

CFR working paper NO. 25-06

**interpretable machine Learning for
earnings forecasts:
Leveraging high-dimensional
financial statement data**

D. Hess • F. Simon • S. Weibels

**centre for financial research
cologne**

Interpretable Machine Learning for Earnings Forecasts: Leveraging High-Dimensional Financial Statement Data^{*}

Dieter Hess[†]

Frederik Simon[‡]

Sebastian Weibels[§]

First Version: Nov. 2023

This Version: Feb. 2025

^{*}We thank Tobias Bauckloh, Mario Hendriock, Alexander Kempf, Cédric Lesage, Roman Liesenfeld, Maximilian Müller, Michael Overesch, Tom Zimmermann and participants of the EAA Meeting 2024 as well as the CAAA Annual Conference 2024 for helpful comments and suggestions.

[†]University of Cologne, Department of Finance and Centre for Financial Research (CFR), hess@wiso.uni-koeln.de

[‡]LIQID Asset Management, frederik.simon@liquid.de

[§]University of Cologne, Institute of Econometrics and Statistics and Centre for Financial Research (CFR), weibels@wiso.uni-koeln.de

Interpretable Machine Learning for Earnings Forecasts: Leveraging High-Dimensional Financial Statement Data

Abstract

We predict earnings for forecast horizons of up to five years by using the entire set of Compustat financial statement data as input and providing it to state-of-the-art machine learning models capable of approximating arbitrary functional forms. Our approach improves prediction one year ahead by an average of 11% compared to the traditional linear approach that performs best. This superior performance is consistent across a variety of evaluation metrics as well as different firm subsamples and translates into more profitable investment strategies. Extensive model interpretation reveals that income statement variables, especially different definitions of earnings, are by far the most important predictors. Conversely, we find that while income statement variables decline in relevance, balance sheet information becomes more significant as the forecast horizon extends. Lastly, we show that the influence of interactions and nonlinearities on the machine learning forecast is modest, but substantial differences between firm subsamples exist.

JEL classification: G11, G12, G17, G31, G32, M40, M41

Keywords: Earnings Forecasts, Cross-Sectional Earnings Models, Machine Learning

1 Introduction

Future earnings play a crucial role in determining the intrinsic value of assets (e.g., Monahan, 2018). Analysts, in particular, use earnings forecasts to derive buy/sell recommendations for stocks (e.g., Schipper, 1991; Brown, 1993). Earnings also play a key role in corporate decision-making, serving as a primary financial metric for external stakeholders (Graham et al., 2005). Lastly, as demonstrated by Ball and Brown (1968), earnings are directly related to stock returns. Return expectations can even be derived directly from predicted earnings in the form of the implied cost of capital (ICC) of a company (e.g., Gordon and Gordon, 1997; Claus and Thomas, 2001; Gebhardt et al., 2001; Easton, 2004; Ohlson and Juettner-Nauroth, 2005).

Earnings forecasts are generally obtained from either analysts or statistical models. Traditionally, statistical approaches have been primarily simple and linear in nature (e.g., Hou et al., 2012; Li and Mohanram, 2014).¹ With the advent of more advanced statistical models, i.e., machine learning approaches in particular, various more flexible approaches have been proposed by researchers (e.g., Cao and You, 2021; Jones et al., 2023).

As outlined by Gu et al. (2020) and Israel et al. (2020), machine learning introduces flexibility along three dimensions: first, it accommodates complex functional forms. That is, in contrast to the traditional linear approaches that have been used to predict earnings, machine learning allows for complex non-linear functional forms. Second, machine learning enables the use of large conditioning information sets, allowing researchers to uncover previously undetected relationships. Third, machine learning incorporates advanced optimization techniques, such as regularization, to prevent overfitting.

To the best of our knowledge, research on earnings prediction has primarily focused on the first dimension. Put differently, researchers have proposed the use of different types of machine learning algorithms to approximate the possibly complex functional form that relates predictors and future earnings (e.g., Cao and You, 2021). In general, more complex machine learning models are likely to yield more accurate earnings predictions. In fact, Kelly et al. (2022) demonstrate that increasing complexity, defined as the ratio of model parameters to data, consistently improves out-of-sample prediction performance in return forecasting. They explicitly recommend to use

¹Hereafter, we refer to these models as traditional models.

rich non-linear model specifications rather than simple linear ones. We argue that further research along this dimension thus bears little additional insights. Moreover, extant studies on earnings prediction typically only assess a rather limited set of predictor variables. An important exception is the study by Chen et al. (2022) which exploits the entirety of Extensible Business Reporting Language (XBRL) data. Yet, they restrict their analysis to the prediction of earnings changes and do not predict earnings per se.² The fact that machine learning allows for the use of large conditioning information sets as mentioned by Israel et al. (2020) is rather unexploited in this context so far.

We address this gap by predicting annual earnings per share using a comprehensive set of variables – namely, the full set of financial statement variables from Compustat. This allows us to thoroughly analyze how fundamental accounting-based information drives and relates to future earnings. We employ a selection of prominent, flexible machine learning models, i.e., a random forest model (RF), a gradient boosted tree model (GBT), a gradient boosted tree model with dropout (DART), a feed-forward neural network (NN) and an ensemble of the aforementioned models (ENML).³ We further show how our approach relates to the most widely used traditional linear approaches, namely a simple model only including earnings as a predictor (L) (Gerakos and Gramacy, 2012), the HVZ-model (Hou et al., 2012), the EP-model (Li and Mohanram, 2014), the RI-model (Li and Mohanram, 2014) and an ensemble of the aforementioned models (ENTD).

A key contribution of our study is extensive model interpretation. Understanding the inner workings of prediction models is a fundamental requirement in asset management applications (Israel et al., 2020). However, due to their complexity, machine learning models are hard to interpret. In the machine learning earnings prediction case, model interpretation has thus far been restricted to metrics of variable importance and partial dependencies. More so, these metrics have usually been applied to a predetermined, restrictive set of predictor variables as mentioned above. Put differently, researchers typically choose or construct a set of predictor variables that they deem important before estimating the model. After estimating the models, they derive the extent to which variables from this predetermined set contribute to the predictions and assess the

²In fact, as an extension to their main analysis, they also predict earnings levels. However, they only use a set of 24 variables for this and not the high-dimensional XBRL data set which they use for their main analysis. Surprisingly, their machine learning approach does worse than a simple random walk model. They conclude that earnings levels are hard to predict and use this as an argument for the fact that they focus on earnings changes in their primary analysis.

³The model selection is based on Bali et al. (2023).

partial dependencies of predicted earnings in regards to them.

This study aims to broaden the limited scope of model interpretation found in the existing literature on earnings forecasts: first, we derive the relative importance of variables using SHAP (SHapley Additive exPlanations) values, an approach based on cooperative game theory (Lundberg and Lee, 2017).⁴ Since we do not select or construct variables beforehand as done in comparable studies (e.g., Hansen and Thimsen, 2020; Cao and You, 2021), we are able to holistically infer which out of all the financial statement variables are important from a statistical perspective. We also derive the relative importance of different groups of financial statement variables, such as cash flow statement (CF/S) variables, income statement (I/S) variables and balance sheet (B/S) variables. Importantly, we conduct this analysis for forecast horizons of up to five years. This allows us to assess how different (groups of) variables might vary in terms of their predictive power, depending on the forecast horizon considered. Second, we analyze non-linearity in the context of earnings prediction. In addition to partial dependencies that extant studies focus on in that context (e.g., Cao and You, 2021; Chen et al., 2022), we infer the degree to which different types of non-linearity play a role. More precisely, we show the degree to which interaction effects across financial statement variables and other forms of non-linearities, i.e., non-linearity of the functional form, are important by means of surrogate modeling. This approach is completely transparent, intuitive and easy to replicate, irrespective of the model or software used. Again, we conduct these analyses for forecast horizons of up to five years. The last component of our model interpretation consists of stratifying the sample into different subsamples. Specifically, we group firms-years according to the life cycle stage, the size tercile, and the industry which the respective firm is in. We then assess differences in accuracy, variable importance as well as the impact of interactions and non-linearities across groups. This enables us to more thoroughly understand how future earnings are related to current fundamental data and to infer which types of firms require more nuanced and complex earnings predictions models than others.

Our results can be summarized as follows: first, in line with previous research on model averaging, we find that ensemble models improve predictive accuracy relative to their individual

⁴SHAP values are a way of explaining the results of any machine learning model. They are based on a game-theoretic approach that measures the contribution of each player to the final outcome. In machine learning, each variable is assigned an importance value that represents its contribution to the model's outcome (Lundberg and Lee, 2017).

components across both traditional and machine learning approaches. For instance, in one-year-ahead predictions, the machine learning ensemble further improves accuracy beyond the best-performing individual machine learning model, achieving an additional 1% gain. Given that the best machine learning model already substantially outperforms the best traditional model, this additional improvement underscores the value of ensemble modeling. Notably, the machine learning ensemble consistently outperforms all component models across all forecast horizons

In terms of bias, the traditional and the machine learning approaches yield very similar results. Both types of models yield mostly unbiased predictions for forecast horizons of up to three years. An interesting exception are the traditional models who tend to underestimate one-year-ahead earnings. For forecasts horizons of four and five years, the models considered begin to systematically overestimate earnings. However, the machine learning models are consistently less biased in terms of levels as compared to their traditional analogues.

Mirroring previous findings on machine learning earnings predictions, we further show that our machine learning approaches constantly outperform traditional linear approaches in terms of accuracy (e.g., Cao and You, 2021). For the one-year forecast horizon, the best performing machine learning model, i.e., the ENML is around 11% more accurate than the best performing traditional model, i.e., the ENTD. Even for long forecast horizons of five years, we find that the ENML beats the ENTD in terms of accuracy by around 6%. Furthermore, assessing accuracy differences across out-of-sample periods shows that model performance converges in the periods following the financial crisis, and diverges in favor of the machine learning approaches afterwards.

Assessing the degree to which the models are able to explain out-of-sample variation in earnings makes an even more convincing case for the ML approaches. In fact, the average out-of-sample R^2 (OOS R^2) of the ENML is between 15% and 23% higher than that of the ENTD for the forecast horizons considered.

Lastly, we show that the more accurate predictions of the machine learning approach translate into more profitable portfolios based on ICC. An investment strategy which invests into the top ICC quintile based on ENML forecasts each year outperforms an analogous strategy which utilizes ICC based on ENTD forecasts. Moreover, we show that both ICC strategies outperform a value-weighted market portfolio in terms of cumulative returns.

Turning to model interpretation of the machine learning approach, we find that current I/S

variables, particularly current earnings, play a dominant role in short-term earnings predictions. This aligns with traditional models, which also heavily rely on lagged earnings as a key predictor. However, a key novel finding is that with increasing forecast horizon, variable importance shifts significantly, becoming more balanced among financial statement types. For one-year-ahead predictions, I/S variables contribute around 66% while B/S variables contribute around 19% to total importance. Over longer horizons, this imbalance diminishes: for $t + 5$ -predictions, I/S variables contribute around 47%, while B/S variables contribute around 37%. CF/S variables consistently contribute around 15% to total importance throughout the forecast horizons considered. Put differently, the longer the forecast horizon, the more important B/S information becomes. This shift suggests that while short-term earnings primarily reflect income statement dynamics, longer-term earnings expectations are increasingly shaped by firms' balance sheet fundamentals.

Further disentangling the effects of different components of financial statement information reveals that certain pieces of financial statement information dominate others. For example, debt and supplemental information resemble the most important pieces of B/S information. Turning to the CF/S, we find that variables related to the operating cash flow are much more relevant than variables related to either the investing cash flow or the financing cash flow. Lastly, turning to the I/S, we find that especially EBIT and net income are important, contributing around 18% and 25% to total importance for one-year-ahead forecasts, respectively. Interestingly, while the importance of net income declines to approximately 12% for five-year forecasts, EBIT maintains a consistent level of significance. This suggests that information which is less exposed to accounting manipulation gains relevance for longer-term forecasts.

We further show that for one-year-ahead predictions, a linear surrogate model is able to explain around 90-94% of the variation in machine learning earnings predictions across out-of-sample periods. Interactions and non-linearities contribute somewhat equally to explaining the remaining variation. In fact, adding interactions increases the R^2 by around 3-4%. We attribute the remaining unexplained variation to other types of non-linearities, which are not captured by interactions, i.e., non-linearity of the functional form. As the forecast horizon increases, the linear surrogate model provides a progressively less accurate approximation of the relationship between predictions and inputs, suggesting that longer-term earnings forecasts involve increasingly complex functional dependencies that require more flexible models to capture effectively.

The last part of our model interpretation consists of stratifying the sample into different types of firms, as outlined above. We find that generally, the ENML dominates the ENT D in terms of accuracy for each firm subsample considered. However, significant differences in accuracy (improvements) across subsamples exist. Most notably, this can be seen for firms in different life cycle stages. Earnings of firms in the beginning and the end of their life cycle are forecasted with substantially lower accuracy than those of firms in the middle of their life cycle. Importantly, it is precisely for these firms – where earnings predictability is inherently more difficult – that the ENML achieves the most significant improvements over the ENT D. Moreover, we find that for these firms, interactions account for a larger portion of the ENML predictions than for firms in the middle of their life cycle. Again, this suggests that when the prediction task is more complex – whether due to longer forecast horizons or greater firm heterogeneity – models that allow for richer interactions and non-linear structures provide a distinct advantage.

The remainder of this paper is structured as follows: In Section 2 we outline the relevant literature that we are contributing to. In Section 3 we describe our empirical approach. We evaluate and compare our approach in Section 4. In Section 5 we provide extensive interpretation. Finally, Section 6 concludes the study.

