temporal Fusion Transformers (TFT) to analyze the raw data directly, bypassing the need for humans to convert raw data into images.

To answer the question, we train LSTM and TFT with firms' most recent eight quarterly earnings (along with the assigned "sell", "hold", or "buy" label) in the same in-sample period. Earnings are divided by lagged assets to enhance cross-sectional comparability. Then, for each of the three models we analogously generate "buy probability" for earnings images in the same out-of-sample period, form decile

[2018]). Hou et al. [2015] propose the q-model to control for market, size (ME), investment (IVA), and profitability (return on equity, ROE), and Hou et al. [2021] further includes an expected growth factor (EG) into the q^5-model. Daniel et al. [2020] propose a 3-factor risk-and-behavioral model that accounts for market, long-term financing (FIN), and short-term earnings surprise (PEAD). Fama-French factors are obtained from Kenneth French's website, q^5-model factors are obtained from Lu Zhang's website, and DHS3 factors are obtained from Lin Sun's website.

[16] The results are qualitatively the same if we assign earnings images with labels based on firms' 21-day or 42-day post-announcement market-adjusted buy-and-hold returns, train CNN on these images, and examine the 21-day or 42-day post-announcement buy-and-hold abnormal returns for *CNN buy* probability deciles.

portfolios accordingly, and report the average return differential in 63-day post-announcement buy-and-hold abnormal returns between the highest and lowest buy probability deciles. We describe the modeling choices for both models in the Internet Appendix.

Panels B and C report the results for LSTM and TFT, respectively. We find that the average return differences in post-announcement buy-and-hold abnormal returns are significantly positive, suggesting that both models can detect useful features from the time-series of earnings that are related to post-announcement performance. However, Panels D and E show that the average hedge returns based on CNN are significantly larger than those based on each of the three models in the out-of-sample period. The results suggest that CNN combined with image representation of historical earnings produces more information useful in prediction post-earnings announcement drift.

5. Incorporating Accounting Knowledge into CNN

5.1 Variation in Bar Charts Using Earnings Quality

Results from the previous sections suggest that CNN outperforms alternative deep learning models in predicting post-earnings announcement drift. Next, we seek to leverage our domain expertise in accounting to create more informative visual representations of earnings images for the machine learning. We believe that the drift-predicting power of CNN can be further enhanced when input bar charts are enriched with information that provides visual clues about the quality of the earnings bars that we graphed. Specifically, we expect that the hedge portfolio return sorted on CNNBP will be larger when input bar charts are constructed to not only reflect earnings levels, but also to reflect accounting-relevant information.

We incorporate the concept of earnings quality into the construction of earnings bar charts. While earnings provide information about an enterprise's financial performance during a given period (Statement of Financial Accounting No.1), they do not, in isolation, convey anything about persistence or quality of the reported figures. According to the literature (e.g., Sloan [1996]; Dechow and Dichev [2002]; Penman and Zhang [2002]; Dechow and Schrand [2004]; Dechow et al. [2010]; Dichev et al. [2013]), high quality earnings accurately reflect a firm's core business operations and tend to repeat themselves in the future. In other words, high-quality earnings should be backed by operating cash flows and serve as useful predictors of future earnings.

Motivated by these prior studies, we consider the following three earnings quality measures (all scaled by lagged assets). The first earnings quality measure is cash flow from continuing operations (CFO), defined as cash flow from operating activities less cash flow from extraordinary items and discontinued operations (e.g. Sloan 1996). Earnings consist of an accrual and a cash flow component, and earnings are more likely to persist if they are backed by higher cash flows from continuing operations. By contrast, accruals are associated with lower persistence and deemed lower quality (Sloan [1996], Dichev et al. [2013]). As a result, higher CFO represents higher earnings quality.[17]

The second and third earnings quality measures are two variables defined in Ball and Nikolaev [2022]: adjusted income before extraordinary items (IBC_A) and Dechow-Dichev operating earnings (OE), with higher values representing higher earnings quality. IBC_A removes non-operating items (e.g., depreciation and amortization) that have no future operating cash flow equivalent from income before extraordinary items

[17] We tired several additional measures of quality to augment the earnings bars, including change in inventory, change in accounts receivable, special item accruals, an indicator for meet-or-beat, and changes in R&D spending. None of these measures significantly improved the CNN+ model's ability to predict returns.

as reported in cash flow statements (IBC_A), while OE is defined as the sum of operating cash flows and working capital accruals following Dechow and Dichev [2002]. The two variables can be viewed as the accruals-based earnings versions that remove transitory items and closely align with operating cash flows. Ball and Nikolaev [2022] find that both variables dominate operating cash flows in predicting future cash flows.

We need to add variation in the bar charts that reflect earnings quality in addition to earnings numbers. We follow the previous approach to create earnings bar-chart images, but now fill each earnings bar with different shades of grey to signal the firm's earnings quality in the given quarter. Specifically, we transform the earnings quality (i.e., CFO, IBC_A, or OE) of a firm's quarterly earnings into deciles, where the cutoff points are based on the distribution of these variables from the same quarter of the previous year.[18] Quarterly earnings in the lowest earnings quality decile (decile 1) are filled with dark grey (pixel value = 26), while quarterly earnings in the highest earnings quality decile (decile 10) are filled with white (pixel value = 255). The pixel value alternately increases by 25 and 26 as the earnings decile increases. If a firm's earnings quality cannot be computed in a given quarter due to missing data, we assign an earnings quality decile of 5.5. Its pixel value is set to 140, which is the midpoint between the pixel value of decile 5 (= 128) and the pixel value of decile 6 (= 153).

Panel A of Figure 3 presents the color used to fill each earnings bar-chart image for each earnings quality decile, with the pixel values being higher for higher earnings quality deciles. Panel B displays three example bar-chart images, based on the provided earnings and earnings quality deciles from the most recent eight quarters, with the most recent eight quarterly earnings of the three images being the same as those in Figure 2 to foster comparison. We see that the new earnings bar-chart image with different

[18]This better reflects the effect of the integral approach to quarterly reporting. See, for example, Rangan and Sloan [1998] and Collins and Hribar [2000].

shadings allows machines to "see" both the evolution of quarters earnings along with their earnings quality in a straightforward way.

5.2 Model Performance: CNN vs. CNN+

In this section, we compare the drift-predicting performance of CNN based on original bar charts with that of CNN based on new bar charts incorporating earnings quality. To distinguish between the two, we refer to the latter as the "CNN+" model and denote the buy probability generated by CNN+ as the *CNN+ buy* probability (CNNBP+).

We separately train CNN and CNN+, using an in-sample period spanning from 1990Q4 to 2009Q4. We begin the training period in 1990Q4 for the following reasons. First, the construction of the three earnings quality variables requires data from cash flow statements, which are largely unavailable prior to 1988Q1 in Compustat. Second, determining the cutoffs for earnings quality in quarter q requires the distribution of earnings quality in quarter q-4. As a result, 1989Q1 (four quarters after 1988Q1) is the first quarter for which earnings quality deciles can be established, and 1990Q4 (eight quarters after 1989Q1) is the first quarter in which we can construct shaded earnings bar charts. We follow the same model architecture and regularization procedures as those in Section 2.

In each quarter between 2010Q2 and 2023Q2, we create earnings images for firms announcing earnings and whose 10-K/10-Q filing is released no later than one day after earnings announcement date.[19] We apply this filter to eliminate potential look-ahead bias because firms do not always release earnings announcements concurrently with

[19]The in-sample training period is based on historical earnings images and labels between 1990Q4 to 2009Q4, which were constructed using information available between 1990Q1 to 2010Q1. Hence, we start creating earnings images and making predictions from 2010Q2 onward.

their 10-Q/10-K filings. This procedure ensures that the new bar charts can be reliably constructed prior to generating out-of-sample predictions, as our post-announcement return accumulation period (i.e., [+2, +64]) commences two days after earnings announcement date. Finally, we apply the stored parameters of the previously trained CNN and CNN+ to these earnings images to generate *CNN buy* probability and *CNN+ buy* probability, respectively.

5.2.1 Portfolio analysis

We follow similar procedures in Section 4 to conduct portfolio analysis. we assign firms announcing earnings into decile portfolios based on their *CNN buy* probability and *CNN+ buy* probability, where the cutoffs are based on the distribution of the previous quarter's *CNN buy* probability and *CNN+ buy* probability, respectively. Our out-of-sample portfolio analysis thus focuses on the period from 2010Q3 to 2023Q2, during which *CNN buy* probability and *CNN+ buy* probability can be reliably transformed into decile ranks. Then, we examine the average return differences in 63-day post-announcement buy-and-hold abnormal returns between the highest and lowest *CNN buy* probability or *CNN+ buy* probability deciles.

Table 3 reports the results. Panel A shows that the average hedge portfolio returns based on the *CNN buy* probability range from 2.1% to 2.5% in a quarter depending on different risk-adjustments, with *t*-statistics all statistically significant at the 1% level. This result provides evidence on the drift-predicting performance of CNN. However, in Panels B to D we find that the average hedge portfolio returns based on the *CNN+ buy* probability are larger in magnitude, ranging from 3.9% to 4.4%, 2.9% to 3.6%, and 3.0% to 3.6% in a quarter when using CFO, IBC_A, and OE as the earnings quality measure, respectively. In Panels E to G, we confirm that the average differences in hedge portfolio returns between portfolios sorted on *CNN buy* probability and those sorted on *CNN+ buy* probability are statistically significant.

Figure 4 depicts the hedge portfolio return (based on FF6) for each model over time during the out-of-sample period. First, we find that the hedge portfolio return is positive in 37, 45, 43, 44 out of the 52 quarters for CNN and CFO, IBC_A, and OE versions of CNN+, respectively. In other words, after controlling for the market, size, value, profitability, investment, and momentum effect, firms in the highest CNNBP+ decile outperform those in the lowest CNNBP+ decile during the 63-day post-announcement period in 82.7% to 86.5% of the quarters. In contrast, firms in the highest CNNBP decile outperform those in the lowest CNNBP decile in only 71.2% of the quarters.

As a robustness check, we employ an alternative in-sample period from 1990Q4 to 2014Q4 and an out-of-sample period from 2015Q3 to 2023Q2, perform the same portfolio analysis, and report the results in Table IA1 in the Internet Appendix. We find that the average hedge portfolio returns based on the *CNN buy* probability range from 1.9% to 2.7%, while those based on the *CNN+ buy* probability generated by CFO, IBC_A, and OE versions of the CNN+ model are significantly larger, ranging from 3.1% to 4.0%, 3.9% to 4.9%, and 3.6% to 4.6%, respectively.[20]

Overall, the findings provide encouraging evidence for the added value of accounting domain knowledge.

5.2.2 Cross-sectional regressions

We next perform a cross-sectional regression analysis to simultaneously control for the firm characteristics that may affect the positive relation between the *CNN buy*

[20] Figure IA1 in the Internet Appendix depicts the hedge portfolio return (based on FF6) for each model over time during this out-of-sample period (2015Q3 to 2023Q2).

probability (or the *CNN+ buy* probability) and post-earnings announcement drift. We first consider three earnings attributes: standardized unexpected earnings (Ball and Brown [1968], Bernard and Thomas [1989], Foster et al. [1984]), earnings acceleration (He and Narayanamoorthy [2020]), and trend in gross profitability (Akbas et al. [2017]). Standardized unexpected earnings (SUE) is the earnings surprise based on a seasonal random walk model, earnings acceleration (EA) captures the change in earnings growth from one quarter to the next, and trend in gross profitability (TREND) characterizes the recent path in a firm's profitability in addition to the profit level.