2 Related Literature

Our work relates to three strands of literature in particular. First, we contribute to the literature on machine learning applications in finance. Machine learning methods have become the prevalent way of conducting prediction exercises, primarily due to their superiority in terms of flexibility as compared to traditional econometric methods and their efficacy in regards to large sets of input data (e.g., Israel et al., 2020; Kelly et al., 2022). For example, Gu et al. (2020) show how different machine learning approaches perform in terms of predicting stock returns and Bali et al. (2023) apply machine learning to the task of predicting option returns. Our study is similar, in the sense that we predict another financial variable, i.e., earnings, with machine learning.

Second, we apply state-of-the-art techniques to interpret our machine learning predictions, thereby explicitly responding to the "need for interpretability" of financial machine learning models as formulated by Israel et al. (2020). We hence also contribute to the literature that aims to

foster transparency and understanding of machine learning methods for prediction in finance research, such as e.g., Bali et al. (2023) who extensively assess machine learning option return predictions.

Third, we contribute to the literature on model-based earnings forecasts. Traditionally, researchers have suggested predicting earnings using time-series regression models (e.g., Ball and Watts, 1972; Albrecht et al., 1977; Watts and Leftwich, 1977), cross-sectional regression models (e.g., Hou et al., 2012; Li and Mohanram, 2014; Harris and Wang, 2019) or even simple random walk models (e.g., Li and Mohanram, 2014). More recently, however, several machine learning earnings prediction approaches have been implemented in the research (e.g., Hansen and Thimsen, 2020; Cao and You, 2021; Chen et al., 2022; Hendriock, 2022; Campbell et al., 2023; Jones et al., 2023; Van Binsbergen et al., 2023). However, the extant literature differs from our study in several key aspects which are outlined in the following.

Hansen and Thimsen (2020) also estimate a range of machine learning methods. They use a more high-dimensional input vector than other studies on predicting level earnings. However, it is still restrictive in the sense that it is based on prior research and not as high-dimensional as our data. Moreover, in contrast to our study, no model interpretation is provided.

The study of Cao and You (2021) also encompasses different machine learning models. However, they use a more restrictive set of input variables than us and provide only limited model interpretation. Moreover, Cao and You (2021) validate their models using traditional cross-validation (and a limited hyperparameter space). This, however, destroys the temporal structure of the observations and introduces information leakage (Gu et al., 2020). We preserve the temporal ordering of the observations by using fixed training, validation and test intervals.

Chen et al. (2022) use a single model (gradient boosted trees) as opposed to our multi-model approach. Furthermore, they predict binary earnings changes, while we focus on predicting level earnings.

Hendriock (2022) suggests predicting earnings by predicting the complete conditional density function. However, he restricts his input variable space to the one as defined by traditional linear models. Moreover, he does not provide model interpretation.

Campbell et al. (2023) benchmark an extensive range of machine learning model specifications with the aim of identifying the ones which compare the best to analyst forecasts. Apart from

the fact that their model choices differ from ours, they use a different, smaller set of inputs, i.e., the Wharton Research Data Services (WRDS) Financial Suite Ratios extended by some additional variables like e.g., the stock return. Furthermore, they provide limited model interpretation as their primary focus is on the aforementioned horse-race between model specifications.

Jones et al. (2023) use a single model, i.e., a gradient boosted tree model algorithm, as opposed to our ensemble approach. In contrast to our study, their target variable is return on net operating assets and they use a set of six ratios as their predictors. Another crucial difference is that Jones et al. (2023) exclusively forecast earnings (changes) in $t + 1$, while we forecast earnings (per share) for horizons $t + 1$ to $t + 5$. Finally, to the best of our knowledge, as the only other study in this context, they assess the impact of interactions. However, their method of doing so and their predictor variables differ from ours. Our surrogate modeling approach is easily applicable to any type of model, in any software and further allows us to explicitly determine the effect of interaction effects and non-linearity in parameters. Interestingly, our findings in regards to interactions differ strongly from the ones provided by Jones et al. (2023). They find substantial importance of interactions, while we find that interactions among financial statement variables are irrelevant.

Van Binsbergen et al. (2023) use a random forest model to predict earnings conditional on financial ratios, similar to Campbell et al. (2023). In contrast, our approach involves estimating a spectrum of machine learning models individually and in ensemble configurations. Moreover, we utilize an entirely distinct set of input data, specifically the comprehensive Compustat financial statement dataset. Finally, unlike their study, which primarily assesses analyst biases, our analysis encompasses detailed explanations of model predictions.

Summing up, to the best of our knowledge, we are the first to predict level earnings per share for forecast horizons of up to five years using the entirety of available Compustat financial statement variables. Furthermore, we contribute novel guidance for future research on earnings (per share) predictions by thoroughly interpreting our state-of-the-art machine learning approaches using model agnostic and easily applicable methods, something that has been not done extensively thus far.

3 Empirical Approach

3.1 General setup

We express earnings E of firm i in period $t + \tau$ as the expectation in period t plus an error ϵ :

$$E_{i,t+\tau} = \mathbb{E}_t[E_{i,t+\tau}] + \epsilon_{i,t+\tau}. \quad (1)$$

Every earnings prediction model can be viewed as an attempt to estimate $\mathbb{E}_t[E_{i,t+\tau}]$, the expected value. More precisely, we assume that expected earnings in $t + \tau$ are a function of a vector of inputs X of firm i known at time t :

$$\mathbb{E}_t[E_{i,t+\tau}] = f(X_{i,t}). \quad (2)$$

It becomes evident that modelling expected earnings consists of three crucial parts. First, one has to determine which inputs enter the model, that is, how X is defined. Second, one has to decide which functional form $f(\cdot)$ takes on. This corresponds to the decision about which empirical model to choose from. Lastly, one has to decide on how to estimate $f(\cdot)$, i.e., which statistical loss function to minimize. The latter is not the explicit focus of this study and thus we keep it simple. We follow the original implementations of the traditional models and estimate them using the mean squared error (MSE). In case of the machine learning models, we follow Gu et al. (2020) and estimate the models both using the MSE and the mean absolute error (MAE) and report the predictions based on the loss function that leads to more accurate forecasts according to the price scaled absolute forecast error (PAFE) at the 1-year horizon.⁵

Thus, apart from the decision regarding the loss-function, the two contrasting *extreme* approaches to predicting earnings are: (1) actively making the decision which variables and which functional form to assume ex ante, and (2) letting the data speak by selecting a model that permits flexible functional forms and providing it with the entire data (or at least a very large set of variables) available. The former corresponds to the traditional prediction approaches suggested by e.g., Hou et al. (2012). To the best of our knowledge, the latter has not yet been implemented

⁵To be precise, Gu et al. (2020) choose either the MSE or the Huber loss, depending on which performs better. We choose either the MSE or the MAE, depending on which performs better.

for earnings per se, a noteworthy exception being Chen et al. (2022) who employ this approach for (binary) earnings *changes*. There are approaches that fall in between (1) and (2). In these approaches, either the functional form or the input vector is restricted significantly (e.g., Hendriock, 2022; Van Binsbergen et al., 2023).

Our study focuses on assessing the second, flexible approach (2) and comparing it to traditional approaches (1). We more thoroughly elaborate on the choice of the input vector and of the model in 3.2 and 3.3, respectively.

3.2 Data

US annual financial statement data is obtained from Compustat. Our sample period ranges from 1988 to 2023. This is due to the fact that CF/S-data is only sparsely available prior to 1988. To conduct the ICC portfolio evaluation, we add price and return data from CRSP to the Compustat data used for model estimation. Moreover, we drop observations with missing prices, prices smaller than 1\$, missing or zero common shares outstanding or missing earnings. This results in a final sample for model estimation that consists of 158,809 observations.

For our machine learning models, we use the Compustat financial statement items as predictors. We drop variables with more than 50% of observations missing or no observations in any of the cross-sections (i.e., estimation years), yielding 193 variables.⁶ An overview over these variables is given in Table B.2 in the Appendix. Analogous to Chen et al. (2022), we include lags and first-order differences of these variables, resulting in a set of 579 predictor variables in total. For the traditional models, we construct input variables according to the respective models. An overview over these variables is given in Table A.1 in the Appendix. All our variables, including our target variable, are scaled by common shares outstanding and winsorized at the 1%- and 99%-level, respectively. Finally, since neural networks are sensitive to scale differences across input variables, we standardize each variable.

⁶We also drop variables already scaled by shares. The reason for that is, that we scale all our variables by shares as mentioned below and hence these variables are redundant. However, this only pertains to five variables in our study.

3.3 Models

We estimate two groups of models. The first group consists of popular simple linear models that have been introduced in the literature thus far. All of these models assume a linear additive relation between earnings and some low-dimensional input vector, i.e.,

$$\mathbb{E}_t[E_{i,t+\tau}] = \beta X_{i,t}, \quad (3)$$

where β denotes a vector of coefficients. The models differ in terms of which variables the input vector consists of. A more detailed description is given in Appendix A.1. A difference to be noted is that the models also slightly differ in how they define the output. While the RI and the EP model use earnings per share, the HVZ model uses earnings. In this study, we define earnings as "income before extraordinary items" (Compustat variable: ib). As mentioned above, we consistently scale our output as well as our input variables by common shares outstanding in all of the models. Our forecast horizon, both for this and the following group of models spans from $\tau = 1$ to $\tau = 5$.

The second group of models consists of flexible models that are able to approximate arbitrary complex functional forms. Put differently, we do not assume any specific type of functional form when estimating these models. Analogous to Bali et al. (2023), we estimate a random forest model (RF), a gradient boosted tree model with (DART) and without dropout (GBT) and a neural net (NN). Importantly, we try to restrict our input vector as little as possible. Specifically, we feed the models the 579 variables as outlined above, including the lags and first-level differences. We argue that this input vector corresponds to a proxy for the entirety of (relevant) financial statement input variables. Note that extending the input vector even further implies more computational effort.⁷

Since studies in other realms of financial forecasting have shown that using an ensemble of models may prove superior to using single models (e.g., Bali et al., 2023), we derive the equally weighted average prediction for both groups of models, i.e., we derive two ensemble model predictions:

⁷We could have also included other variables like price, analyst forecasts, etc. (e.g., Campbell et al., 2023; Van Binsbergen et al., 2023). However, the focus of our study is the thorough analysis of the relationship between fundamental accounting-based information and future earnings.

$$\mathbb{E}_t^{(En)}[E_{i,t+\tau}] = \frac{1}{J} \sum_{j=1}^J \mathbb{E}_t^{(j)}[E_{i,t+\tau}], \quad (4)$$

where $j \in J$ denotes the respective single model and En denotes the respective ensemble model.

The ensemble of the traditional models is denoted by ENT D and the ensemble of the fully flexible machine learning prediction models is denoted by ENML.

3.4 Out-of-sample approach

We employ a rolling window strategy to obtain our out-of-sample prediction results. Specifically, for the machine learning models, we divide our data into training, validation and test sets. For each forecast horizon τ , the process for generating forecasts as of t proceeds as follows: we train our models using earnings from $t - 11$ to $t - 2$ as output and corresponding financial statement data lagged by τ as predictors. Next, we tune the machine learning models using earnings from $t - 1$ to t as output and lagged financial statement data by τ as predictors. Tuning involves determining the optimal hyperparameter values for the model, as detailed in Table B.1 in the Appendix. Subsequently, earnings predictions for $t + \tau$ are derived by inputting variables from t into the optimized models. This process is repeated recursively, advancing one year at a time. We follow the approach of Hou et al. (2012) and Li and Mohanram (2014), estimating models at the end of June each year, under the assumption of a reporting lag of three to fourteen months for financial statements.⁸

In contrast, traditional linear approaches do not require a tuning window. Hence, we only partition the data into training and test sets when estimating these models. The subsequent steps of the procedure remain unchanged. More precisely, for each forecast horizon τ , models are trained using earnings from $t - 11$ to t as output and corresponding lagged financial statement data by τ as predictors. Earnings predictions for $t + \tau$ are then derived by utilizing predictor variables from t .

⁸Specifically, data from April of year $t - 1$ to March of year t is considered the most recent fiscal year-end data available as of June in year t , capturing the information as of t .

3.5 Evaluation

We evaluate the predictive performance of the models across a range of evaluation metrics. First, we compute the error metrics that are common in the earnings prediction literature (e.g., Hou et al., 2012). For each forecast horizon $\tau \in [1, 2, 3, 4, 5]$, these include the price scaled forecast error (PFE) or bias:

$$PFE_{i,t+\tau} = \frac{E_{i,t+\tau} - \hat{E}_{i,t+\tau}}{Price_{i,t}}, \quad (5)$$

and the price scaled absolute forecast error (PAFE) or accuracy:

$$PAFE_{i,t+\tau} = \frac{|E_{i,t+\tau} - \hat{E}_{i,t+\tau}|}{Price_{i,t}}, \quad (6)$$

where $E_{i,t+\tau}$ denotes actual earnings for firm i in period $t + \tau$, $\hat{E}_{i,t+\tau}$ denotes the respective forecast and $Price_{i,t}$ is the firm's stock price at the end of June in the respective estimation year.

Second, we assess the out-of-sample R^2 ($OOSR^2$) of each individual as well as the ensemble forecasts, i.e., for every out-of-sample period we calculate

$$OOS R^2_{t+\tau} = 1 - \frac{\sum_{i=1}^{n_{t+\tau}} (E_{i,t+\tau} - \hat{E}_{i,t+\tau})^2}{\sum_{i=1}^{n_{t+\tau}} (E_{i,t+\tau} - \bar{E}_{t+\tau})^2}, \quad (7)$$

for each model. Here, $\bar{E}_{i,t+\tau}$ denotes average earnings of firms in period $t + \tau$. Albeit not commonly used in the earnings prediction literature (e.g., Hendriock (2022) being an exception), this evaluation metric is of particular importance for the typical use case of earnings predictions, i.e., long-short ICC portfolios. In this context, predicting cross-sectional variation is much more important than accurately predicting earnings per se, since an investor goes long (short) the stocks which's ICC is high (low) in cross-sectional comparison.

Lastly, we derive the ICC based on the two ensemble forecasts. We follow the literature and calculate ICC following the methods of Gordon and Gordon (1997), Claus and Thomas (2001), Gebhardt et al. (2001), Easton (2004) and Ohlson and Juettner-Nauroth (2005).⁹ More precisely, our ICC estimates are derived as the average of the five aforementioned methods using both the traditional ensemble and the machine learning ensemble earnings predictions. We then construct

⁹A description of the models is provided in Appendix C.

equally weighted long-short zero investment portfolios based on ICC and assess their average performance across the out-of-sample periods.

3.6 Interpretation

A primary contribution of this study is the comprehensive interpretation of the machine learning approach, addressing a key issue in machine learning applications in finance and accounting (Israel et al., 2020). Our study fills a gap in the existing literature, which either lacks model interpretation entirely or provides only limited insights, as discussed above. More precisely, we derive the variable importance and the degree to which different types of non-linearity play a role for our best performing machine learning model, i.e., the machine learning ensemble. Further, we assess differences in accuracy, variable importance as well as the degree of non-linearity across different subsamples of firms.

Variable importance

We determine the importance of the different variables with respect to the earnings prediction. To do so, we compute their SHAP values, a state-of-the-art approach for assessing the importance of input variables which is based on cooperative game theory (Lundberg and Lee, 2017). In essence, SHAP values approximate how a model's prediction changes when knowing the value of a respective input variable. The approach is model-agnostic and allows us to evaluate the importance of input variables irrespective of the model used. We conduct these analyses at both the individual variable level and the grouped-variable level. Specifically, we determine the relative importance of predictors grouped into balance sheet, cash flow statement, and income statement data as well as predictors grouped into current, lagged, and difference variables. Furthermore, we provide an in-depth accounting perspective on which specific types of financial statement information are important by breaking the financial statements down into schematic components. Importantly, we conduct these analyses per forecast horizon.

Non-linearity

In addition to deriving the variable importance, we also evaluate the extent to which non-

linearity plays a role in our machine learning model.

First, we assess the degree to which non-linearity in terms of variables and non-linearity in terms of functional form play a role. In a first step, we regress predicted earnings on the 20 most important input variables using a linear regression model.¹⁰ In a second step, we add all possible two-way interactions to the surrogate regression model from the prior step. Assuming there is some degree of non-linearity present in the fully flexible model, this allows us to disentangle the degree to which non-linearity in terms of variables (i.e., interactions across inputs) and non-linearity in terms of functional form play a role. More specifically, we assess the adjusted in-sample R^2 s of the two surrogate models. The R^2 of the first-step surrogate model indicates the degree to which the predictions are linear. The difference between the R^2 of the first-step surrogate model and the second-step surrogate model indicates the degree to which non-linearity in terms of variables, i.e., (two-way) interactions across financial statement variables, plays a role. We attribute the portion that remains unexplained by the second-step surrogate model to non-linearity in functional form.¹¹

Second, we assess the partial dependence of earnings with respect to the most important input variables. As typically done in the literature, we evaluate the partial dependencies graphically via so-called partial dependence plots. To do so, we fit a non-parametric lowess model (locally weighted linear regression) to the SHAP values of a predictor value of interest (the output) and the associated predictor values (the input) and plot the result. This allows us to approximate the effect of the respective predictor variable on future earnings.

Again, we conduct the surrogate modeling and the partial dependence analyses per forecast horizon.

Subsample analysis

The last part of our model interpretation comprises of a subsample analysis. More precisely, we briefly assess differences in accuracy, variable importance as well as non-linearity across different subsets of firm-year observations. Inter alia, this assessment allows us to determine

¹⁰We only use the 20 most important variables, because otherwise, including all possible two-way interactions in the second step requires an excessive amount of computing power. More so, we find in undocumented results that adding more variables does not significantly change the results.

¹¹Theoretically, the unexplained portion also includes effects of interaction terms of order three and higher. However, we assume that these can be neglected and find evidence for this assumption in undocumented analyses.

whether the superiority of the machine learning model stems from specific subsamples specifically and whether there exist differences in how current fundamental information is related to future earnings across different subsets firms.

We consider three ways of stratifying the sample: first, we stratify predictions according to the life-cycle stage the respective firm is in in the period in which the forecast is conducted. We define life-cycle stages following Dickinson (2011).¹² Second, we stratify predictions according to the firm size, measured via market capitalization as of the period in which the forecast is conducted. Lastly, we stratify predictions according to the industry, defined via the Fama-French 5 industries, a firm is in as of forecast date.¹³

4 Evaluation

4.1 Accuracy and bias

Chen et al. (2022) report that their machine learning approach does worse than a simple random walk type model when predicting level earnings. They explicitly state that level earnings are hard to predict and thus resort to the prediction of earnings changes in their main analysis. In contrast, mirroring findings by e.g., Cao and You (2021) our flexible machine learning models for earnings (per share) level prediction outperform the traditional linear models by a significant margin.

Price scaled forecast error

Table 1 shows the time-series averages of the median PFE for the four traditional, the four machine learning and the two ensemble models. The PFE provides insight into whether the estimated earnings are systematically over- or underestimated (biased) relative to actual earnings.

In general, the two ensemble models appear to be the least biased models for each group of models considered. Both models yield insignificant biases in most cases. This is in line with extant studies on earnings prediction, which report that earnings predictions by statistical models do not exhibit biases as opposed to those by analysts (Hou et al., 2012). In fact, only the traditional

¹²The classification scheme is summarized in Table D.1 in the Appendix.

¹³Industries are defined following the Fama-French 5 industry portfolios classification, available on Kenneth French's website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html) website. The classification scheme is summarized in Table D.2 in the Appendix.

ensemble yields significant biases for two out of five forecast horizons considered. Interestingly, it systematically underestimates earnings in $t + 1$. Both ensemble models tend to systematically overestimate earnings with increasing forecast horizons. However, the machine learning ensemble achieves between 44.80-87.50% lower biases, depending on the forecast horizon.

The results further indicate that at the non-ensemble level across forecast horizons $t + 1$ to $t + 3$, some models yield statistically significant PFEs, with the most biased machine learning model for $t + 1$ being the RF model (0.0059) and the most biased traditional model for $t + 1$ being the RI model (0.0063). Interestingly, the three more sophisticated traditional models systematically underestimate earnings in $t + 1$, reflecting the aforementioned results regarding the ensembles. For $t + 4$ no single model yields statistically significant PFEs, while for $t + 5$, most non-ensemble models yield PFEs that are significantly different from zero. Overall, our findings suggest that non-ensemble machine learning models generally yield lower PFEs compared to their traditional counterparts.

In conclusion, machine learning models generally exhibit lower bias. Additionally, earnings prediction models tend to systematically overestimate earnings as the forecast horizon increases, regardless of whether traditional linear or machine learning methods are employed

[TABLE 1 ABOUT HERE]

Price scaled absolute forecast error

Turning to the next evaluation metric, Table 2 reports the time-series averages of the median PAFEs for the four traditional, the four machine learning and the two ensemble models. The PAFE is a measure for the accuracy of a model, with values closer to zero indicating higher accuracy.

First, we observe a positive effect of model stacking. Overall, among the traditional models, the best-performing one is the traditional ensemble, while among the machine learning models, the best-performing one is the machine learning ensemble. Specifically, the machine learning ensemble outperforms each of its individual component models for every forecast horizon. The same is true for the traditional ensemble for the horizons $t + 1$ to $t + 2$. Especially in the traditional case, this result is surprising, since the models differ very little in terms of the predictor variables. For forecast horizons $t + 3$ to $t + 5$, the traditional ensemble performs slightly worse than the best

performing traditional component model(s) (0.0396 vs 0.0394 in $t + 3$, 0.0445 vs 0.0438 in $t + 4$ and 0.0505 vs 0.0494/0.0504 in $t + 5$). Second, we find that in most cases, all machine learning models, including the ensemble, outperform all traditional models, including the ensemble, for all prediction horizons. However, for forecast horizons $t + 3$ to $t + 5$, the RI as well as the ENTD model are more accurate than some of the machine learning component models. This stresses the benefit of model averaging for the non-linear machine learning models specifically. The difference in accuracy between the machine learning and the linear ensemble is at a statistically significant level between -0.0020 and -0.0032. This translates into a relative difference of 11.44% for earnings in $t + 1$ to 6.34% for earnings in $t + 5$. The ENML thus provides not only statistically significant, but also economically meaningful gains in accuracy over the traditional models.

[TABLE 2 ABOUT HERE]

Figure 1 illustrates the median PAFE of the ENML and ENTD for each out-of-sample year and for forecast horizons of $t + 1$ and $t + 5$, respectively. The plots illustrate that the overall PAFE levels strongly increase with increasing forecast horizon. Furthermore, they reveal that the machine learning ensemble is more accurate in all years and for both forecast horizons, except for 2009, in which the ENTD yields slightly more accurate $t + 5$ forecasts than the ENML. In general, the difference between the ENML-PAFE and the ENTD-PAFE varies across out-of-sample periods.

[FIGURE 1 ABOUT HERE]

4.2 Out-of-sample R^2

The next metric considered is the out-of-sample R^2 ($OOS R^2$). The results are reported in Table 3. The $OOS R^2$ allows us to assess how the models perform in terms of explaining out-of-sample variation in future earnings.

As expected, the $OOS R^2$ decreases with increasing forecast horizon. Moreover, the results generally confirm the positive effect of model stacking. Our analysis demonstrates that the machine learning ensemble consistently outperforms its individual components across all forecast horizons considered. In contrast, there are instances in which the traditional ensemble exhibits

slightly lower performance than the RI and the EP model for specific forecast horizons throughout forecast horizons $t + 1$ to $t + 3$. Nevertheless, the traditional ensemble consistently outperforms all of its component models for the forecast horizons $t + 4$ to $t + 5$. This outperformance is particularly strong for $t + 5$ forecasts, amounting to 28.30% when comparing it to the second-best performing traditional model.

Furthermore, our findings validate the notion that machine learning approaches surpass traditional linear models in terms of predictive performance. To be more specific, when we compare the $OOS R^2$ of the machine learning models, including the ensemble, against that of the traditional models, including the ensemble, we find that the machine learning models outperform the traditional ones in most cases. In fact, just assessing the best-performing models, i.e., the ensembles, we find that the machine learning ensemble beats its traditional counterpart for every forecast horizon. The difference in $OOS R^2$ between the ensemble models is statistically significant at the 1% level for forecast horizons $t + 1$ to $t + 3$. Further, while the relative PAFE difference between the ensembles decreases with increasing forecast horizon, the difference in $OOS R^2$ increases from 14.91% for $t + 1$ predictions to 23.40% for $t + 5$ predictions, in relative terms.

[TABLE 3 ABOUT HERE]

Figure 2 plots the $OOS R^2$ for the ENML and ENTND for each year and for $t + 1$ and $t + 5$, respectively. Again, it is evident that the machine learning ensemble outperforms the traditional ensemble in the majority of years and for both forecast horizons. Remarkably, the $OOS R^2$ of both ensemble models exhibits a noticeable dip around 2009, particularly pronounced in forecasts for $t + 5$. Notably, for $t + 5$ forecasts, it takes some years for the $OOS R^2$ to rebound. This result underscores the delayed integration of new information, such as the financial crisis in this case, into longer-term forecasts. Such delayed adaptation is inherent in the rolling window approach we employ.

[FIGURE 2 ABOUT HERE]

4.3 Implied cost of capital

A common application of earnings forecasts is the derivation of the implied cost of capital (ICC), for which earnings predictions serve as a crucial input. We restrict this analysis to the best performing traditional earnings forecast model, i.e., the traditional ensemble, as well as the best performing machine learning earnings forecast model, i.e., the machine learning ensemble.

We evaluate the return predictions via ICC as follows. At the end of June in a given year t we estimate the ICC as an average of five commonly used methods.¹⁴ We then sort stocks into quintiles according to the estimated ICC and buy the top quintile.¹⁵ We then evaluate the return from July in t to June in $t + 1$ and repeat the process. This long-only strategy corresponds to a theoretically feasible and easily implementable strategy. In Figure 3 we show how an initial investment of 1\$ would have developed over the out-of-sample periods. We plot the \$-portfolio value based on the ENTD-ICC and the ENML-ICC. As an additional benchmark, we also plot the value of a value-weighted market portfolio for the same period.¹⁶

The plot reveals two key insights. First, (long-only) ICC investment strategies seem to work well in general. Both ICC portfolios outperform the value-weighted benchmark portfolio that we include. Second, the ENML-ICC portfolio outperforms the ENTD-ICC portfolio by a significant margin. More precisely, at the end of the out-of-sample periods, the initial investment of the ENML-ICC portfolio grew to a value of around \$ 6, outperforming the ENTD-ICC portfolio by around \$ 1.2.

[FIGURE 3 ABOUT HERE]

We conclude that the improved accuracy of machine learning predictions translates into more profitable investment strategies, thereby stressing the practical importance of earnings prediction accuracy.

¹⁴An overview over these methods is provided in Appendix C.1.

¹⁵Note that this quintile corresponds to the stocks for which the return expectation is the highest based on the ICC.

¹⁶Value-weighted returns are retrieved from CRSP. More specifically, we annualize the monthly value-weighted market returns including dividends.

5 Interpretation

5.1 Variable importance

As outlined above, we assess the degree to which financial statement variables matter in the machine learning ensemble by computing their SHAP values.¹⁷ More precisely, we compute SHAP values per out-of-sample period and derive their respective averages for each variable.