In addition to the three earnings attributes, we also compare to a host of known anomalies: market capitalization (Fama and French [1992], [1993]), book-to-market ratio (Fama and French [1992], [1993]), pre-announcement return (Carhart [1997]), earnings persistence (Francis et al. [2004]), earnings volatility (Cao and Narayanamoorthy [2012]), gross profitability (Novy-Marx [2013]), operating profitability (Ball et al. [2016]), total accruals (Richardson et al. [2005]), operating accruals (Sloan [1996], Hribar and Collins [2002]), and asset growth (Cooper et al. [2008]).

Specifically, we estimate Fama and MacBeth [1973] regressions in which the dependent variable is the firm's post-announcement 63-day Fama-French five-factor and momentum-adjusted buy-and-hold return (FF6). Using FF6 ensures that the return-predicting ability of the *CNN buy* probability (or the *CNN buy*+ probability) is not driven by the market, size, value, profitability, investment, or momentum factors.[21] In each quarter, we run the following cross-sectional regressions for the *CNN buy* probability

[21]To conserve space, we only report the results for FF6. The results are similar when using the other post-announcement buy-and-hold abnormal returns.

$$\text{FF6}_{i,q+1} = \alpha_q + \beta_q \textit{CNN buy} \text{probability}_{i,q} + \sum\beta_{c,q}\text{Controls}_{i,q} + \varepsilon_{i,q+1}, \quad (3)$$

and the following cross-sectional regression for each of the three *CNN+ buy* probability

$$\text{FF6}_{i,q+1} = \alpha_q + \beta_q \textit{CNN+ buy} \text{probability}_{i,q} + \sum\beta_{c,q}\text{Controls}_{i,q} + \varepsilon_{i,q+1}, \quad (4)$$

where i refers to the stock, q refers to the quarter, and $\text{FF6}_{i,q+1}$ is the post-announcement Fama-French five-factor and momentum-adjusted buy-and-hold return over the windows [+2, +64] in trading days relative to firm i's earnings announcement date in quarter q. The *CNN buy* probability, *CNN+ buy* probability, and control variables are converted into scaled ranks ranging from −0.5 to 0.5 with a mean of zero. Then, we average the cross-sectional coefficients across all quarters, and multiply them by 100 (so the coefficients are reported in percent).

Table 4 reports the results. First, the coefficient on the *CNN buy* probability in column 1 is 0.341 (*t*-statistic = 0.544), suggesting that the *CNN buy* probability does not yield incremental drift-predicting power beyond the 13 firm characteristics. However, the coefficient on the *CNN+ buy* probability in column 2 is 1.658 and statistically significant at the 1% level (*t*-statistic = 3.509), suggesting that a long-short strategy of going long on the highest *CNN buy* probability decile and short on the lowest decile generates an incremental 63-day FF6 of around 1.658%, controlling for other anomalies. We find similar results in columns 3 and 4: the coefficients on the *CNN+ buy* probability based on the other two CNN+ models are positive and statistically significant at the 5% level.

The findings in Table 4 indicate that CNN is picking up features that relate to the existing anomalies in predicting post-announcement returns, while a significant fraction of the *CNN+ buy* probability's drift-predicting power remains largely orthogonal to that of the existing anomalies.

5.3 Explaining the Drift Predictability of CNNBP/CNNBP+

Having demonstrated the superior return-predicting power of the *CNN+ buy* probability relative to that of the *CNN buy* probability using both portfolio analysis and cross-sectional regressions, in this section we seek to better understand the underlying sources of their return predictability via a decomposition exercise.

We adopt the econometric approach in Hou and Loh [2016] to evaluate a number of candidate variables that may potentially explain the return-predicting ability of the *CNN buy* probability and that of the *CNN+ buy* probability. Specifically, their methodology allows us to quantify the extent to which each candidate explanation accounts for the return predictability of the *CNN buy* probability, either in isolation or after controlling for other competing explanations. To conserve space, we describe the methodology using the *CNN buy* probability in the following discussions; the same process applies to the *CNN+ buy* probability.

First, we estimate Fama and MacBeth [1973] cross-sectional regressions to examine the relation between the *CNN buy* probability and post-earnings announcement returns. For each quarter q between 2010Q3 to 2023Q2, we run the following cross-sectional regression:

$$\mathrm{FF6}_{i,q+1} = \alpha_q + \beta_q \mathrm{CNNBP}_{i,q} + \varepsilon_{i,q+1}, \tag{5}$$

where i refers to the stock, q refers to the quarter, and $\mathrm{FF6}_{i,q+1}$ is the post-announcement Fama-French five-factor and momentum-adjusted buy-and-hold return over the windows [+2, +64] in trading days relative to firm i's earnings announcement date in quarter q. Using FF6 ensures that the return-predicting ability of the *CNN buy* probability is not driven by the market, size, value, profitability, investment, or momentum factors.[22] The *CNN buy* probability is converted into scaled ranks ranging from −0.5 to 0.5 with a mean of zero. The average coefficient (multiplied by 100 and reported in percent) on the *CNN buy* probability across all quarters equals 1.281% with a t-statistic of 2.617, suggesting a positive relation between the *CNN buy* probability and post-announcement returns.

Next, for each quarter q, we regress the *CNN buy* probability on a candidate explanatory variable:

$$\text{CNNBP}_{i,q} = a_q + \delta_q \text{Candidate}_{i,q} + \mu_{i,q}, \tag{6}$$

where both The *CNN buy* probability and the candidate variables are converted into scaled ranks ranging from −0.5 to 0.5 with a mean of zero. After obtaining the regression coefficient estimates, we decompose $\text{CNNBP}_{i,q}$ into two orthogonal components: the candidate component ($= \delta_q \text{Candidate}_{i,q}$) is the component of $\text{CNNBP}_{i,q}$ related to the candidate variable, and the residual component ($= a_q + \mu_{i,q}$) is the component of $\text{CNNBP}_{i,q}$ unrelated to the candidate variable.

Lastly, we use the linearity of covariances to decompose the β_q from equation (5) into two components, β_q^{C} and β_q^{R}, where the former is the component of β_q related to the candidate variable and the latter is the component of β_q unrelated to the candidate variable. Specifically,

[22]The results are similar when using the other post-announcement buy-and-hold abnormal returns.

$$
\begin{aligned}
\beta_q &= \frac{Cov[\text{FF6}_{i,q+1}, \text{CNNBP}_{i,q}]}{Var(\text{CNNBP}_{i,q})} \\
&= \frac{Cov[\text{FF6}_{i,q+1}, (\delta_q \text{Candidate}_{i,q} + \alpha_q + \mu_{i,q})]}{Var(\text{CNNBP}_{i,q})} \\
&= \frac{Cov[\text{FF6}_{i,q+1}, \delta_q \text{Candidate}_{i,q}]}{Var(\text{CNNBP}_{i,q})} + \frac{Cov[\text{FF6}_{i,q+1}, (\alpha_q + \mu_{i,q})]}{Var(\text{CNNBP}_{i,q})} \\
&= \beta_q^{\text{C}} + \beta_q^{R}. \qquad (7)
\end{aligned}
$$

Therefore, $\beta_q^{\text{C}}/\beta_q$ measures the fraction of the CNNBP's drift predictability explained by the candidate variable, and $\beta_q^{\text{R}}/\beta_q$ measures the "residual" fraction of the CNNBP's drift predictability left unexplained by the candidate variable. By construction, these two fractions sum to one. While the means and variances of the two fractions do not have closed forms, Hou and Loh [2016] derive their approximations using the multivariate delta method based on Taylor series expansions. As a result, we can test the statistical significance of both fractions.

For a candidate variable to have positive contribution towards explaining the CNNBP's drift predictability (i.e., $\frac{\beta_q^{\text{C}}}{\beta_q} > 0$), the correlation between the CNNBP and the candidate variable (sign captured by δ_q) and the correlation between FF6 and the candidate variable (i.e., $Cov[\text{FF6}_{i,q+1}, \text{Candidate}_{i,q}]$) should have the same sign. In contrast, if one of the correlations is close to zero or if the two correlations have opposite signs, then the candidate variable may contribute little or negatively to explaining the CNNBP's drift predictability.

5.3.1 Evaluating candidate explanations one at a time

Panel A of Table 5 presents the results for this decomposition exercise using each of the 13 variables as the candidate variable to explain the CNNBP's drift predictability. We start off by using standardized unexpected earnings (SUE) as an example to illustrate the decomposition analysis (column 1 in Panel A, Table 4). Step 1 is the baseline regression of regressing FF6 on CNNBP, and the average coefficient on CNNBP is 1.281% (t-statistic = 2.617). Note that the average coefficients on CNNBP in Step 1 are the same across all candidate variables, since we require the 13 firm variables to be non-missing in the out-of-sample dataset.

In Step 2, we regress CNNBP on SUE each quarter. The average coefficient on SUE is 44.200 with a t-statistic of 31.333, suggesting that CNNBP is significantly related to SUE (moving from the lowest decile to the highest decile of CNNBP on average leads to a 44.2% increase in the decile of *CNN buy* probability). The adjusted R-squared further shows that 19.9% of the variation in CNNBP can be explained by SUE. In addition, Step 2 allows us to decompose CNNBP each quarter into two orthogonal components: the candidate component ($\delta_q \mathrm{SUE}_{i,q}$) is the component of CNNBP related to SUE and the residual component ($a_q + \mu_{i,q}$) is the component of CNNBP unrelated to SUE.

In Step 3, we use the above two components to decompose the coefficient on CNNBP (β_q) in Step 1 into a component that is related to SUE (β_q^{C}) and a residual component (β_q^{R}), as shown in equation (7). The time-series averages of β_q^{C} and β_q^{R} are 0.818% and 0.463%, respectively, and they sum to $\beta_q\,(= 1.281\%)$ by construction. Since $\frac{\beta_q^{\mathrm{C}}}{\beta_q} = 63.8\%$ and is statistically significant at the 1% level (t-statistic = 2.815), we conclude that SUE alone explains 63.8% CNNBP's drift predictability (i.e., the relation between CNNBP_q and $\mathrm{FF6}_{q+1}$). On the other hand, the fraction left unexplained is 36.2% ($= \frac{\beta_q^{\mathrm{R}}}{\beta_q}$), which is statistically indistinguishable from zero.

Turning to the other candidate variables, we find that in Step 2 CNNBP is positively related to earnings acceleration (EA), trend in gross profitability (TREND), past returns (PASTRET), earnings persistence (PERSIST), earnings volatility (VOL), book-to-market ratio (BM), while negatively related to market capitalization (SIZE), gross profitability (GP), operating profitability (OP), operating accruals (OA), total accruals (TA), and asset growth (AG). Step 3 suggests that SUE and TREND explains 37.1% and 9.5% of the CNNBP's drift predictability, respectively, while the other candidate variables' contributions are not statistically significant at the 10% level.

Panels B to D presents the analogous decomposition exercise for the *CNN+ buy* probability generated by the CFO, IBC_A, and OE versions of the CNN+ model, respectively. First, Step 1 in Panels B to D shows that the average coefficients on CNNBP+ across all quarters are 2.088% (t-statistic = 3.670), 1.716% (t-statistic = 3.311), and 1.760% (t-statistic = 3.905), respectively, suggesting a positive relation between each of the three *CNN+ buy* probability and the post-announcement returns.

Next, while the three CNN+ models are based on different earnings quality measures, Step 2 in Panels B to D suggests that the three *CNN+ buy* probabilities exhibit certain similarities—they are all positively related to SUE, EA, TREND, PASTRET, PERSIST, VOL, and GP, and negatively related to OA, TA, and AG. In contrast, a striking difference emerges when compared to Step 2 in Panel A: gross profitability (GP) is negatively associated with CNNBP, but positively associated with CNNBP+.