Figure 4 shows the mean absolute SHAP values of the ENML, averaged over all out-of-sample periods and scaled so that variable importance per forecast horizon sums to one. We show the twenty most important variables for predicting earnings in $t + 1$ and sort them according to their importance.¹⁸ The higher the SHAP value, the more important the variable. For $t + 1$ predictions, *ib*, i.e., current earnings, is the most important variable by far.¹⁹ This comes as no surprise, considering that a simple model including only current earnings as a predictor performs comparably well in predicting future earnings.

A striking finding is that the remaining 19 most important variables are primarily different definitions of earnings. For example, the second most important variable *oiadp* resembles "operating income after depreciation" and the third most important variable *ebit* resembles "earnings before interest and taxes". The only top-twenty variables that do not originate from the income statement are *oancf* and *fopo*.

In general, few variables dominate across forecast horizons $t + 1$ to $t + 5$. Another finding regarding the different forecast horizons is that the significance of *ib* gradually diminishes with increasing horizon. Instead, *oiadp*, another earnings variable, emerges as the most important variable. Moreover, one of the variables not stemming from the income statement, i.e., *oancf*, becomes increasingly important with increasing forecast horizon.

Lastly, the results suggest that current data is more important than lagged data or first order differences. We revisit this claim below.

[FIGURE 4 ABOUT HERE]

¹⁷We focus on the ensemble model as it is the best performing machine learning approach.

¹⁸The variable definitions are provided in Table B.2 in the Appendix.

¹⁹Note that this is the earnings definition that we use as our target variable. Further note that all of our variables are scaled by common shares outstanding. Thus, strictly speaking, we refer to earnings per share when talking about *ib*.

We now explore whether the most important predictor variables act as substitutes or complements. This assessment is conducted by examining the absolute Pearson correlation coefficients of the top-twenty variables reported in Figure 5. If the variables are substitutes, one would expect high coefficient values. In general, we find mixed results. *pi*, *ibcom*, *ibadj*, *ni*, *ibc* and *niadj* are correlated quite strongly with *ib* and each other. All of these are variations of earnings definitions which do not differ strongly from each other. This observation leads us to consider these variables as substitutes for *ib*, indicating that they do not necessarily possess significant independent predictive power. Other variables, that are not as closely related to *ib*, either because they explicitly exclude major income statement items, such as *oiadjp*, or because they are not income statement items at all, such as *oancf*, do not correlate as strongly with earnings. We interpret this as evidence that these items are complements to *ib* and hence possess stand-alone predictive power.

[FIGURE 5 ABOUT HERE]

5.2 Group importance

Table 4 reports variable importance per group. More precisely, we group the variables according to the financial statement they originate from (Panel A), whether they are current, lagged or change information (Panel B), and according to the two aforementioned categories (Panel C). Grouping the variables according to the financial statement they originate from reveals that income statement (I/S) variables are the most important variables. On average, for one-year-ahead predictions, I/S variables contribute approximately 67% to the total importance, while balance sheet (B/S) variables and cash flow statement (CF/S) variables contribute around 19% and 15%, respectively. This finding aligns with the analysis of the most important variables, indicating that I/S variables significantly outweigh others in importance for predicting earnings. However, we also find that I/S variables become less important with increasing forecast horizon. In fact, the importance of I/S variables decreases to around 47% for $t + 5$ forecasts. In contrast, B/S variables become more important with increasing forecast horizon (around 37% for $t + 5$ predictions) while CF/S variables stay at a constant level.

[TABLE 4 ABOUT HERE]

Grouping variables according to whether they are current, lagged or difference variables in Panel B reveals that current data is by far the most important group out of these categories, contributing around 74% of total importance for $t + 1$ predictions. Lagged and difference data each contribute around 13 – 14% to total importance for $t + 1$ predictions. Again, importance becomes more evenly distributed among the groups with increasing forecast horizon. More precisely, current variables become less important while lagged variables become more important. This might be the case because short-term forecasts are heavily influenced by current information due to their sensitivity to recent developments. Longer-term forecasts benefit from a combination of current and lagged information to capture the interplay of short-term dynamics and longer-term trends.

Further breaking down the groups according to the two aforementioned categories stresses the findings above. Overall, current I/S variables contribute around 55% to total importance for $t + 1$ forecasts and hence represent the most important group of variables by a significant margin. This is intuitive and supports the finding that simple earnings forecasts models only considering current earnings items, like the L model or the EP model, perform comparably well in predicting future earnings.

We conclude our variable importance assessment by more thoroughly analyzing the variable importance per financial statement type in Table 5. More precisely, we assess the importance of each financial statement type per schematic financial statement component, such as e.g., current assets, fixed assets or equity, in the B/S case. This analysis provides an intuitive accounting perspective on which components of financial statements are important and how the importance might change across forecast horizons.

[TABLE 5 ABOUT HERE]

The table provides several key insights: first, assessing the B/S, we find that the debt and supplemental items are the most important pieces of B/S information, with both contributing around 5-6% to total importance for $t + 1$ forecasts. Moreover, all pieces of B/S information

consistently increase in importance with increasing forecast horizon.

Second, variables associated with the operating cash flow resemble the most important category of CF/S variables. This comes as no surprise, since the operating cash flow closely relates to earnings. The relevance of the investing cash flow slightly increases with increasing forecast horizon. However, overall, the differences are minor.

Third, different definitions of earnings, i.e., EBITDA, EBIT, EBT and net income, resemble the most important I/S categories. Out of these categories, EBIT and net income are the most important categories with around 18% and 25% share in total importance, respectively. Interestingly, the importance of EBIT only slightly declines with increasing forecast horizon, whereas the importance of the net income consistently strongly declines with increasing forecast horizon to around 12% for $t + 5$ forecasts. This dynamic might be attributable to the fact that net income is more strongly exposed to accounting manipulation than EBIT and hence less reliable in the long-term. Moreover, we find that while sales contribute very little overall, they consistently increase in importance with increasing forecast horizon. This supports the notion that items which are less exposed to discretionary accounting gain predictive value when considering longer forecast horizons. Revisiting the aforementioned finding that operating cash flow variables maintain consistent importance across forecast horizons further reinforces this notion. Unlike earnings, cash flows include no discretionary accrual items and are hence not exposed to earnings management (e.g., Jones, 1991). Consequently, their predictive value does not decrease for longer-term forecasts. In summary, these findings suggest that the variations in importance across forecast horizons are primarily driven by the presence of earnings management. Future research endeavors could offer additional insights into these dynamics.

5.3 Non-linearity

We now approximate the degree to which non-linearity of the functional form and non-linearity of variables, i.e., interactions among financial statement variables considered, play a role in predicting earnings.

Surrogate model

We find that for our flexible ENML approach, around 94% of the variation in predicted earnings can be explained by the linear surrogate model for $t + 1$ predictions on average. This indicates that the ENML predictions can still be approximated quite accurately by a linear model. More so, it indicates that interaction effects as well as non-linearity of the functional form do not play a major role, on average. In fact, including two-way interactions increases the average adjusted R^2 by around 2 percentage points, on average. The remaining unexplained portion of variance of the ENML predictions is attributable to the non-linear functional form of the ENML.²⁰ We depict this graphically in Figure 6. The figure shows the adjusted in-sample R^2 per out-of-sample period, derived by regressing predicted ENML earnings on a linear surrogate model and a linear surrogate model including two-way interactions. The figure also includes the surrogate models for predictions for $t + 5$. With increasing horizon, a slightly larger portion of the earnings-predictor relation can be attributed to interaction effects and non-linearity of the functional form. Nonetheless, the linear surrogate model still explains around 90% of the $t + 5$ predictions on average.

[FIGURE 6 ABOUT HERE]

Partial dependence plots

We now turn to how the aforementioned degree of non-linearity is expressed at the variable-level. Figure 7 shows the partial dependence plots for *ib*, *oiadp*, *ebit* and *oancf* for all forecast horizons.²¹ The partial dependence measures the sensitivity of the predicted earnings to the individual financial statement variables.

The upper-left panel shows the effect of *ib* on the model output. Remarkably, the sensitivity appears to be linear for both positive and negative values of *ib*. However, there is a distinction in the slope of the line for positive and negative values, suggesting varying sensitivities of future earnings to current earnings for profit and loss firms, respectively. This may explain why the EP model and the RI model by Li and Mohanram (2014) yield comparably good forecasting results,

²⁰Theoretically, it can also be attributed to higher-order interactions and variables which we did not include. However, in undocumented results, we find that this is not the case.

²¹We plot the partial dependence of the three most important variables and the most important variable not stemming from the income statement.

especially for short forecast horizons. The two models include a dummy for negative earnings and the interaction between earnings and the negative earnings dummy, which essentially allows for different slopes of *ib* for profit and loss firms.

For *ebit* in the lower-left panel we find a similar trend as for *ib*. For *oiadp*, there is also a difference in slopes between profit and loss firms, especially for longer forecast horizons. However, the kink appears slightly below zero. Also, interestingly, for $t + 1$ forecasts, the change relationship appears to be concave. In contrast *oancf* is essentially linearly related to future earnings across all forecast horizons.

[FIGURE 7 ABOUT HERE]

5.4 Subsamples

We now turn to the subsample analysis. More precisely, we assess the ENML predictions per life cycle stage, firm size tercile and industry. This analysis allows an assessment of where exactly overall accuracy improvements stem from and whether the relation between current fundamental data and future earnings differs for different subsets of firms. For the sake of comprehensiveness, we are restricting this analysis to the two ensemble models and focus on the key evaluation and interpretation metrics, i.e., the median PAFE, variable importance as well as the surrogate models.

Life cycle

First, similar to Easton et al. (2024), we stratify predictions according to the life cycle stage that a firm is in as of the estimation date. Importantly, life cycle classifications are only used for evaluation, not for estimation – meaning that the ML models do not directly "see" which life cycle stage a firm belongs to. Despite this, the ML models produce much better predictions for firms at the beginning and at the end of their life cycle, suggesting that they implicitly detect patterns in the underlying financial statement data that distinguish these firms.

Assessing accuracy differences using the median PAFE reveals interesting patterns. First, ENML outperforms ENTND in every life cycle stage and for every forecast horizon. Second, firms at the beginning of their life cycle (labelled "Intro") and at the end (labelled "Decline") exhibit the

lowest overall accuracy by far. However, these are also the stages where ML provides the largest improvement: for $t + 1$ forecasts, ENML reduces median PAFE by 0.0110 and 0.0086 for firms in their "Intro" and "Decline" stages, respectively – while improvements for "Growth," "Mature," and "Shake-out" firms range from 0.0022 to 0.0034. These results are visualized in Figure 8.²²

[Figure 8 ABOUT HERE]

Turning to variable importance, we find no significant differences across life cycle stages. As summarized in Table 6, the importance of financial statement information in forecasting future earnings remains consistent across life cycle stages.²³ More precisely, for firms in every life cycle stage, income statement data is the most important group of predictors for $t + 1$ forecasts, while balance sheet data becomes more important for longer horizons.

[TABLE 6 ABOUT HERE]

Lastly, we examine the degree to which interaction effects and non-linearities play a role for different life cycle stages. Table 7 reports the results. In general, we find a decrease in the linear surrogate model's R^2 with increasing forecast horizon, indicating that non-linearities and interactions become more important over time. This decrease is observed for all life cycle stages. Nevertheless, the effects differ strongly across life cycle stages. For "Intro" and "Decline" firms, a simple linear surrogate model without interactions explains only 86-87% of ENML predictions, whereas for "Growth," "Mature," and "Shake-out" firms, it explains 94-95%. However, once we introduce interaction terms, the surrogate model explains 97-99% of the ENML predictions for all life cycle stages. This suggests that interaction effects are particularly important for firms at the beginning and end of their life cycle, which aligns with the stronger accuracy improvements observed for these firms. Importantly, these effects are uncovered by the ENML despite the fact that life cycle stages are not provided as an input. Instead, it autonomously captures complex patterns in financial statement data that distinguish these firms and enhance earnings predictability.

²²The results for all forecast horizons are summarized in Table D.3 in the Appendix.

²³In untabulated results, we find that this is also true for importance on the variable as well as the financial statement component level.

[TABLE 7 ABOUT HERE]

Firm size

Next, we stratify predictions according to the size of the firm as of estimation date. In line with the literature, we find that the larger the firms, the higher the prediction accuracy (Li and Mohanram, 2014). Nonetheless, the ENML outperforms the ENT D for each size subsample considered. Further, we find that the accuracy improvements increase with decreasing firm size. More precisely, for $t + 1$ forecasts, the ENML yields a median PAFE improvement of 0.0073 for small firms and of 0.0030-0.0033 for medium and large firms. We visualize these findings in Figure 9.²⁴

[FIGURE 9 ABOUT HERE]

In terms of variable importance, we find no significant differences between the findings regarding the full sample and the sample stratified by size. We summarize the results with respect to the financial statements in Table 8.²⁵

[TABLE 8 ABOUT HERE]

Turning to the surrogate models, we again find a decrease in linear surrogate R^2 , i.e., an increasing influence of non-linearities and interactions in the ENML, with increasing forecast horizon. Moreover, as in the life cycle analysis, the surrogate R^2 s match the differences in accuracy improvements across subsamples. The ENML achieves the highest accuracy improvements for the small firm subsample and it is this firm subsample for which adding interactions leads to the largest gains in surrogate R^2 . Again, the effect of nonlinearities, which we infer from the portion of the variance in ENML predictions which is not explained by the surrogate model including interactions, is constant across subsamples. The results are summarized in Table 9.

²⁴The results for all forecast horizons are summarized in Table D.4 in the Appendix.

²⁵In untabulated results, we find that this is also true for importance on the variable as well as the financial statement component level.