Finally, Step 3 shows that the drift predictability of CNNBP+ can be significantly explained by SUE (31.4%), EA (14.1%), TREND (5.9%), and GP (15.7%); by SUE (36.5%), EA (20.2%), and SIZE (11.3%); and by SUE (37.3%), EA (19.3%), TREND (7.6%), SIZE (12.7%), BM (9.3%), and GP (7.2%) in Panels A, B, and C, respectively.

5.3.2 Evaluating all candidate explanations simultaneously

Next, we proceed to examine the total fraction of the CNNBP's (or CNNBP+'s) drift predictability that the 13 candidate variables can collectively explain and assess the marginal contribution of each candidate variable after controlling for the other candidate variables. Table 6 reports the results of this multivariate analysis.

Step 1 is the same as in Table 5. In Step 2, we find the 13 variables explain 32.5%, 21.5%, 26.4%, and 27.2% of the variation in CNNBP, CFO, IBC_A, and OE versions of CNNBP+, respectively. The relation between CNNBP (or CNNBP+) and each of the 13 variables is similar as before. For example, we continue to find that both CNNBP and CNNBP+ are positively related to SUE, while negatively related to OA. Notably, CNNBP is insignificantly related to GP, while the three versions of CNNBP+ are positively related to GP.

Step 3 shows that the 13 variables collectively explain 85.4% (= 100% -14.6%) of CNNBP's drift predictability. Consistent with the findings in Panel A of T able 5, SUE and EA stand out, explaining 59.1% (t-statistic = 2.817) and 14.3% (t-statistic = 1.952) of the return predictability, respectively. The fraction left unexplained is 14.6% and is statistically indistinguishable from zero (t-statistic = 0.407). The results explain why the coefficient of regressing post-announcement returns (FF6) on the *CNN buy* probability is insignificant in column 1 of Table 3: the *CNN buy* probability's drift-predicting ability largely resembles that of SUE and EA, and thus disappears after controlling for SUE and EA in the regression.

On the other hand, while SUE continues to play an important role in explaining each of the three *CNN+ buy* probability's drift-predicting ability (ranging from 28.6% to 35.1%), gross profitability (GP) also has a decent contribution (ranging from 7.7% to 12.5%). Overall, we see the fraction of the CNNBP+'s drift predictability left unexplained is 60.6% (t-statistic = 4.820), 44.9% (t-statistic = 2.983), and 41.8% (t-statistic = 2.399) when CNNBP+ is generated by CFO, IBC_A, and OE versions of the CNN+ models, respectively. The results explain why the coefficient of regressing post-announcement buy-and-hold abnormal returns on the *CNN+ buy* probability all shrinks[23] after controlling for the 13 variables, but remains significantly positive: SUE and GP can partially, but not fully, explain the *CNN+ buy* probability's return-predicting power.

Overall, the decomposition analyses in Tables 5 and 6 offer insights into the drift-predictive power of both the CNN and *CNN+ buy* probabilities. Specifically, the findings indicate that CNN is able to extract relevant information from earnings images that is predictive of future returns. However, its limitation lies in its inability to identify features beyond the earnings data itself, as most of the drift predictability associated with the *CNN buy* probability is already captured by standardized unexpected earnings (SUE) and earnings acceleration (EA), both of which are transformations of the underlying earnings figures.

In contrast, the drift-predictive features captured by the three CNN+ models extend beyond SUE to also encompass gross profitability (GP), an anomaly that is not based on earnings. This finding is particularly noteworthy given that GP is not directly employed as the earnings quality measure in constructing the variation earnings bar charts; instead, we rely on cash flow from continuing operations (CFO) and two operating earnings variables—IBC_A and OE. Furthermore, the 13 firm characteristics collectively explain around 40% to 60% of the drift-predictive power of the *CNN+ buy* probability, leaving a substantial fraction unexplained. In other words, after we utilize

[23]When going from the univariate regression (the first two columns in Table 4) to the multivariate regression in Table 3, the coefficient on the *CNN+ buy* probability shrinks from 2.088% to 1.658%, from 1.716% to 1.110%, and from 1.760% to 1.059% when employing CFO, IBC_A, and OE as the earnings quality measure to create shaded earnings images for the CNN+ models, respectively.

the concept of earnings quality to create variation earnings bar-chart images, the three CNN+ models capture drift-predictive features that are incremental to the existing anomalies in predicting post-announcement returns. The capacity of the CNN+ models to uncover novel sources of return predictability may potentially explain their superior performance relative to the baseline CNN.

6 Conclusion

Our research explores the potential of applying AI to visualized earnings data in financial analysis. First, we transform time series quarterly earnings into earnings bar-chart images, and train CNNs on these images to extract features predictive of post-earnings announcement drift. We find that its out-of-sample drift predictability outperforms that of alternative trend-detectable models.

Next, we employ accounting domain expertise to enrich the earnings bar charts with earnings quality information. Specifically, we borrow three established earnings quality measures from the literature, adjust the shading of earnings bars to reflect earnings quality, and train CNNs on these images accordingly. We find the out-of-sample drift-predicting performance based on the shaded earnings images is superior to that based on the unshaded earnings images, and remains significant after controlling for previously documented anomalies and earnings attributes.

Using a decomposition framework, we show that the drift predictability based on CNN alone can be largely attributed to earnings-based drift predictors such as standardized unexpected earnings and earnings acceleration. In contrast, incorporating human expertise into the CNN model extends its drift-predictive power beyond standardized unexpected earnings to also capture gross profitability. Importantly, a significant portion of the drift predictability based on the combined wisdom of humans and CNN remains unexplained by existing return anomalies.

Overall, our paper highlights the usefulness of applying AI along with human domain expertise to visualized data in accounting research. Although AI increasingly challenges the role of human expertise in accounting, our findings provide a modest example in which AI achieves superior performance when supported by human insights—reinforcing the unique value accountants bring to the table.

References

Akbas, F., C. Jiang, and P. D. Koch (2017). The trend in firm profitability and the cross-section of stock returns. *The Accounting Review 92*(5), 1–32.

Ball, R. and P. Brown (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research 6*(2), 159–178.

Ball, R., J. Gerakos, J. T. Linnainmaa, and V. Nikolaev (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics 121*(1), 28–45.

Ball, R. and V. V. Nikolaev (2022). On earnings and cash flows as predictors of future cash flows. *Journal of Accounting and Economics*, *73*(1), 101430.

Bao, Y., B. Ke, B. Li, and J. Yu. (2020). Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach. *Journal of Accounting Research 58*(1), 199–235.

Bernard, V. L. and J. K. Thomas (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research 27*, 1–36.

Brown, N. R. Crowley, B. Elliott (2020). What are you saying? Using topic to detect financial misreporting. *Journal of Accounting Research 58*(1), 237–291.

Cao, S., Y. Cheng, M. Wang, Y. Xia, and B. Yang (2024). Visual information and AI divide: Evidence from corporate executive presentations. *Available at SSRN.*

Cao, S., W. Jiang, J. Wang, and B. Yang (2024). From man vs. machine to man+ machine: *The art and AI of stock analyses. Journal of Financial Economics 160*, 103910.

Cao, S. S. and G. S. Narayanamoorthy (2012). Earnings volatility, post-earnings announcement drift, and trading frictions. *Journal of Accounting Research 50*(1), 41–74.

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

Chen, L., M. Pelger, and J. Zhu (2024). Deep learning in asset pricing. *Management Science 70*(2), 714–750.

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

Christensen, T. E., K. E. Fronk, J. A. Lee, and K. K. Nelson (2024). Data visualization in 10-k filings. *Journal of Accounting and Economics 77* (2-3), 101631.

Collins, D. W. and P. Hribar (2000). Earnings-based and accrual-based market anomalies: one effect or two? *Journal of Accounting and Economics 29*(1), 101–123.

Cooper, M. J., H. Gulen, and M. J. Schill (2008). Asset growth and the cross-section of stock returns. *The Journal of Finance 63*(4), 1609–1651.

Daniel, K., D. Hirshleifer, and L. Sun (2020). Short-and long-horizon behavioral factors. *The Review of Financial Studies 33*(4), 1673–1736.

Dechow, P. M. and I.D. Dichev (2002). The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review*, *77*(s-1), 35–59.

Dechow, P. M. and C. Schrand (2004). Earnings quality. *The Research Foundation of CFA Institute.*

Dechow, P., W. Ge, and C. Schrand (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, *50*(2-3), 344–401.

Dichev, I. D., J. R. Graham, C. R. Harvey, and S. Rajgopal (2013). Earnings quality: Evidence from the field. *Journal of Accounting and Economics*, *56*(2-3), 1–33.

Engelberg, J., C. Sasseville, and J. Williams (2012). Market madness? The case of mad money. *Management Science 58*(2), 351–364.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *The Journal of Finance 47*(2), 427–465.

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

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fama, E. F. and K. R. French (2018). Choosing factors. *Journal of Financial Economics 128*(2), 234–252.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

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

Foster, G., C. Olsen, and T. Shevlin (1984). Earnings releases, anomalies, and the behavior of security returns. *The Accounting Review,* 574–603.

Francis, J., R. LaFond, P. M. Olsson, and K. Schipper (2004). Costs of equity and earnings attributes. *The Accounting Review 79*(4), 967–1010.

Frank, M. Z. and A. Sanati (2018). How does the stock market absorb shocks? *Journal of Financial Economics 129*(1), 136–153.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies 33*(5), 2326–2377.

Garfinkel, J. A. and J. Sokobin (2006). Volume, opinion divergence, and returns: A study of post–earnings announcement drift. *Journal of Accounting Research 44*(1), 85–112.

Gu, M., S. H. Teoh, and S. Wu (2023). GIF sentiment and stock returns. *Available at SSRN.*

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Gu, S., B. Kelly, and D. Xiu (2021). Autoencoder asset pricing models. *Journal of Econometrics 222*(1), 429–450.

He, S. and G. G. Narayanamoorthy (2020). Earnings acceleration and stock returns. *Journal of Accounting and Economics 69*(1), 101238.

Hou, K. and R. K. Loh (2016). Have we solved the idiosyncratic volatility puzzle? *Journal of Financial Economics 121*(1), 167–194.

Hou, K., H. Mo, C. Xue, and L. Zhang (2021). An augmented q-factor model with expected growth. *Review of Finance 25*(1), 1–41.

Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies 28*(3), 650–705.

Hribar, P. and D. W. Collins (2002). Errors in estimating accruals: Implications for empirical research. *Journal of Accounting Research 40*(1), 105–134.

Hu, A. and S. Ma (2024). Persuading investors: A video-based study. *Forthcoming at The Journal of Finance.*

Jiang, J., B. Kelly, and D. Xiu (2023). (Re-) imag (in) ing price trends. *The Journal of Finance 78*(6), 3193–3249.

Kapadia, N. and M. Zekhnini (2019). Do idiosyncratic jumps matter? *Journal of Financial Economics 13*(3), 666–692.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

Kim, A. M. Muhn, V. Nikolaev (2024). Financial Statement Analysis with Large Language Models. *Available at arXiv.*

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135* (2), 271–292.

Leippold, M., Q. Wang, and W. Zhou (2022). Machine learning in the Chinese stock market. *Journal of Financial Economics 145*(2), 64–82.

Livnat, J. and R. R. Mendenhall (2006). Comparing the post-earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research 44*(1), 177–205.

Moss, A. (2022). How do brokerages' digital engagement practices affect retail investor information processing and trading? *Ph.D. thesis, The University of Iowa*.

Murray, S., Y. Xia, and H. Xiao (2024). Charting by machines. *Journal of Financial Economics 153*, 103791.

Nekrasov, A., S. H. Teoh, and S. Wu (2022). Visuals and attention to earnings news on twitter. *Review of Accounting Studies 27*(4), 1233–1275.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation. *Econometrica 55*(3), 703–708.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics 108*(1), 1–28.

Obaid, K. and K. Pukthuanthong (2022). A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics 144*(1), 273–297.

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

Rangan, S. and R. G. Sloan (1998). Implications of the integral approach to quarterly reporting for the post-earnings-announcement drift. *The Accounting Review*, 353–

371.

Rapach, D. E., J. K. Strauss, and G. Zhou (2013). International stock return predictability: What is the role of the United States? *The Journal of Finance 68*(4), 1633–1662.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and I. Tuna (2005). Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics 39*(3), 437–485.

Savor, P. G. (2012). Stock returns after major price shocks: The impact of information. *Journal of Financial Economics 106*(3), 635–659.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review*, 289–315.

Vega, C. (2006). Stock price reaction to public and private information. *Journal of Financial Economics 82*(1), 103–133.

Appendix A. Plotting Earnings Images.

First, let $E_1, E_2, \dots, E_8$ denote the most recent eight quarterly earnings corresponding to quarter $q-7, q-6, \dots, q, E_{\text{MAX}}$ and E_{MIN} denote the maximum and minimum of the eight quarterly earnings, and $r(\ )$ denote the function that rounds the input value to the nearest whole number. We set the bottom-left vertex of the image as the origin of a two-dimensional coordinate system, so a rectangular area in the image can be represented as $([x_1, x_2], [y_1, y_2])$. Next, we classify firms' most recent eight quarterly earnings into one of the three types, determine the values corresponding to the top and bottom of the image, and plot each quarterly earnings into bars accordingly. The three types are as follows:

- Type I ($E_{\text{MIN}} \geq 0$; the most recent eight quarterly earnings are all non-negative): In this case, we set E_{MAX} and 0 as the top and bottom of the image, respectively. E_{i} is plotted as the area of

$$\left([3i-2, 3i-1], \left[0, r\left(24 * \frac{E_i}{E_{\text{MAX}}}\right)\right]\right), \tag{1}$$

  for i = 1, ..., 8. Image 1 in Figure 2 displays an example earnings image of this type. The maximum earnings is E_7 and thus it occupies a whole column. All other quarterly earnings are plotted upward, and their heights are determined using E_7 as the reference point.

- Type II ($E_{\text{MAX}} > 0$ and $E_{\text{MIN}} < 0$; the maximum quarterly earnings is positive while the minimum earnings is negative): In this case, E_{MAX} and E_{MIN} coincide with the top and bottom of the image, respectively. The implicit "zero-earnings line" corresponds to $r\left(24 * \frac{-E_{MIN}}{E_{MAX}-E_{MIN}}\right)$, and E_{i} is plotted above or below the zero-earnings line as follows:

$$\begin{cases} [3i-2,3i-1], \left[r\left(24*\frac{-E_{MIN}}{E_{MAX}-E_{MIN}}\right), r\left(24*\frac{-E_{MIN}}{E_{MAX}-E_{MIN}}\right)+r\left(24*\frac{E_i}{E_{MAX}-E_{MIN}}\right)\right] & if\ E_i > 0, \\ [3i-2,3i-1], \left[r\left(24*\frac{-E_{MIN}}{E_{MAX}-E_{MIN}}\right)-r\left(24*\frac{-E_i}{E_{MAX}-E_{MIN}}\right), r\left(24*\frac{-E_{MIN}}{E_{MAX}-E_{MIN}}\right)\right] & if\ E_i \leq 0, \end{cases} \tag{2}$$

for i = 1, ..., 8. Image 2 in Figure 2 displays an example earnings image of this type. Here we see the advantage of using bar charts as opposed to line graphs when plotting earnings. Bars can represent positive, zero, or negative earnings without further specifying numbers on the vertical axis. Positive earnings are plotted upward while negative earnings are plotted downward, and the bar lengths (in pixels) are computed as the rounded value of 24 multiplied by the absolute values of E_{i} scaled by $E_{\mathrm{MAX}} - E_{\mathrm{MIN}}$.

- Type III ($E_{\mathrm{MAX}} \leq 0$; the most recent eight quarterly earnings are all non-positive): In this case, 0 and E_{MIN} coincide with the top and bottom of the image, respectively. E_{i} is plotted as the area of

$$[3i-2,3i-1], \left[24-r\left(24*\frac{E_i}{E_{\mathrm{MIN}}}\right), 24\right], \tag{3}$$

for i = 1, ..., 8. Image 3 in Figure 2 displays an example earnings image of this type. The minimum earnings is E_5 and thus it occupies a whole column. All earnings are plotted downward, and their heights are plotted using E_5 as the reference point.

Note that in all three types, it is possible for E_{i} to be very close to zero after scaling and thus does not occupy a full pixel in the image after rounding, i.e., $y_1 = y_2$.[24] In addition, the distance between two neighboring earning of pixel between is

[24]One extreme case is that all eight quarterly earnings are very close to each other so that when plotting earnings on a bar chart, each earnings bar occupies a whole column. In this case, one cannot tell from the image whether all earnings are positive or negative. However, we checked all earnings images and did not find this extreme case.

greater than the distance between the leftmost (or rightmost) earnings and the border of the image, which is consistent with the default setup of a bar chart for most statistical software.

Appendix B. Variable Definitions. This table summarizes variable definitions.

Variables	Descriptions
MAR	Market-adjusted return (MAR) is defined as the difference between the buy-and-hold return of an announcing firm and that of the CRSP value-weighted market portfolio over the 63-day windows [+2, +64] following its earnings announcement date.
SAR	Size-adjusted return (SAR) is defined as the difference between the buy-and-hold return of an announcing firm and that of a size-matched portfolio over the 63-day window ([+2, +64]) following its earnings announcement date. We use the monthly NYSE size decile breakpoints at the end of June in year t to determine the size-matched portfolio for a firm whose earnings announcement date is between July of year t to June of year $t+1$.
FF4	Fama-French three-factor and momentum-adjusted buy-and-hold return during the 63-day window ([+2, +64]) following earnings announcement date, with factor loadings estimated using the 120-day window ([-150, -31], 90 days minimum) prior to the earnings announcement date. The factors are market, size, value, and momentum.
FF6	Fama-French five-factor and momentum-adjusted buy-and-hold return during the 63-day window ([+2, +64]) following earnings announcement date, with factor loadings estimated using the 120-day window ([-150, -31], 90 days minimum) prior to the earnings announcement date. The factors are market, size, value, operating profitability, investment, and momentum.
HMXZ5	q^5-factor-adjusted buy-and-hold return during the 63-day window ([+2, +64]) following earnings announcement date, with factor loadings estimated using the 120-day window ([-150, -31], 90 days minimum) prior to the earnings announcement date. The factors are market, size, investment, return on equity, and expected growth.
DHS3	Behavioral-factor-adjusted buy-and-hold return during the 63-day window ([+2, +64]) following earnings announcement date, with factor loadings estimated using the 120-day window ([-150, -31], 90 days minimum) prior to the earnings announcement date. The factors are market, financing, and post earnings announcement drift.

CFO	Quarterly cash flow from continuing operations, defined as the quarterly change in year-to-date net cash flow from operating activities (OANCFY) less the quarterly change in year-to-date cash flow from extraordinary items and discontinued operations (XIDOCY). The latter is set to zero if missing. CFO is scaled by total assets (ATQ) in the previous quarter.
IBC_A	Quarterly adjusted income before extraordinary items, defined as the sum of the quarterly change in year-to-date income before extraordinary items reported in the cash flow statement (IBCY); the quarterly change in year-to-date depreciation and amortization (DPCY); the quarterly change in year-to-date cash flow from extraordinary items and discontinued operations (XIDOCY); the quarterly change in year-to-date sale of property, plant and investments gain (SPPIVY); the quarterly change in year-to-date net loss earnings (ESUBCY); and the quarterly change in year-to-date other items involved in the calculation of funds from operations (FOPOY). All missing items are set to zero except for the quarterly change in year-to-date income before extraordinary items reported in the cash flow statement. IBC_A is scaled by total assets (ATQ) in the previous quarter.
OE	Quarterly operating earnings, defined as the quarterly change in year-to-date net cash flow from operating activities (OANCFY) minus the sum of the quarterly change in year-to-date accounts receivable decrease (RECCHY); the quarterly change in year-to-date inventory decrease (INVCHY); the quarterly change year-to-date accounts payable and accrued liabilities increase (APALCHY); the quarterly change in year-to-date income taxed accrued increase (TXACHY); and the quarterly change in year-to-date net change in other assets and liabilities (AOLOCHY). All missing items are set to zero except for the quarterly change in year-to-date net cash flow from operating activities. OE is scaled by total assets (ATQ) in the previous quarter.
SUE	Standardized unexpected earnings, defined as quarter q's EPS minus quarter q-4's EPS, scaled by the standard deviation of EPS in the most recent eight quarters (six quarters minimum). EPS is computed as income before extraordinary items (IBQ), divided by shares outstanding (CSHOQ). Shares are adjusted for stock splits.
EA	Earnings acceleration. For firm i in quarter q, we use $$\frac{\mathrm{EPS}_{i,q}-\mathrm{EPS}_{i,q-4}}{\text{Stock Price}_{i,q-1}}-\frac{\mathrm{EPS}_{i,q-1}-\mathrm{EPS}_{i,q-5}}{\text{Stock Price}_{i,q-2}},$$ where $\mathrm{EPS}_{i,q}$ is earnings per share for firm i in quarter q. Shares are adjusted for stock splits.
TREND	Trend in quarterly gross profitability. For firm i in quarter q, we use $\beta_{i,q}$ estimated from the following time-series regression:

$$GPQ_{i,q} = \alpha_{i,q} + \beta_{i,q} t + \lambda_{1,i,q} D1 + \lambda_{2,i,q} D2 + \lambda_{3,i,q} D3 + \epsilon_{i,q} ,$$

where $t = 1, 2, \dots, 8$ and represents a deterministic time trend covering quarter $q - 7$ through q, and D1 to D3 represent quarterly dummy variables. GPQ is calculated as sales revenue (SALEQ) minus costs of goods sold (COGSQ), divided by total assets (ATQ). If SALEQ is unavailable, we use quarterly revenue (REVTQ). If COGSQ is unavailable, we use quarterly total operating expenses (XOPRQ) minus quarterly selling, general and administrative expenses (XSGAQ, zero if missing).

PASTRET — Past return, defined as the value-weighted market-adjusted stock return during the [−30, −2] window prior to earnings announcement date.

PERSIST — Earnings persistence. For firm i in quarter q, we use $\beta_{i,q}$ estimated from the following time-series regression:

$$\text{EARNINGS}_{i,q} = \alpha_{i,q} + \beta_{i,q} \text{EARNINGS}_{i,q-1} + \epsilon_{i,q} ,$$

with the most recent eight quarters (quarter $q - 7$ to q) of earnings (IBQ).

VOL — Earnings volatility. We use the standard deviation of ROA in the most recent eight quarters (quarter $q - 7$ to q). ROA is defined as quarterly earnings (IBQ) divided by total assets (ATQ) in the previous quarter.