[TABLE 9 ABOUT HERE]

Industry

Lastly, we stratify the predictions according to the industry the respective firm is in as of the estimation date. We find that the ENML outperforms the ENTD for nearly all forecast horizons and industries. In fact, the ENML improve the median PAFE by 0.0032-0.0050, depending on the industry. The aforementioned range highlights the substantial variations in accuracy enhancements across different industries. The results are visualized in Figure 10.²⁶

[FIGURE 10 ABOUT HERE]

Turning to variable importance, we again find that stratification of the sample does not lead to results that differ from that for the full sample. We report the importance per financial statement in Table 10.²⁷

[TABLE 10 ABOUT HERE]

Interestingly, the surrogate R^2 s do not match the results regarding accuracy as in the two cases above. The surrogate modeling results are reported in Table 11. For example, the ENML achieves the highest accuracy gains over the ENTD for firms within industry "4". Yet, the degree to which interaction terms and non-linearities play a role for the ENML predictions are not the highest for this industry, as indicated by the surrogate R^2 s. We conclude that when stratifying by industries, it is primarily the additional variables which we feed to the ENML which lead to differences in predictive performance. Again, the effect of the non-linear functional form appears to be modest and slightly lower than the effect of interactions, in general.

[TABLE 11 ABOUT HERE]

²⁶The results for all forecast horizons are summarized in Table D.5 in the Appendix.

²⁷In untabulated results, we find that this is also true for importance on the variable as well as the financial statement component level.

6 Conclusion

We show that earnings per share predictions based on state-of-the-art machine learning approaches using high-dimensional financial statement data are more accurate than those based on traditional linear approaches. These improvements hold across all evaluation metrics assessed, i.e., commonly used error metrics, the $OOS R^2$ as well as the performance of long-short ICC portfolios based on the predictions.

Importantly, we provide an intuitive breakdown of how important the different pieces of fundamental accounting information are for predicting earnings. We find that current I/S variables, especially current earnings, are the most important predictors. However, with increasing forecast horizon, variable importance becomes more balanced. More precisely, B/S information becomes much more important whereas I/S information becomes less important with increasing forecast horizon. Thoroughly disentangling the different financial statements suggests that this dynamic may be attributable to earnings management.

As the first study to thoroughly decompose the effects of non-linearity in the earnings prediction context we find that especially for short term-horizons, the relationship as approximated by the best performing machine learning model, i.e., the machine learning ensemble, can still be described by a linear surrogate model to a large extent. More precisely, we find that on average around 94% of the variance in machine learning ensemble predictions can be explained by a linear model, depending on the forecast horizon. Interactions and non-linearity of the functional form somewhat equally contribute to the remaining small unexplained portion of predictions. As the forecast horizon increases, the linear surrogate R^2 decreases slightly.

We conclude our empirical analysis by stratifying firms into different subsets and assessing differences in accuracy, variable importance and non-linearity. Overall, the ENML beats the ENT D for every subset of firms considered. However, significant differences in terms of accuracy and improvements of accuracy of the ENML as compared to the ENT D exist across firm subsets. Moreover, the degree to which the ENML exploits interactions across fundamental variables appears to vary across life cycle stages and firm size terciles.

Our findings provide important guidance for future research. First, we show that machine learning approaches are an excellent tool for earnings predictions. We hence argue that research

which uses (model) earnings predictions in some way or another should resort to machine learning methods, if high accuracy is desired. Second, we show which financial statement variables and groups thereof are important. Future research may build upon that when building models and deciding which variables to include. Importantly, this includes the differences in terms of variable importance across forecast horizons. For example, if one is interested in an earnings prediction model including only a small number of variables for computation-related reasons, employing distinct (small) sets of variables for different forecast horizons might be beneficial. Lastly, we show that model-based forecast performance as well as the degree to which complex forecast models are necessary differs between subsets of firms. Additional research that further aims to disentangle these relations and their causes is an interesting avenue for future research and could ultimately lead to further forecast model improvements.

References

- Albrecht, W. S., L. L. Lookabill, and J. C. McKeown (1977). The time-series properties of annual earnings. *Journal of Accounting Research* 15(2), 226–244.
- Bali, T. G., H. Beckmeyer, M. Moerke, and F. Weigert (2023). Option return predictability with machine learning and big data. *Working Paper*.
- Ball, R. and P. Brown (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6(2), 159–178.
- Ball, R. and R. Watts (1972). Some time series properties of accounting income. *The Journal of Finance* 27(3), 663–681.
- Brown, L. D. (1993). Earnings forecasting research: its implications for capital markets research. *International Journal of Forecasting* 9(3), 295–320.
- Campbell, J., H. Ham, Z. G. Lu, and K. Wood (2023). Expectations matter: When (not) to use machine learning earnings forecasts. *Working Paper*.
- Cao, K. and H. You (2021). Fundamental analysis via machine learning. *Working Paper*.
- Chen, X., Y. H. T. Cho, Y. Dou, and B. Lev (2022). Predicting future earnings changes using machine learning and detailed financial data. *Journal of Accounting Research* 60(2), 467–515.
- Claus, J. and J. Thomas (2001). Equity premia as low as three percent? evidence from analysts' earnings forecasts for domestic and international stock markets. *The Journal of Finance* 56(5), 1629–1666.
- Dickinson, V. (2011, 11). Cash flow patterns as a proxy for firm life cycle. *The Accounting Review* 86(6), 1969–1994.
- Easton, P. D. (2004). Pe ratios, peg ratios, and estimating the implied expected rate of return on equity capital. *The Accounting Review* 79(1), 73–95.
- Easton, P. D., M. M. Kapons, S. J. Monahan, H. H. Schütt, and E. H. Weisbrod (2024). Forecasting earnings using k-nearest neighbors. *The Accounting Review* 99(3), 115–140.

- Gebhardt, W. R., C. M. C. Lee, and B. Swaminathan (2001). Toward an implied cost of capital. *Journal of Accounting Research* 39(1), 135–176.
- Gerakos, J. J. and R. B. Gramacy (2012). Regression-based earnings forecasts. *Chicago Booth Research Paper No. 12-26*.
- Gordon, J. R. and M. J. Gordon (1997). The finite horizon expected return model. *Financial Analysts Journal* 53(3), 52–61.
- Graham, J. R., C. R. Harvey, and S. Rajgopal (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40(1), 3–73.
- Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies* 33(5), 2223–2273.
- Hansen, J. W. and C. Thimsen (2020). Forecasting corporate earnings with machine learning. *Working Paper*.
- Harris, R. D. F. and P. Wang (2019). Model-based earnings forecasts vs. financial analysts' earnings forecasts. *British Accounting Review* 51(4), 424–437.
- Hendriock, M. (2022). Forecasting earnings with predicted, conditional probability distribution density functions. *Working Paper*.
- Hou, K., M. A. van Dijk, and Y. Zhang (2012). The implied cost of capital: A new approach. *Journal of Accounting and Economics* 53(3), 504–526.
- Israel, R., B. Kelly, and T. Moskowitz (2020). Can machines "learn" finance? *Journal of Investment Management* 18(2).
- Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of Accounting Research* 29(2), 193–228.
- Jones, S., W. J. Moser, and M. M. Wieland (2023). Machine learning and the prediction of changes in profitability. *Contemporary Accounting Research* n/a(n/a).
- Kelly, B., S. Malamud, and K. Zhou (2022). The virtue of complexity in return prediction. *Working Paper*.

- Li, K. K. and P. Mohanram (2014). Evaluating cross-sectional forecasting models for implied cost of capital. *Review of Accounting Studies* 13(3), 1152–1185.
- Liaw, R., E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica (2018). Tune: A research platform for distributed model selection and training. *Working Paper*.
- Lundberg, S. M. and S.-I. Lee (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems* 30, pp. 4765–4774.
- Monahan, S. J. (2018). Financial statement analysis and earnings forecasting. *Foundation and Trends in Accounting* 12, 105–215.
- Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3).
- Ohlson, J. A. and B. E. Juettner-Nauroth (2005). Expected eps and eps growth as determinants of value. *Review of Accounting Studies* 10, 349–365.
- Schipper, K. (1991). Analysts' forecasts. *Accounting Horizons* 5(4), 105–121.
- Van Binsbergen, J. H., X. Han, and A. Lopez-Lira (2023). Man vs. machine learning: The term structure of earnings expectations and conditional biases. *Review of Financial Studies* 36(6), 2361–2396.
- Watts, R. L. and R. W. Leftwich (1977). The time series of annual accounting earnings. *Journal of Accounting Research* 15(2), 253–271.

Figures

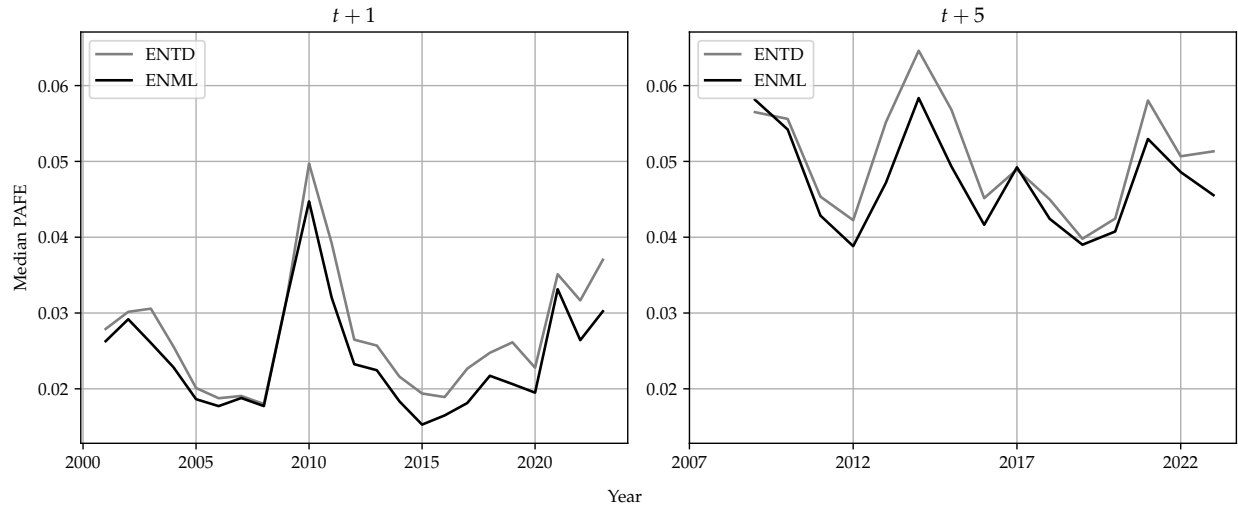


Figure 1: PAFE across out-of-sample periods

This figure shows the median price scaled absolute forecast errors (PAFEs) of the machine-learning ensemble (ENML) and the traditional ensemble (ENT) per out-of-sample period for forecast horizons $t + 1$ and $t + 5$.

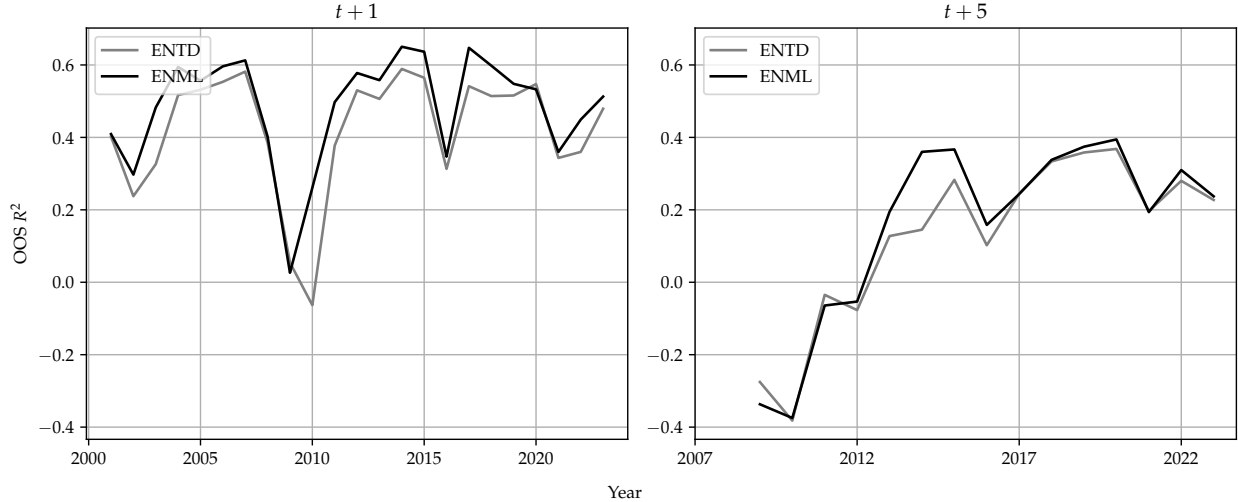


Figure 2: R^2 across out-of-sample periods

This figure shows the OOS R^2 of the machine-learning ensemble (ENML) and the traditional ensemble (ENT) per out-of-sample period for forecast horizons $t + 1$ and $t + 5$.

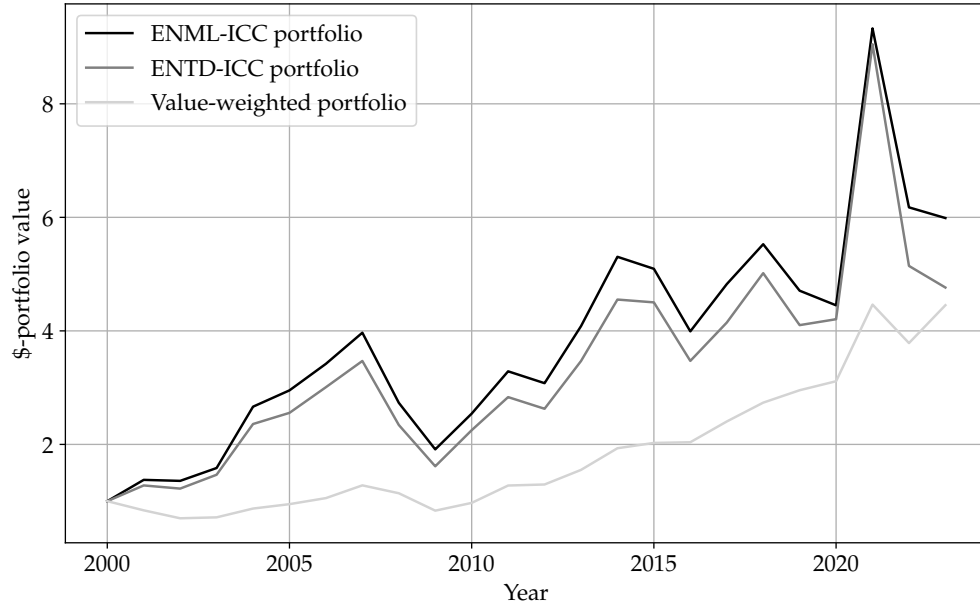


Figure 3: Portfolio evaluation

This figure shows the development of an initial investment of 1\$ investment across out of sample periods. We assess three strategies. First, we invest based on implied cost of capital (ICC). At the end of June of a respective year t , we sort stocks based on their ICC and buy the top quintile. We then collect the returns from July in t to June in $t + 1$ and repeat the process. We do this based on ICC derived via ENML forecasts (ENML-ICC) and based on ICC derived via ENT-D forecasts (ENT-D-ICC). Second, we invest into the value weighted market portfolio. Monthly value-weighted market portfolio returns are directly derived via CRSP.

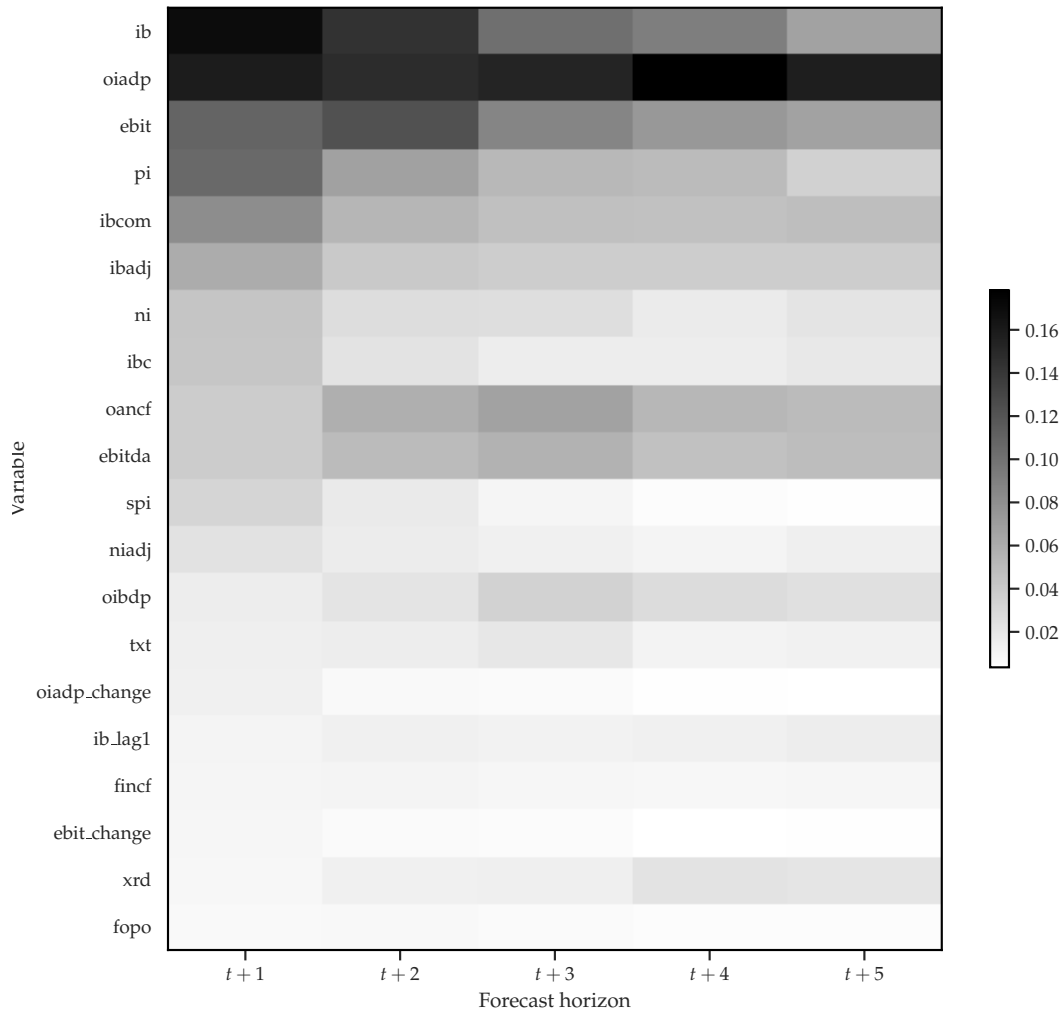


Figure 4: Variable importance for the machine learning forecast ensemble

This figure depicts the absolute SHAP values of the 20 most important variables for the machine learning ensemble, averaged over out-of-sample periods and scaled so they sum up to one within each forecast horizon. In this context, importance is defined as the ranking of the respective variable according to the aforementioned metric for forecast horizon $t + 1$.

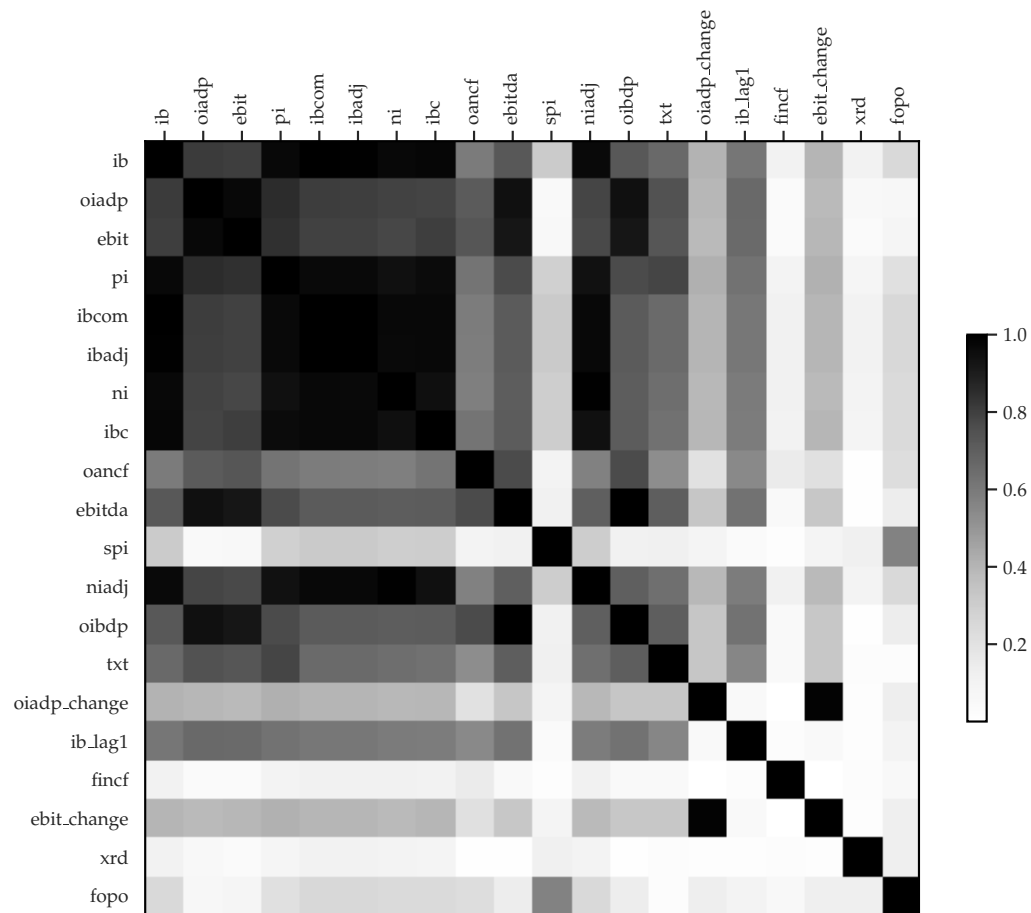


Figure 5: Correlation heatmap for the most important variables

This figure shows the absolute Pearson correlation coefficients for the 20 most important variables for the machine learning ensemble.

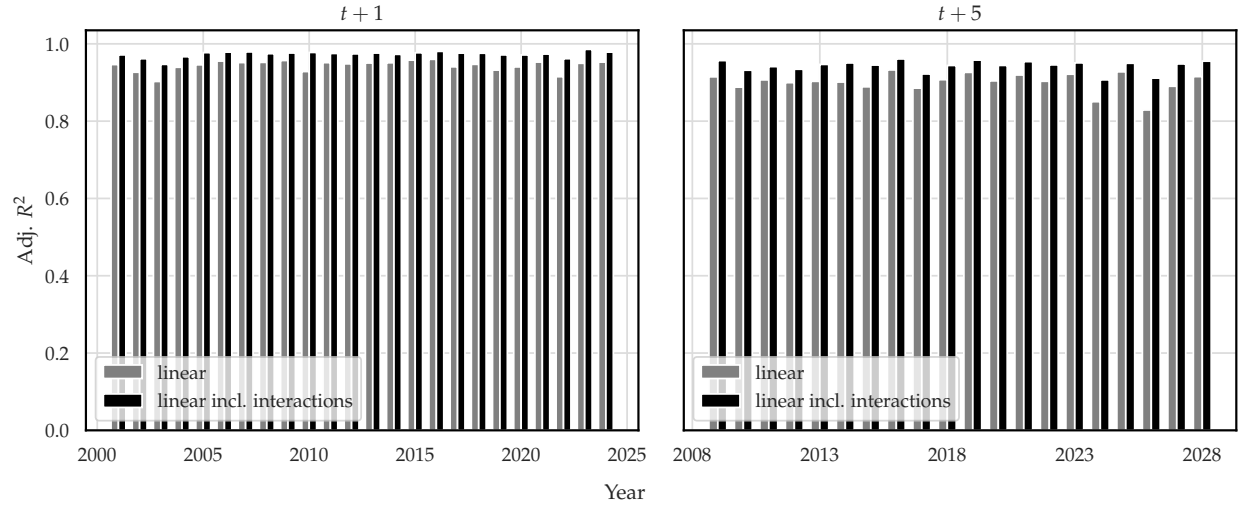


Figure 6: Surrogate models

This figure shows the adjusted R^2 of the surrogate models that we fit to our machine learning ensemble predictions for forecast horizons $t + 1$ and $t + 5$. The linear model (linear) is a simple linear model in which we regress the respective predictions on the 20 most important predictor variables according to their average absolute SHAP values for $t + 1$ forecasts. The linear model including interactions (linear incl. interactions) is a linear model in which we use the same set of predictors as well as all possible two-way interactions.

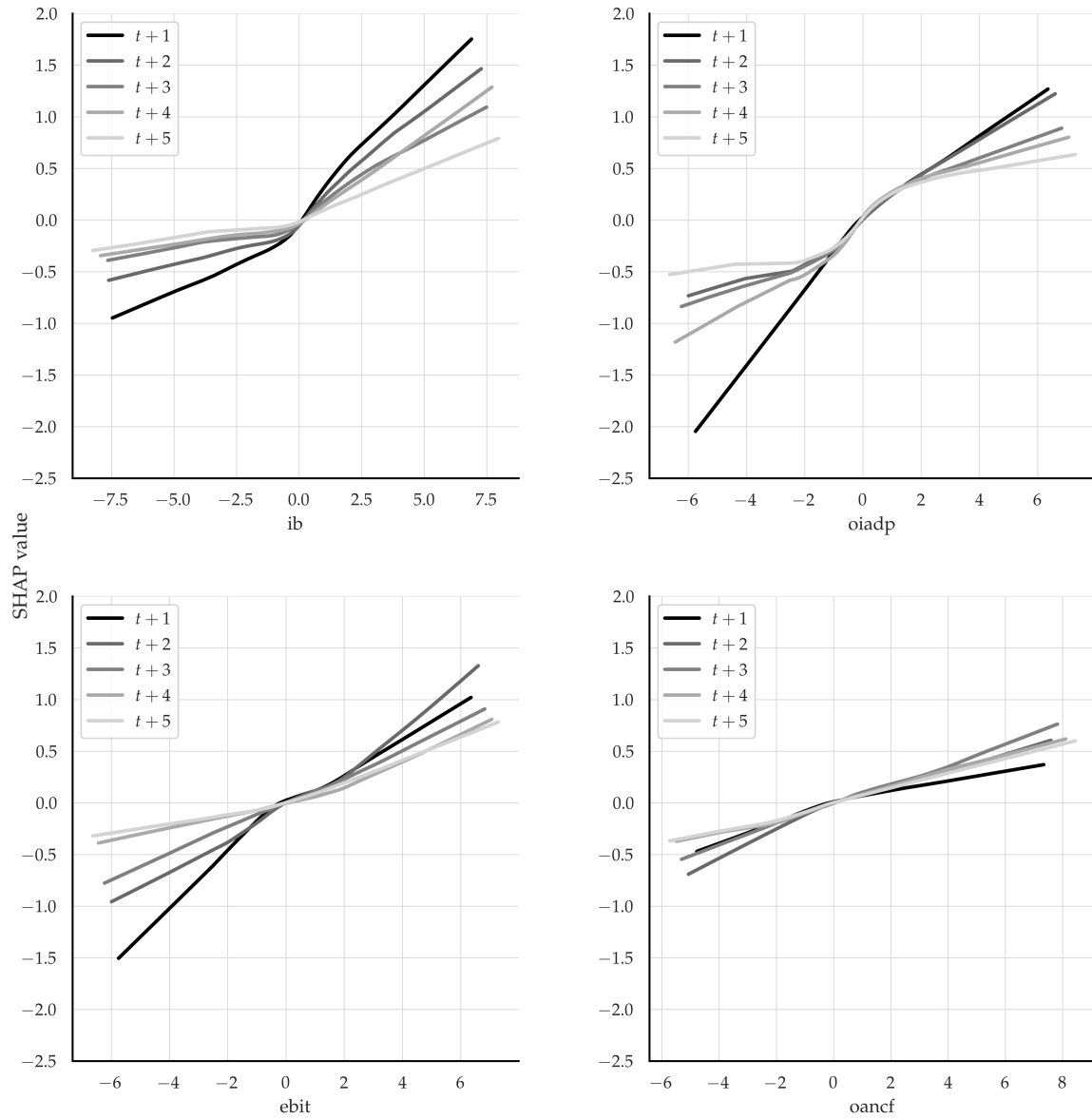


Figure 7: Partial dependence plots

The panels show the sensitivity of expected future earnings to the respective variable for all forecast horizons. More specifically, we fit a nonparametric lowess model (locally weighted linear regression) to the SHAP values of the respective variable.