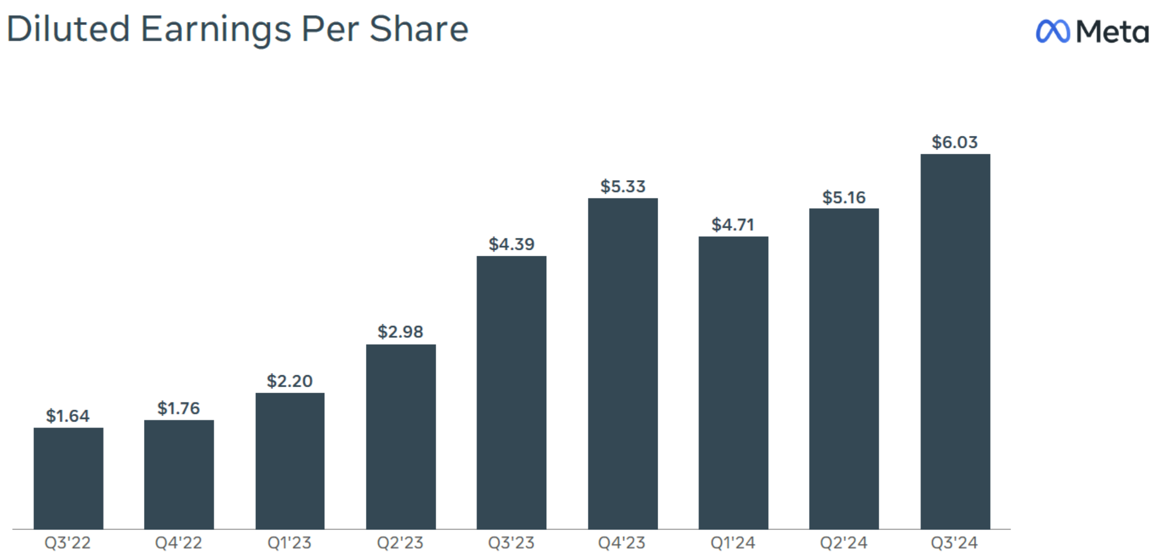
SIZE — Firm size for July of year t to June of year $t + 1$ is defined as June market capitalization (from CRSP) of year t.

BM — Book-to-market ratio for July of year t to June of year $t + 1$ is defined as book equity for the fiscal year ending in calendar year $t - 1$ divided by the market capitalization at the end of December of $t - 1$. Book equity is computed as stockholders' book equity (SEQ), plus deferred taxes (TXDB, zero if missing) and investment tax credit (ITCB, zero if missing), minus the book value of preferred stock (depending on availability, we use redemption (PSTKRF), carrying (PSTKL), or par value (PSTK)).

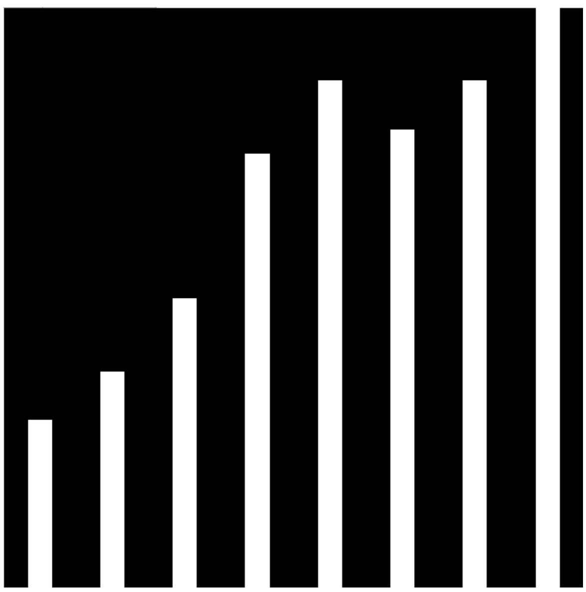
GP — Gross profitability for July of year t to June of year $t + 1$ is defined as sales revenue (SALE) minus cost of goods sold (COGS), divided by total assets (AT) for the fiscal year ending in calendar year $t - 1$. If SALE is unavailable, we use revenue (REVT). If COGS is unavailable, we use total operating expenses (XOPR) minus selling, general and administrative expenses (XSGA, zero if missing).

OP — Operating profitability for July of year t to June of year $t + 1$ is defined as sales revenue (SALE) minus cost of goods sold (COGS), minus selling, general, and administrative expenses (XSGA), and plus research and development expenditures (XRD, zero if missing), scaled by total assets (AT) for the fiscal year ending in calendar year $t - 2$. If SALE is unavailable, we use revenue (REVT). If COGS is unavailable, we use total

	operating expenses (XOPR) minus selling, general and administrative expenses (XSGA, zero if missing).
TA	Total accruals for July of year t to June of year $t+1$ is defined as net income (NI) minus operating, investing, and financing net cash flows (OANCF, IVNCF, and FINCF) plus sales of stocks (SSTK, zero if missing) minus stock repurchases and dividends (items PRSTKC and DV, zero if missing) for the fiscal year ending in calendar year $t-1$, scaled by total assets (AT) for the fiscal year ending in $t-2$.
OA	Operating accruals for July of year t to June of year $t+1$ is defined as net income (NI) minus net cash flow from operations (OANCF) for the fiscal year ending in calendar year $t-1$, scaled by total assets (AT) for the fiscal year ending in $t-2$.
AG	Asset growth for July of year t to June of year $t+1$ is defined as total assets (AT) for the fiscal year ending in calendar year $t-1$ minus total assets for the fiscal year ending in $t-2$, scaled by total assets for the fiscal year ending in $t-2$.



Panel A. Actual earnings chart from Meta's Q3 2024 Earnings Presentation



Panel B. Transformed image of Meta's earnings

Figure 1. Examples of visualized earnings information. This figure displays the actual visualized earnings provided in Meta's Q3 2024 Earnings presentation and the transformed image that is given to the CNN to predict the buy probability.

Image 1 Image 2 Image 3

	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8
Image 1	0.163	0.669	1.457	1.788	1.727	1.913	2.752	2.825
Image 2	0.643	1.067	1.637	1.181	-0.911	1.377	1.047	0.493
Image 3	-2.887	-3.934	-3.812	-4.613	-16.378	-7.156	-7.919	-5.768

Figure 2. Plotting Earnings Images. This figure displays three example earnings images. $E_1, E_2, \ldots$, and E_8 represent quarterly earnings in quarter $q-7$, $q-6$, ..., and q, respectively. Image 1 is a type I earnings image whose quarterly earnings in the most recent eight quarters are all non-negative. Image 2 is a type I earnings image whose maximum quarterly earnings in the most recent eight quarters is positive, and the minimum quarterly earnings in the most recent eight quarters is negative. Image 3 is a type III earnings image whose quarterly earnings in the most recent eight quarters are all non-positive.

Panel A. Pixel value by earnings quality decile

	Earnings quality decile										
	1	2	3	4	5	5.5	6	7	8	9	10
Pixel value	26	51	77	102	128	140	153	179	204	230	255
Color											

Panel B. Earnings images with earnings quality shading

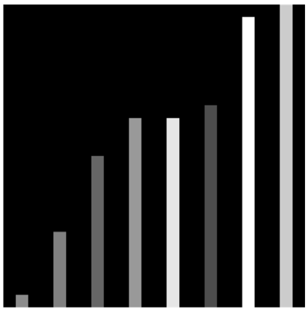


Image 1

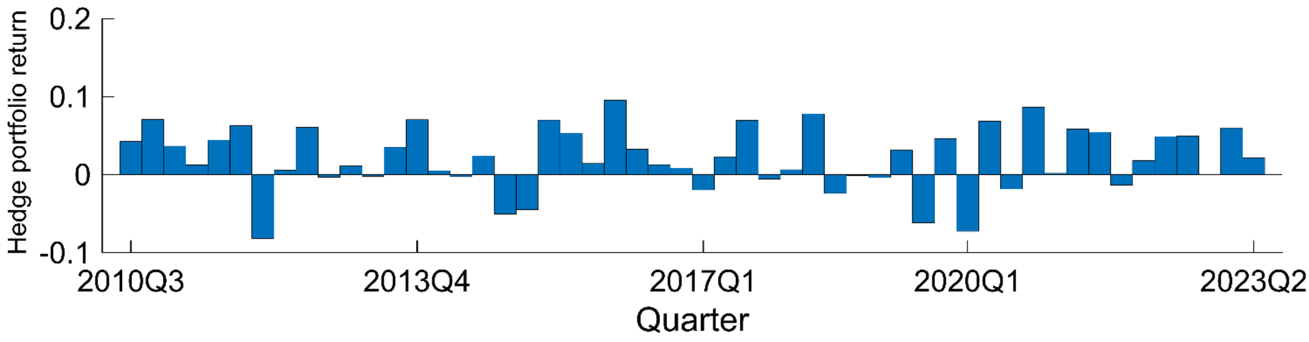
Image 2

Image 3

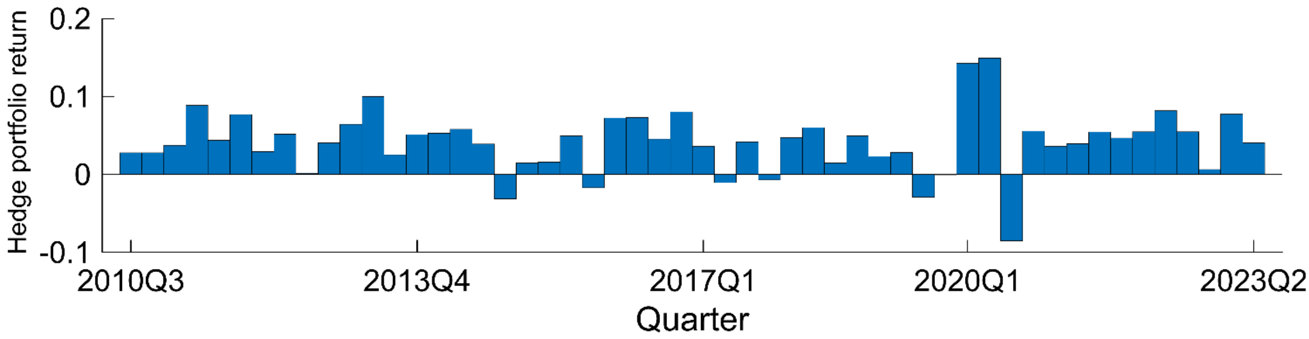
	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8
Image 1	0.163	0.669	1.457	1.788	1.727	1.913	2.752	2.825
Image 2	0.643	1.067	1.637	1.181	-0.911	1.377	1.047	0.493
Image 3	-2.887	-3.934	-3.812	-4.613	-16.378	-7.156	-7.919	-5.768
	EQ_1	EQ_2	EQ_3	EQ_4	EQ_5	EQ_6	EQ_7	EQ_8
Image 1	5.5	5	4	6	9	3	10	8
Image 2	5	9	4	5	9	9	3	5
Image 3	5.5	2	1	1	2	2	1	5

Figure 3. Plotting Earnings Images after incorporating Earnings Quality. Panel A presents for each earnings quality decile the pixel values used to color the earnings bars. Panel B displays three example earnings images incorporated with earnings quality. $E_1, E_2, \ldots,$ and E_8 represent the quarterly earnings in quarter $q-7$, $q-6$, ..., and q, respectively. $EQ_1, EQ_2, \ldots,$ and EQ_8 represent the earnings quality decile for $E_1, E_2, \ldots,$ and E_8, respectively.