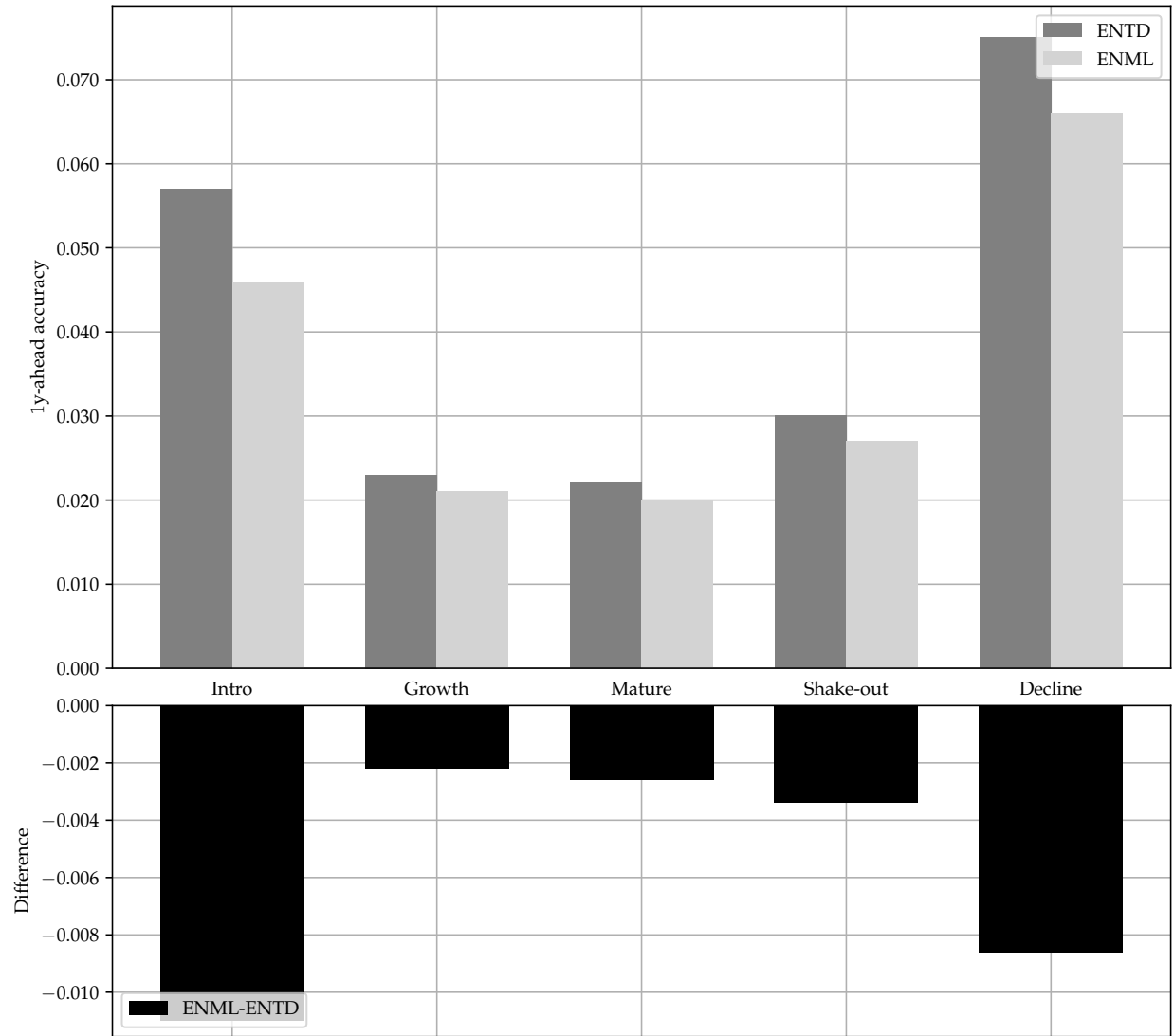


Figure 8: Accuracy differences per life cycle stage

This figure reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) per life cycle stage for the two ensemble models and for $t + 1$ forecasts. Life cycle stages are defined following Dickinson (2011) and as of estimation date. An overview is given in Table D.1 in the Appendix. ENTD denotes the traditional ensemble and ENML denotes the machine learning ensemble. The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year.

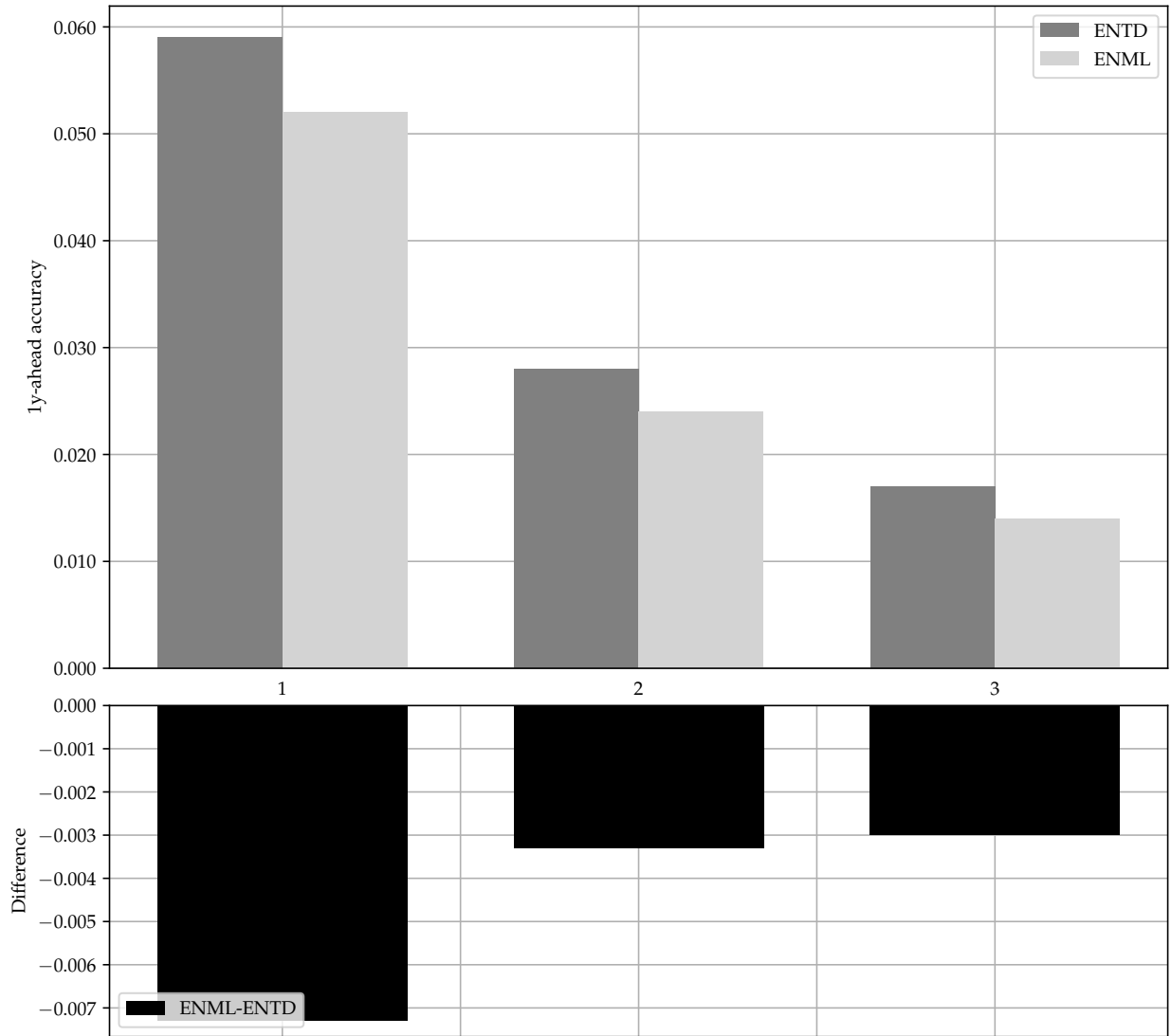


Figure 9: Accuracy differences per size tercile

This figure reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) per size tercile for the two ensemble models and for $t + 1$ forecasts. Size is defined as the market capitalization as of estimation date. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. ENT D denotes the traditional ensemble and ENML denotes the machine learning ensemble. The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year.

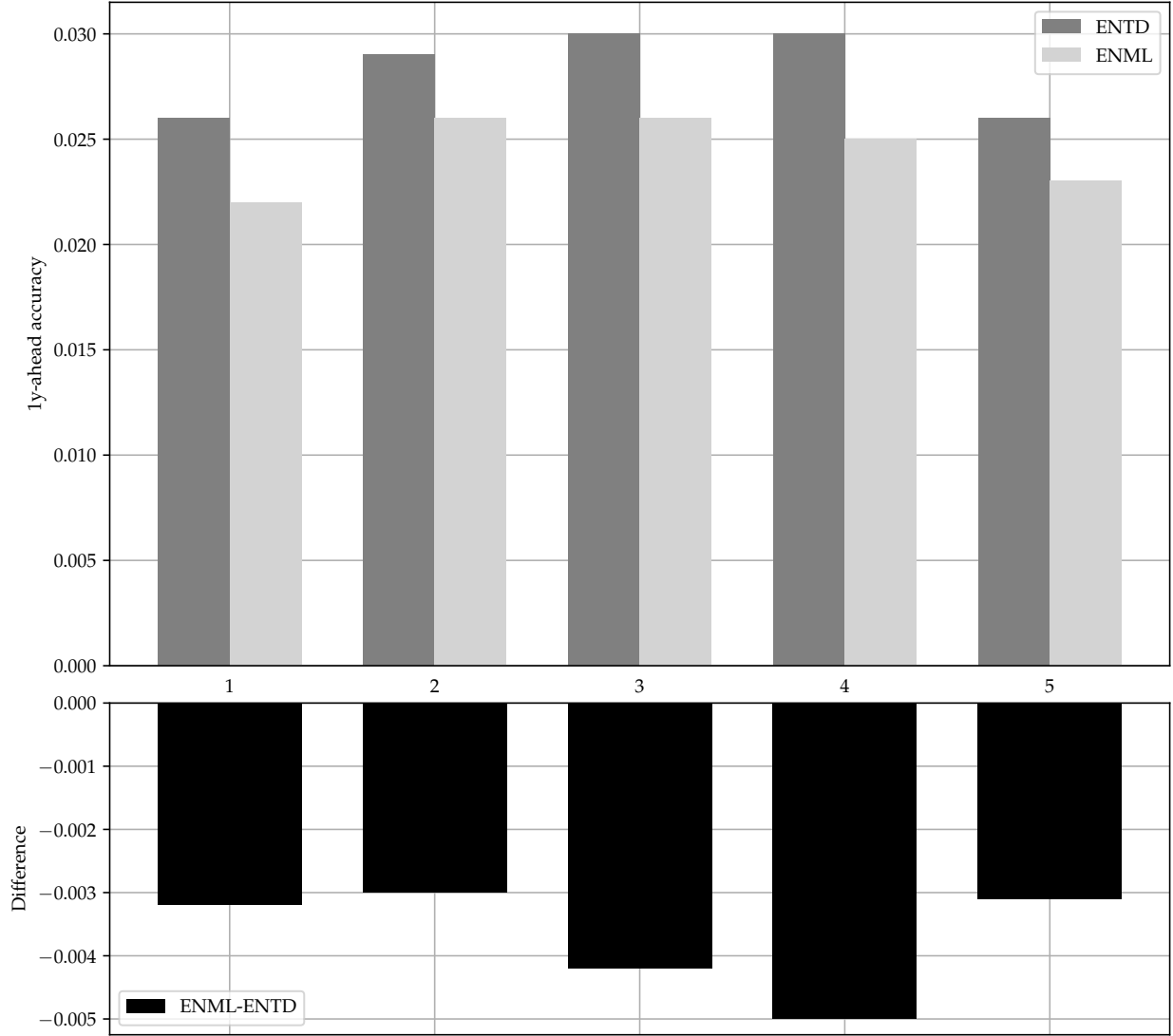


Figure 10: Accuracy differences per industry

This figure reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) per industry for the two ensemble models and for $t + 1$ forecasts. Industries are defined following the Fama-French 5 industry portfolios classification, available on Kenneth French's website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html), and as of estimation date. An overview is given in D.2 in the Appendix. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. ENTND denotes the traditional ensemble and ENML denotes the machine learning ensemble. The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year.