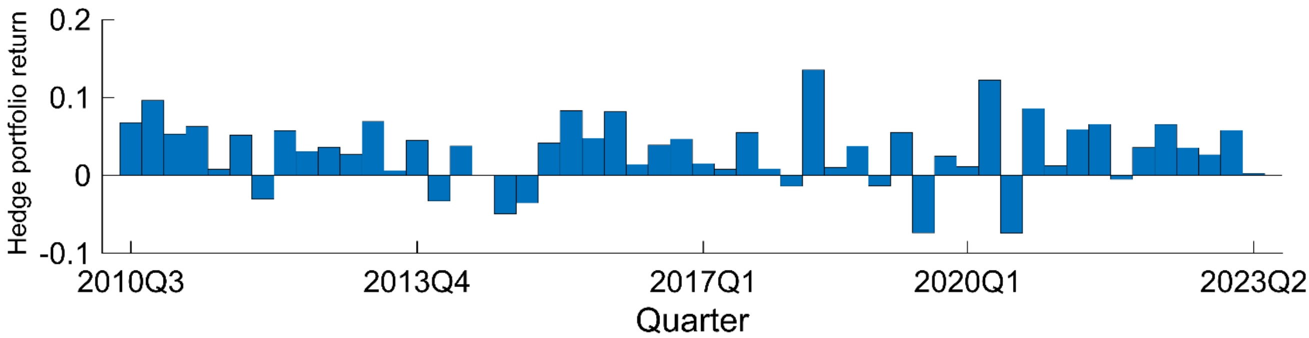
Panel A: CNN



Panel B: CNN+ (CFO version)



Panel C: CNN+ (IBC$_A$ version)



Panel D: CNN+ (OE version)

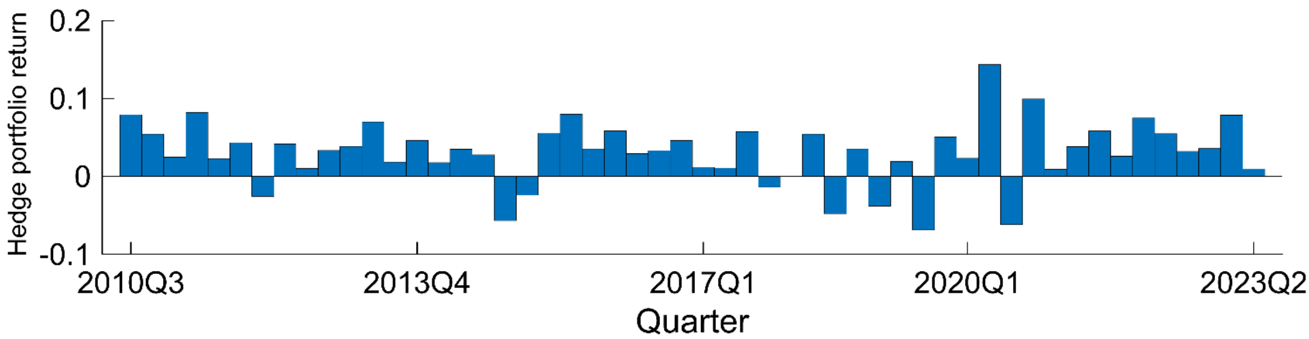


Figure 4. Drift-Predicting Performance Over Time: CNN vs. CNN+. This figure depicts the hedge portfolio return based on CNN and CFO, IBC$_A$, and OE versions of CNN+ in Panels A to D, respectively, during the out-of-sample period (2010Q3-2023Q2). The hedge portfolio returns based on CNN (CNN+) are computed as the post-announcement 63-day Fama-French five-factor and momentum-adjusted buy-and-hold returns between the highest and lowest CNNBP (CNNBP+) deciles. See Appendix B for variable definitions.

Table 1.

Sample Selection

All Compustat firm-quarters with matched CRSP Permno (SHRCD = 10 or 11; EXCHCD = 1, 2 or 3) whose earnings announcement date (Compustat item RDQ) is between 1974Q1 and 2023Q2	814,230
Drop observations with missing RDQ in the most recent eight quarters	(86,536)
Drop observations with earnings announcements on the same date for the same firm in the most recent eight quarters	(3,347)
Drop observations with RDQ less than 30 days away from the previous quarter RDQ in the most recent eight quarters	(16,942)
Drop observations with RDQ before or more than 180 days after the quarter fiscal period end date in the most recent eight quarters	(2,966)
Drop observations with missing earnings (Compustat item IBQ) in the most recent eight quarters	(74,377)
Drop observations whose CRSP daily price at the current quarter RDQ is missing or ≤$1	(102,201)
Drop financial firms (SIC codes between 6000 and 6999) and utility firms (SIC codes between 4900 and 4949)	(109,372)
Drop observations with non-positive book-to-market ratio (BM) or missing market capitalization (SIZE)	(14,585)
Drop observations with more than 30 missing CRSP daily returns in the 120-day window ([-150, -31]) prior to the current quarter RDQ	(24)
Total observations	403,880
Table 2	
In-sample dataset: observations between 1974Q1 and 1993Q4	124,341
Out-of-sample dataset: observations between 1994Q3 and 2023Q2 with non-missing SUE, EA, TREND, PASTRET, PERSIST, VOL, GP, OP, OA, TA, and AG	239,012
Tables 3 to 6	
In-sample dataset: observations between 1990Q4 and 2009Q4	191,118
Out-of-sample dataset: observations between 2010Q3 and 2023Q2 with non-missing SUE, EA, TREND, PASTRET, PERSIST, VOL, GP, OP, OA, TA, and AG and whose 10-K/10-Q filing is released no later than one day after RDQ	43,734

This table reports the sample selection procedures. The in-sample dataset is for model training. The out-of-sample dataset is for testing the out-of-sample model performance. See Appendix B for variable definitions.

Table 2.

Drift-Predicting Performance: CNN vs. LSTM and TFT

	MAR	SAR	FF4	FF6	HMXZ5	DHS3
Panel A: CNN						
	0.035***	0.035***	0.031***	0.032***	0.034***	0.034***
	(6.990)	(6.812)	(8.331)	(8.961)	(8.086)	(7.809)
Panel B: LSTM						
	0.030***	0.029***	0.026***	0.026***	0.027***	0.025***
	(5.684)	(5.349)	(6.931)	(7.631)	(6.351)	(6.001)
Panel C: TFT						
	0.024***	0.022***	0.020***	0.019***	0.020***	0.018***
	(4.802)	(4.495)	(5.853)	(5.934)	(5.513)	(5.024)
Panel D: CNN vs. LSTM						
	0.006*	0.006**	0.006**	0.006***	0.006***	0.008***
	(1.831)	(1.998)	(2.522)	(2.735)	(2.721)	(3.373)
Panel E: CNN vs. TFT						
	0.011**	0.013**	0.011**	0.013***	0.014***	0.015***
	(2.058)	(2.512)	(2.498)	(2.968)	(3.577)	(3.574)

Panels A to C report the average hedge portfolio returns during the out-of-sample period (1994Q3 to 2023Q2) for CNN, LSTM, and TFT, respectively. The hedge portfolio returns based on CNN are computed as the differences in 63-day buy-and-hold abnormal returns (BHAR) following earnings announcements—including market-adjusted return (MAR), size-adjusted return (SAR), and factor-adjusted returns (FF4, FF6, HMXZ5, and DHS3)—between the highest and lowest CNNBP deciles. CNNBP is the *CNN buy* probability generated by CNN. The CNNBP decile cutoffs are based on the distribution of the previous quarter's CNNBP. The average hedge portfolio returns for LSTM and TFT are computed analogously. Panels E and F report the average differences in hedge portfolio returns between CNN and LSTM, and between CNN and TFT, respectively. See Appendix B for variable definitions. Newey and West [1987] t-statistics with three lags are reported in parentheses, and ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 3.

Drift-Predicting Performance: CNN vs. CNN+

	MAR	SAR	FF4	FF6	HMXZ5	DHS3
Panel A: CNN						
	0.021***	0.021***	0.022***	0.021***	0.023***	0.025***
	(2.869)	(2.858)	(4.532)	(4.333)	(4.263)	(3.444)
Panel B: CNN+ (using CFO as the earnings quality measure)						
	0.039***	0.039***	0.042***	0.041***	0.042***	0.044***
	(7.232)	(7.296)	(9.116)	(9.153)	(7.745)	(6.779)
Panel C: CNN+ (using IBC_A as the earnings quality measure)						
	0.031***	0.029***	0.030***	0.031***	0.034***	0.036***
	(4.588)	(4.394)	(5.773)	(5.915)	(6.556)	(5.331)
Panel D: CNN+ (using OE as the earnings quality measure)						
	0.031***	0.030***	0.030***	0.030***	0.032***	0.036***
	(5.276)	(5.468)	(5.908)	(5.636)	(5.929)	(4.915)
Panel E: CNN+ (using CFO as the earnings quality measure) vs. CNN						
	0.018***	0.019***	0.020***	0.020***	0.018***	0.019***
	(2.696)	(3.098)	(3.805)	(3.564)	(3.120)	(0.409)
Panel F: CNN+ (using IBC_A as the earnings quality measure) vs. CNN						
	0.009*	0.008*	0.008**	0.010***	0.011**	0.011**
	(1.871)	(1.727)	(2.649)	(3.148)	(2.524)	(2.571)
Panel G: CNN+ (using OE as the earnings quality measure) vs. CNN						
	0.010**	0.009**	0.008**	0.009**	0.008*	0.011**
	(2.060)	(2.047)	(2.544)	(2.592)	(1.875)	(2.639)

Panels A to D report the average hedge portfolio returns in the out-of-sample period (2010Q3-2023Q2) for CNN and the CFO, IBC_A, and OE versions of CNN+, respectively. Earnings images are unshaded for CNN, while for each CNN+ model, earnings images are shaded according to one of three earnings quality measures (all scaled by lagged assets): CFO (cash flow from continuing operations), IBC_A (adjusted income before extraordinary items), and OE (Dechow-Dichev operating earnings). The hedge portfolio returns based on CNN (CNN+) are computed as the return differences in 63-day buy-and-hold abnormal returns (BHAR) after earnings announcements—including market-adjusted returns (MAR), size-adjusted returns (SAR), and factor-adjusted returns (FF4, FF6, HMXZ5, and DHS3)—between the highest and lowest CNNBP (CNNBP+) deciles. CNNBP (CNNBP+) is the *CNN buy* probability (*CNN+ buy* probability) generated by CNN (CNN+). The CNNBP (CNNBP+) decile cutoffs are based on the distribution of the previous quarter's CNNBP (CNNBP+). Panels E to G report the average differences in hedge portfolio returns between CNN and each of the three CNN+, respectively. See Appendix B for variable definitions. Newey and West [1987] t-statistics with three lags are reported in parentheses, and ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 4.

CNNBP/CNNBP+ and Post-Announcement Returns: Regression Analysis

	(1)	(2)	(3)	(4)
Intercept	0.396	0.366	0.363	0.369
	(0.878)	(0.825)	(0.822)	(0.833)
CNNBP	0.341			
	(0.544)			
CNNBP+		1.658***	1.110**	1.059**
		(3.509)	(2.322)	(2.438)
SUE	0.918*	0.543	0.691	0.707
	(1.881)	(1.122)	(1.508)	(1.494)
EA	1.207**	1.184**	1.148**	1.172**
	(2.588)	(2.503)	(2.433)	(2.521)
TREND	0.560	0.398	0.481	0.457
	(0.995)	(0.724)	(0.865)	(0.812)
PASTRET	0.077	0.022	0.050	0.057
	(0.161)	(0.045)	(0.104)	(0.119)
PERSIST	0.086	0.022	0.020	0.015
	(0.188)	(0.049)	(0.045)	(0.034)
VOL	−0.362	−0.543	−0.620	−0.572
	(−0.519)	(−0.835)	(−0.918)	(−0.839)
SIZE	−1.853*	−1.906*	−1.872*	−1.884*
	(−1.753)	(−1.768)	(−1.739)	(−1.736)
BM	1.022	0.865	0.889	0.904
	(0.926)	(0.807)	(0.838)	(0.838)
GP	2.374*	2.237	2.359*	2.325*
	(1.702)	(1.618)	(1.709)	(1.678)
OP	0.873	0.600	0.770	0.808
	(0.865)	(0.609)	(0.785)	(0.820)
OA	0.550	0.713	0.620	0.628
	(0.696)	(0.896)	(0.790)	(0.797)
TA	−0.931	−0.961	−0.935	−0.949
	(−1.287)	(−1.339)	(−1.298)	(−1.322)
AG	−0.417	−0.304	−0.359	−0.359
	(−0.786)	(−0.552)	(−0.662)	(−0.659)
Adj. R^2	0.023	0.023	0.023	0.023
obs.	43,734	43,734	43,734	43,734

The table presents results of Fama and MacBeth [1973] regressions in the out-of-sample period (2010Q3-2023Q2) using the Fama-French five-factor and momentum-adjusted buy-and-hold return (FF6) during the 63-day window ([+2, +64]) following earnings announcement date as the dependent variable. The *CNN buy* probability (CNNBP) in column 1 is generated by CNN, while the *CNN+ buy* probability (CNNBP+) in columns 2 to 4 is generated by the CFO, IBC_A, and OE versions of CNN+, respectively. The control variables are standardized unexpected earnings (SUE), earnings acceleration (EA), trend in gross profitability (TREND), market capitalization (SIZE), book-to-market ratio (BM), past returns (RET[-30, -2]), earnings persistence (PERSIST), earnings volatility (VOL), gross profitability (GP), operating profitability (OP), total accruals (TA), operating accruals (OA), and asset growth (AG). All variables except for FF6 are converted into scaled ranks ranging from −0.5 to 0.5 with a mean of zero. See Appendix B for variable definitions. Time-series averages of coefficients are multiplied by 100. Newey and West [1987] *t*-statistics with three lags are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 5.

Decomposing the Drift Predictability of CNNBP/CNNBP+: Univariate Analysis

Panel A: CNN													
	Candidate variables												
	SUE	EA	TREND	PASTRET	PERSIST	VOL	SIZE	BM	GP	OP	OA	TA	AG
Step 1: Regress FF6 on CNNBP													
Intercept	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412
	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)
CNNBP	1.281**	1.281**	1.281**	1.281**	1.281**	1.281**	1.281**	1.281**	1.281**	1.281**	1.281**	1.281**	1.281**
	(2.617)	(2.617)	(2.617)	(2.617)	(2.617)	(2.617)	(2.617)	(2.617)	(2.617)	(2.617)	(2.617)	(2.617)	(2.617)
Step 2: Regress CNNBP on a candidate variable													
Intercept	-0.280	-0.380	-0.359	-0.298	-0.357	-0.281	-0.317	-0.319	-0.317	-0.317	-0.317	-0.316	-0.317
	(-1.435)	(-1.191)	(-1.022)	(-0.855)	(-1.078)	(-0.904)	(-0.971)	(-0.977)	(-0.967)	(-0.970)	(-0.969)	(-0.969)	(-0.970)
Candidate	44.200***	29.108***	16.127***	2.429**	5.009***	19.796***	-10.097***	15.224***	-4.061***	-10.989***	-10.400***	-13.293***	-11.117***
	(31.333)	(41.273)	(11.999)	(2.149)	(4.561)	(9.810)	(-7.306)	(17.067)	(-2.831)	(-4.876)	(-10.810)	(-10.151)	(-10.430)
Adj. R^2	19.9%	8.8%	2.9%	0.3%	0.5%	4.7%	1.3%	2.5%	0.5%	2.0%	1.3%	2.1%	1.5%
Step 3: Decompose the CNNBP coefficient from Step 1													
Candidate	0.818	0.475	0.122	0.001	-0.015	0.025	0.238	0.224	-0.077	-0.14	-0.047	0.024	0.14
	63.8%***	37.1%**	9.5%*	0.1%	-1.2%	1.9%	18.6%	17.5%	-6.0%	-10.9%	-3.7%	1.9%	10.9%
	(2.815)	(2.053)	(1.719)	(0.022)	(-0.518)	(0.122)	(1.499)	(1.350)	(-1.029)	(-1.140)	(-0.532)	(0.270)	(1.370)
Residual	0.463	0.806	1.159	1.28	1.296	1.256	1.043	1.057	1.358	1.421	1.328	1.257	1.141
	36.2%	62.9%***	90.5%***	99.9%***	101.2%***	98.1%***	81.4%***	82.5%***	106.0%***	110.9%***	103.7%***	98.1%***	89.1%***
	(1.442)	(3.236)	(12.203)	(31.197)	(34.227)	(6.345)	(6.213)	(6.911)	(15.392)	(9.854)	(13.140)	(11.325)	(9.559)
Panel B: CNN+ (using CFO as the earnings quality measure)													
	Candidate variables												
	SUE	EA	TREND	PASTRET	PERSIST	VOL	SIZE	BM	GP	OP	OA	TA	AG
Step 1: Regress FF6 on CNNBP+													
Intercept	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412
	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)	(0.907)
CNNBP+	2.088***	2.088***	2.088***	2.088***	2.088***	2.088***	2.088***	2.088***	2.088***	2.088***	2.088***	2.088***	2.088***
	(3.670)	(3.670)	(3.670)	(3.670)	(3.670)	(3.670)	(3.670)	(3.670)	(3.670)	(3.670)	(3.670)	(3.670)	(3.670)
Step 2: Regress CNNBP+ on a candidate variable													
Intercept	-0.049	-0.106	-0.129	-0.014	-0.151	-0.034	-0.087	-0.087	-0.084	-0.088	-0.086	-0.087	-0.086
	(-0.199)	(-0.324)	(-0.330)	(-0.039)	(-0.445)	(-0.103)	(-0.259)	(-0.259)	(-0.250)	(-0.262)	(-0.254)	(-0.257)	(-0.256)

Candidate	35.983***	16.662***	18.246***	5.219***	4.742***	8.483***	0.913	0.797	12.042***	9.737***	-13.048***	-3.209***	-5.673***
	(32.186)	(23.101)	(12.490)	(5.701)	(5.944)	(4.248)	(0.521)	(0.984)	(9.032)	(5.505)	(-12.843)	(-2.929)	(-5.801)
Adj. R^2	13.1%	3.0%	3.6%	0.4%	0.4%	1.4%	0.5%	0.1%	1.8%	1.4%	1.9%	0.3%	0.5%
Step 3: Decompose the CNNBP+ coefficient from Step 1													
Candidate	0.655	0.295	0.123	0.008	-0.015	-0.157	-0.137	0.021	0.328	-0.108	-0.063	-0.117	0.046
	31.4%***	14.1%**	5.9%*	0.4%	-0.7%	-7.5%	-6.6%	1.0%	15.7%**	-5.2%	-3.0%	-5.6%	2.2%
	(3.594)	(2.496)	(1.743)	(0.181)	(-0.458)	(-1.528)	(-0.838)	(0.622)	(2.266)	(-0.590)	(-0.676)	(-1.127)	(0.927)
Residual	1.434	1.793	1.965	2.08	2.103	2.245	2.226	2.067	1.76	2.169	2.151	2.205	2.042
	68.6%***	85.9%***	94.1%***	99.6%***	100.7%***	107.5%***	106.6%***	99.0%***	84.3%***	105.2%***	103.0%***	105.6%***	97.8%***
	(6.507)	(11.609)	(18.826)	(27.845)	(45.370)	(17.916)	(14.390)	(43.947)	(9.549)	(14.332)	(16.101)	(23.643)	(24.563)
Panel C: CNN+ (using IBC_A as the earnings quality measure)													
	Candidate variables												
	SUE	EA	TREND	PASTRET	PERSIST	VOL	SIZE	BM	GP	OP	OA	TA	AG
Step 1: Regress FF6 on CNNBP+													
Intercept	0.395	0.395	0.395	0.395	0.395	0.395	0.395	0.395	0.395	0.395	0.395	0.395	0.395
	(0.899)	(0.899)	(0.899)	(0.899)	(0.899)	(0.899)	(0.899)	(0.899)	(0.899)	(0.899)	(0.899)	(0.899)	(0.899)
CNNBP+	1.716***	1.716***	1.716***	1.716***	1.716***	1.716***	1.716***	1.716***	1.716***	1.716***	1.716***	1.716***	1.716***
	(3.311)	(3.311)	(3.311)	(3.311)	(3.311)	(3.311)	(3.311)	(3.311)	(3.311)	(3.311)	(3.311)	(3.311)	(3.311)
Step 2: Regress CNNBP+ on a candidate variable													
Intercept	-0.246	-0.324	-0.348	-0.267	-0.331	-0.220	-0.275	-0.276	-0.273	-0.275	-0.273	-0.274	-0.274
	(-0.954)	(-0.865)	(-0.844)	(-0.649)	(-0.845)	(-0.621)	(-0.711)	(-0.715)	(-0.707)	(-0.712)	(-0.708)	(-0.710)	(-0.711)
Candidate	35.352***	20.593***	15.909***	2.812**	5.528***	25.667***	-8.660***	6.654***	3.140**	-2.923	-14.070***	-9.966***	-8.407***
	(23.102)	(26.074)	(9.656)	(2.499)	(6.987)	(10.858)	(-5.740)	(6.460)	(2.573)	(-1.423)	(-10.762)	(-6.832)	(-8.618)
Adj. R^2	12.9%	4.6%	2.9%	0.3%	0.5%	7.5%	1.1%	0.6%	0.3%	0.7%	2.3%	1.3%	0.9%
Step 3: Decompose the CNNBP+ coefficient from Step 1													
Candidate	0.626	0.347	0.104	0.001	0.001	-0.028	0.195	0.162	0.102	-0.12	-0.051	-0.022	0.103
	36.5%***	20.2%**	6.1%	0.1%	0.1%	-1.6%	11.3%*	9.4%	5.9%	-7.0%	-2.9%	-1.3%	6.0%
	(3.332)	(2.462)	(1.617)	(0.034)	(0.030)	(-0.109)	(1.866)	(1.645)	(1.398)	(-1.427)	(-0.445)	(-0.348)	(1.560)
Residual	1.09	1.368	1.612	1.714	1.715	1.743	1.521	1.554	1.614	1.836	1.766	1.737	1.613
	63.5%***	79.8%***	93.9%***	99.9%***	99.9%***	101.6%***	88.7%***	90.6%***	94.1%***	107.0%***	102.9%***	101.3%***	94.0%***
	(5.112)	(7.964)	(19.772)	(32.517)	(30.012)	(6.026)	(9.834)	(13.253)	(17.782)	(18.189)	(13.682)	(17.763)	(13.944)
Panel D: CNN+ (using OE as the earnings quality measure)													
	Candidate variables												
	SUE	EA	TREND	PASTRET	PERSIST	VOL	SIZE	BM	GP	OP	OA	TA	AG
Step 1: Regress FF6 on CNNBP+													
Intercept	0.394	0.394	0.394	0.394	0.394	0.394	0.394	0.394	0.394	0.394	0.394	0.394	0.394

	(0.894)	(0.894)	(0.894)	(0.894)	(0.894)	(0.894)	(0.894)	(0.894)	(0.894)	(0.894)	(0.894)	(0.894)	(0.894)
CNNBP+	1.760***	1.760***	1.760***	1.760***	1.760***	1.760***	1.760***	1.760***	1.760***	1.760***	1.760***	1.760***	1.760***
	(3.905)	(3.905)	(3.905)	(3.905)	(3.905)	(3.905)	(3.905)	(3.905)	(3.905)	(3.905)	(3.905)	(3.905)	(3.905)
Step 2: Regress CNNBP+ on a candidate variable													
Intercept	-0.211	-0.295	-0.307	-0.279	-0.298	-0.200	-0.251	-0.253	-0.251	-0.252	-0.250	-0.251	-0.251
	(-0.828)	(-0.781)	(-0.756)	(-0.682)	(-0.763)	(-0.571)	(-0.653)	(-0.657)	(-0.650)	(-0.654)	(-0.650)	(-0.651)	(-0.652)
Candidate	36.416***	20.014***	18.358***	3.105***	5.343***	25.192***	-10.173***	6.771***	3.750***	-4.089**	-16.517***	-11.235***	-8.986***
	(25.742)	(29.071)	(11.416)	(2.916)	(6.043)	(10.532)	(-6.675)	(6.998)	(2.948)	(-2.139)	(-11.110)	(-7.606)	(-9.151)
Adj. R^2	13.4%	4.2%	3.7%	0.3%	0.5%	7.3%	1.4%	0.6%	0.4%	0.7%	3.1%	1.6%	1.0%
Step 3: Decompose the CNNBP+ coefficient from Step 1													
Candidate	0.656	0.339	0.134	-0.013	-0.006	-0.009	0.224	0.164	0.127	-0.105	-0.058	0.004	0.099
	37.3%***	19.3%**	7.6%*	-0.7%	-0.3%	-0.5%	12.7%*	9.3%*	7.2%*	-6.0%	-3.3%	0.3%	5.6%
	(3.563)	(2.583)	(1.944)	(-0.295)	(-0.178)	(-0.040)	(1.88 0)	(1.856)	(1.738)	(-1.247)	(-0.444)	(0.064)	(1.470)
Residual	1.104	1.421	1.626	1.773	1.766	1.77	1.536	1.596	1.633	1.865	1.818	1.756	1.661
	62.7%***	80.7%***	92.4%***	100.7%***	100.3%***	100.5%***	87.3%***	90.7%***	92.8%***	106.0%***	103.3%***	99.7%***	94.4%***
	(5.344)	(9.684)	(16.390)	(34.286)	(26.433)	(7.104)	(10.217)	(12.660)	(17.547)	(15.949)	(12.617)	(16.898)	(15.260)

Using Fama-MacBeth [1973] cross-sectional regressions, the relation between CNNBP (or CNNBP+) and the post-announcement 63-day Fama-French five-factor and momentum-adjusted buy-and-hold returns (FF6) is decomposed into a component that is related to a candidate variable and a residual component. In Panel A, Step 1 regresses FF6 on CNNBP (i.e., $\text{FF6}_{i,q+1} = \alpha_q + \beta_q \text{CNNBP}_{i,q} + \varepsilon_{i,q+1}$). Step 2 regresses CNNBP on a candidate variable (i.e., $\text{CNNBP}_{i,q} = a_q + \delta_q \text{Candidate}_{i,q} + \mu_{i,q}$) to decompose $\text{CNNBP}_{i,q}$ into two orthogonal components: the candidate component ($\delta_q \text{Candidate}_{i,q}$) and the residual component ($a_q + \mu_{i,q}$). Step 3 decomposes the β_q coefficient from Step 1 as: $\beta_q = \frac{Cov[\text{FF6}_{i,q+1}, \text{CNNBP}_{i,q}]}{Var(\text{CNNBP}_{i,q})} = \frac{Cov[\text{FF6}_{i,q+1}, \delta_q \text{Candidate}_{i,q}]}{Var(\text{CNNBP}_{i,q})} + \frac{Cov[\text{FF6}_{i,q+1}, (\alpha_q + \mu_{i,q})]}{Var(\text{CNNBP}_{i,q})} = \beta_q^{C} + \beta_q^{R}$. The time-series average of β_q^{C} divided by the time-series average of β_q measures the fraction of the CNNBP's drift predictability explained by the candidate variable and the time-series average of β_q^{R} divided by the time-series average of β_q measures the fraction of the CNNBP's drift predictability unexplained by the candidate variable, with the standard errors of the fractions being determined using the multivariate delta method. In Panels B to D, we replace CNNBP with CNNBP+ generated by the CFO, IBC_{A}, and OE versions of the CNN+ model, respectively, and perform the Steps 1 to 3 in an analogous manner. The candidate variables include standardized unexpected earnings (SUE), earnings acceleration (EA), trend in gross profitability (TREND), market capitalization (SIZE), book-to-market ratio (BM), past returns (RET[-30, -2]), earnings persistence (PERSIST), earnings volatility (VOL), gross profitability (GP), operating profitability (OP), total accruals (TA), operating accruals (OA), and asset growth (AG). All variables except for FF6 are converted into scaled ranks ranging from −0.5 to 0.5 with a mean of zero. See Appendix B for variable definitions. Time-series averages of coefficients are multiplied by 100 and reported with t-statistics in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Table 6:

Decomposing the Drift Predictability of CNNBP/CNNBP+: Multivariate Analysis

	CNN			CNN+ (using CFO as the earnings quality measure)			CNN+ (using IBC_A as the earnings quality measure)			CNN+ (using OE as the earnings quality measure)		
	Coeff.	Fraction	t-stat	Coeff.	Fraction	t-stat	Coeff.	Fraction	t-stat	Coeff.	Fraction	t-stat
Step 1: Regress FF6 on CNNBP (or CNNBP+)												
Intercept	0.412		(0.907)	0.412		(0.907)	0.395		(0.899)	0.394		(0.894)
CNNBP	1.281**		(2.617)									
CNNBP+				2.088***		(3.670)	1.716***		(3.311)	1.760***		(3.905)
Step 2: Regression CNNBP/CNNBP+ on candidate variables												
Intercept	-0.313*		(-1.686)	-0.037		(-0.137)	-0.281		(-1.167)	-0.256		(-1.082)
SUE	39.904***		(33.860)	32.771***		(37.964)	32.520***		(26.375)	33.385***		(27.930)
EA	11.419***		(14.648)	2.408***		(3.312)	6.418***		(7.600)	5.464***		(7.134)
TREND	3.296***		(4.727)	10.861***		(12.907)	6.657***		(8.611)	8.739***		(10.731)
PASTRET	0.971		(1.614)	1.910***		(5.137)	1.258**		(2.457)	1.453***		(2.839)
PERSIST	6.199***		(6.903)	5.060***		(9.413)	6.396***		(10.157)	6.022***		(9.471)
VOL	24.780***		(16.766)	15.116***		(10.391)	31.176***		(18.492)	29.333***		(16.226)
SIZE	3.475***		(2.849)	2.628*		(1.911)	2.141*		(1.844)	0.456		(0.369)
BM	20.550***		(22.906)	10.940***		(11.308)	14.546***		(17.422)	14.295***		(15.732)
GP	1.824		(1.407)	8.162***		(6.536)	3.777***		(3.259)	4.893***		(3.483)
OP	2.493**		(2.491)	15.185***		(17.377)	9.821***		(10.380)	8.643***		(11.384)
OA	-2.850***		(-4.744)	-9.911***		(-18.425)	-6.731***		(-8.666)	-9.365***		(-9.250)
TA	-4.881***		(-10.087)	1.195		(1.557)	-1.683**		(-2.180)	-1.218*		(-1.764)
AG	-1.996**		(-2.481)	-3.522***		(-4.362)	-2.538***		(-3.209)	-2.336***		(-3.986)
Adj. R^2	32.5%			21.5%			26.4%			27.2%		
Step 3: Decompose the CNNBP/CNNBP+ coefficient from Step 1												
SUE	0.756	59.1%***	(2.817)	0.598	28.6%***	(3.575)	0.592	34.5%***	(3.337)	0.617	35.1%***	(3.524)
EA	0.183	14.3%*	(1.952)	0.063	3.0%	(1.656)	0.117	6.8%**	(2.009)	0.097	5.5%*	(1.922)
TREND	0.029	2.2%	(1.413)	0.064	3.1%	(1.385)	0.035	2.1%	(1.095)	0.061	3.5%*	(1.676)
PASTRET	−0.017	−1.3%	(−0.587)	−0.008	−0.4%	(−0.325)	−0.018	−1.1%	(−0.637)	−0.027	−1.5%	(−0.981)
PERSIST	−0.031	−2.4%	(−0.848)	−0.022	−1.1%	(−0.650)	−0.015	−0.9%	(−0.441)	−0.020	−1.2%	(−0.596)

VOL	−0.021	−1.7%	(−0.089)	−0.154	−7.4%	(−1.383)	−0.044	−2.6%	(−0.144)	−0.030	−1.7%	(−0.109)
SIZE	−0.148	−11.5%	(−1.176)	−0.113	−5.4%	(−0.842)	−0.083	−4.9%	(−1.050)	−0.062	−3.5%	(−0.928)
BM	0.296	23.1%	(1.274)	0.156	7.5%	(1.081)	0.248	14.5%	(1.449)	0.244	13.9%	(1.477)
GP	0.083	6.5%	(1.489)	0.261	12.5%**	(2.568)	0.131	7.7%*	(1.980)	0.162	9.2%**	(2.077)
OP	−0.041	−3.2%	(−0.805)	−0.005	−0.2%	(−0.029)	−0.017	−1.0%	(−0.160)	−0.020	−1.2%	(−0.244)
OA	−0.062	−4.8%	(−1.235)	−0.027	−1.3%	(−0.301)	−0.025	−1.5%	(−0.380)	−0.049	−2.8%	(−0.582)
TA	0.034	2.6%	(0.839)	−0.035	−1.7%	(−0.639)	−0.018	−1.1%	(−0.484)	0.009	0.5%	(0.435)
AG	0.033	2.5%	(0.745)	0.045	2.2%	(1.136)	0.042	2.5%	(1.239)	0.041	2.3%	(1.506)
Residual	0.187	14.6%	(0.407)	1.266	60.6%***	(4.820)	0.771	44.9%***	(2.983)	0.736	41.8%**	(2.399)
Total	1.281***	100.0%	(2.617)	2.088***	100.0%	(3.670)	1.716***	100.0%	(3.311)	1.760***	100.0%	(3.905)

Using Fama-Macbeth [1973] cross-sectional regressions, the relation between CNNBP (or CNNBP+) and the post-announcement 63-day Fama-French five-factor and momentum-adjusted buy-and-hold returns (FF6) is decomposed into 13 components each related to a candidate variable and a residual component. The standard errors of the fractions of the relation explained are determined using the multivariate delta method. The 13 candidate variables are standardized unexpected earnings (SUE), earnings acceleration (EA), trend in gross profitability (TREND), market capitalization (SIZE), book-to-market ratio (BM), past returns (RET[−30, −2]), earnings persistence (PERSIST), earnings volatility (VOL), gross profitability (GP), operating profitability (OP), total accruals (TA), operating accruals (OA), and asset growth (AG). All variables except for FF6 are converted into scaled ranks ranging from −0.5 to 0.5 with a mean of zero. See Appendix B for variable definitions. Time-series averages of coefficients are multiplied by 100 and reported with t-statistics in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.