Tables

Table 1: Median PFE

	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
L^{MSE}	0.0007	-0.0001	-0.0027	-0.0083	-0.0146**
HVZ^{MSE}	0.0039***	0.0033	0.0006	-0.0054	-0.0129***
EP^{MSE}	0.0056***	0.0055**	0.0033	-0.0027	-0.0091*
RI^{MSE}	0.0063***	0.0059**	0.0028	-0.0032	-0.0103**
ENTD	0.0043***	0.0035	0.0008	-0.0055	-0.0125**
RF^{MSE}	0.0059***	0.0062***	0.0043*	0.0001	-0.0052
GBT^{MAE}	-0.0011	-0.0019	-0.0019	-0.0040	-0.0083*
$DART^{MAE}$	-0.0003	-0.0011	-0.0020	-0.0038	-0.0061
NN^{MAE}	-0.0006	-0.0010	-0.0019	-0.0043	-0.0081*
ENML	0.0014	0.0009	0.0001	-0.0029	-0.0070

This table reports the time-series averages of the median price scaled forecasting errors (PFEs) for all models. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. L is a model with only current earnings as a predictor, HVZ is the model by Hou et al. (2012), EP and RI are the models by Li and Mohanram (2014), ENTD is an equally weighted ensemble of L, HVZ, EP, and RI, RF is a random forest model, GBT and DART are gradient boosted tree models without and with dropout, NN is a neural net, and ENML is an equally weighted ensemble of RF, GBT, DART, and NN. The superscript MAE (MSE) indicates that the respective model is estimated using the mean absolute error (mean squared error) as its loss function. We follow the literature and estimate the traditional models using the MSE. In untabulated results we find that our results are robust to estimating the traditional models using the MAE. We decide on which loss function to report for the ML models depending on which one yields more accurate predictions (as indicated by the price scaled absolute forecast error (PAFE) for the $t + 1$ horizon). The PFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year. ***, **, and * denote statistical significance at the 1%, the 5% and the 10% level, respectively. Standard errors used for deriving statistical significance are adjusted following Newey and West (1987) assuming a lag length of three years.

Table 2: Median PAFE

	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
L^{MSE}	0.0297***	0.0380***	0.0439***	0.0499***	0.0568***
HVZ^{MSE}	0.0282***	0.0360***	0.0416***	0.0473***	0.0534***
EP^{MSE}	0.0280***	0.0356***	0.0407***	0.0451***	0.0504***
RI^{MSE}	0.0278***	0.0350***	0.0394***	0.0438***	0.0494***
ENTD	0.0271***	0.0345***	0.0396***	0.0445***	0.0505***
RF^{MSE}	0.0268***	0.0344***	0.0398***	0.0452***	0.0494***
GBT^{MAE}	0.0266***	0.0341***	0.0394***	0.0449***	0.0499***
$DART^{MAE}$	0.0243***	0.0325***	0.0383***	0.0444***	0.0489***
NN^{MAE}	0.0247***	0.0339***	0.0380***	0.0444***	0.0481***
ENML	0.0240***	0.0317***	0.0372***	0.0425***	0.0473***
ENML-ENTD	-0.0031***	-0.0028***	-0.0024***	-0.0020***	-0.0032***

This table reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) for all models. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. L is a model with only current earnings as a predictor, HVZ is the model by Hou et al. (2012), EP and RI are the models by Li and Mohanram (2014), ENTD is an equally weighted ensemble of L, HVZ, EP, and RI, RF is a random forest model, GBT and DART are gradient boosted tree models without and with dropout, NN is a neural net, and ENML is an equally weighted ensemble of RF, GBT, DART, and NN. The superscript *MAE* (*MSE*) indicates that the respective model is estimated using the mean absolute error (mean squared error) as its loss function. We follow the literature and estimate the traditional models using the MSE. In untabulated results we find that our results are robust to estimating the traditional models using the MAE. We decide on which loss function to report for the ML models depending on which one yields more accurate predictions (as indicated by the price scaled absolute forecast error (PAFE) for the $t + 1$ horizon). The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year. ***, **, and * denote statistical significance at the 1%, the 5% and the 10% level, respectively. Standard errors used for deriving statistical significance are adjusted following Newey and West (1987) assuming a lag length of three years.

Table 3: Average out-of-sample R^2

	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
L^{MSE}	0.3631	0.1932	0.1417	0.0856	0.0262
HVZ^{MSE}	0.4198	0.2875	0.2315	0.1596	0.0901
EP^{MSE}	0.4232	0.2843	0.2285	0.1554	0.0934
RI^{MSE}	0.4326	0.3022	0.2501	0.1713	0.0986
ENTD	0.4220	0.2891	0.2440	0.1807	0.1265
RF^{MSE}	0.4659	0.3147	0.2582	0.1792	0.1293
GBT^{MAE}	0.4417	0.2985	0.2458	0.1673	0.1146
$DART^{MAE}$	0.4760	0.3237	0.2486	0.1653	0.1154
NN^{MAE}	0.4625	0.2857	0.2527	0.1560	0.1222
ENML	0.4849	0.3447	0.2872	0.2100	0.1561
ENML-ENTD	0.0629***	0.0556**	0.0433**	0.0293	0.0296

This table reports the time-series averages of the out-of-sample R^2 s (OOS R^2 s) for all models. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. L is a model with only current earnings as a predictor, HVZ is the model by Hou et al. (2012), EP and RI are the models by Li and Mohanram (2014), ENTD is an equally weighted ensemble of L, HVZ, EP, and RI, RF is a random forest model, GBT and DART are gradient boosted tree models without and with dropout, NN is a neural net, and ENML is an equally weighted ensemble of RF, GBT, DART, and NN. The superscript MAE (MSE) indicates that the respective model is estimated using the mean absolute error (mean squared error) as its loss function. We follow the literature and estimate the traditional models using the MSE. In untabulated results we find that our results are robust to estimating the traditional models using the MAE. We decide on which loss function to report for the ML models depending on which one yields more accurate predictions (as indicated by the price scaled absolute forecast error (PAFE) for the $t + 1$ horizon). ***, **, and * denote statistical significance at the 10%, the 5% and the 1% level, respectively. Standard errors used for deriving statistical significance are adjusted following Newey and West (1987) assuming a lag length of three years. We only test for statistical significance of the difference between the ensemble models (ENML - ENTD).

Table 4: Variable importance by groups

Panel A: Financial statement type					
	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
B/S	0.1928	0.2599	0.3128	0.3361	0.3719
CF/S	0.1503	0.1609	0.1641	0.1529	0.1543
I/S	0.6569	0.5792	0.5231	0.5110	0.4738

Panel B: Variable type					
	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Current	0.7352	0.6875	0.6506	0.6492	0.6207
Lagged	0.1252	0.1734	0.2080	0.2167	0.2492
Change	0.1396	0.1391	0.1413	0.1341	0.1301

Panel C: Financial statement type \times variable type					
	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
B/S current	0.0874	0.1255	0.1563	0.1727	0.1971
B/S lagged	0.0434	0.0677	0.0878	0.0963	0.1069
B/S change	0.0621	0.0668	0.0687	0.0671	0.0679
CF/S current	0.0954	0.0974	0.0924	0.0829	0.0774
CF/S lagged	0.0280	0.0360	0.0430	0.0386	0.0491
CF/S change	0.0269	0.0275	0.0287	0.0314	0.0277
I/S current	0.5525	0.4646	0.4020	0.3936	0.3462
I/S lagged	0.0537	0.0697	0.0772	0.0818	0.0933
I/S change	0.0506	0.0448	0.0439	0.0356	0.0344

Panel A reports the relative variable importance per financial statement group. B/S, CF/S and I/S denote balance sheet, cash flow statement and income statement, respectively. The variables are grouped according to Table B.2 in the Appendix. Panel B reports the relative variable importance per variable type. Panel C reports the relative variable importance per financial statement type \times variable type group. Importance per group in each Panel is defined as the fraction that the respective group contributes to total importance, measured as the sum of absolute SHAP values.

Table 5: Variable importance: financial statements

Panel A: Balance sheet					
	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Current assets	0.0283	0.0410	0.0465	0.0495	0.0595
Fixed assets	0.0199	0.0213	0.0270	0.0287	0.0294
Total assets	0.0037	0.0090	0.0123	0.0159	0.0196
Debt	0.0467	0.0635	0.0743	0.0785	0.0929
Equity	0.0336	0.0438	0.0532	0.0596	0.0605
Total debt & equity	0.0029	0.0052	0.0087	0.0101	0.0122
Supplemental	0.0577	0.0761	0.0908	0.0939	0.0979
Sum B/S importance	0.1928	0.2599	0.3128	0.3361	0.3719
Panel B: Cash flow statement					
	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Operating cash flow	0.1080	0.1081	0.1084	0.0949	0.0976
Investing cash flow	0.0134	0.0184	0.0208	0.0217	0.0211
Financing cash flow	0.0247	0.0288	0.0292	0.0311	0.0302
Total cash flow	0.0042	0.0057	0.0056	0.0052	0.0054
Sum CF/S importance	0.1503	0.1609	0.1641	0.1529	0.1543
Panel C: Income statement					
	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Sale	0.0072	0.0129	0.0169	0.0211	0.0219
Operating expenses	0.0217	0.0310	0.0355	0.0482	0.0470
EBITDA	0.0380	0.0440	0.0517	0.0432	0.0420
Depr. & Amort.	0.0071	0.0076	0.0072	0.0085	0.0108
EBIT	0.1766	0.1625	0.1397	0.1426	0.1240
Interest expenses	0.0542	0.0462	0.0406	0.0328	0.0325
EBT	0.0656	0.0416	0.0311	0.0299	0.0208
Tax expenses	0.0318	0.0395	0.0438	0.0406	0.0455
Net Income	0.2460	0.1818	0.1430	0.1313	0.1174
Dividends	0.0087	0.0120	0.0136	0.0128	0.0118
Sum I/S importance	0.6569	0.5792	0.5231	0.5110	0.4738

This table reports the relative variable importance per financial statement group. B/S, CF/S and I/S denote balance sheet, cash flow statement and income statement. EBITDA denotes earnings before interest, taxes and depreciation and amortization. EBIT denotes earnings before interest and taxes. EBT denotes earnings before taxes. The variables are grouped according to Table B.2 in the Appendix. Importance per financial statement component is defined as the fraction that the respective component contributes to total importance, measured as the sum of absolute SHAP values.

Table 6: Variable importance by financial statement type per life cycle stage

Firm size	Financial statement	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Intro	ENTD	0.2199	0.2810	0.3330	0.3742	0.4062
	ENML	0.1578	0.1657	0.1538	0.1479	0.1418
	ENML - ENTD	0.6222	0.5532	0.5133	0.4779	0.4520
Growth	ENTD	0.1975	0.2558	0.3107	0.3384	0.3775
	ENML	0.1554	0.1617	0.1559	0.1524	0.1596
	ENML - ENTD	0.6471	0.5825	0.5334	0.5093	0.4629
Mature	ENTD	0.2156	0.2943	0.3468	0.3902	0.4089
	ENML	0.1506	0.1538	0.1534	0.1458	0.1510
	ENML - ENTD	0.6338	0.5519	0.4998	0.4640	0.4401
Shake-out	ENTD	0.2091	0.2718	0.3237	0.3470	0.3773
	ENML	0.1515	0.1604	0.1590	0.1497	0.1420
	ENML - ENTD	0.6394	0.5678	0.5174	0.5033	0.4807
Decline	ENTD	0.2066	0.2844	0.3558	0.3933	0.4378
	ENML	0.1706	0.1757	0.1703	0.1678	0.1605
	ENML - ENTD	0.6228	0.5399	0.4739	0.4388	0.4017

This table reports the relative variable importance per financial statement and per life cycle stage. Life cycle stages are defined following Dickinson (2011) and as of estimation date. An overview is given in Table D.1 in the Appendix. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. B/S, CF/S and I/S denote balance sheet, cash flow statement and income statement respectively. The variables are grouped according to Table B.2 in the Appendix. Importance per financial statement type is defined as the fraction that the respective group contributes to total importance, measured as the sum of absolute SHAP values.

Table 7: Surrogate R^2 per life cycle stage

Life cycle stage	Interactions	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Intro	No	0.8682	0.8461	0.8245	0.8226	0.8274
	Yes	0.9811	0.9671	0.9447	0.9565	0.9492
Growth	No	0.9466	0.9205	0.9176	0.9173	0.9138
	Yes	0.9798	0.9683	0.9609	0.9602	0.9606
Mature	No	0.9464	0.9246	0.9243	0.9180	0.9058
	Yes	0.9799	0.9676	0.9612	0.9593	0.9513
Shake-out	No	0.9447	0.9181	0.9144	0.9141	0.9009
	Yes	0.9864	0.9781	0.9680	0.9666	0.9629
Decline	No	0.8604	0.8341	0.8002	0.7979	0.8021
	Yes	0.9711	0.9183	0.8573	0.8952	0.9025

This table reports the average adjusted R^2 s of the surrogate models per industry. Life cycle stages are defined following Dickinson (2011) and as of estimation date. An overview is given in Table D.1 in the Appendix. The surrogate model without (with) interactions is a linear model which regresses the ENML predictions on the twenty most important variables (plus all possible two-way interactions) as of estimation date. Importance is defined as the fraction that the respective variable contributes to total importance, measured as the sum of absolute SHAP values. For each forecast horizon and industry, we repeat this regression per ENML out-of-sample prediction and report the average adjusted R^2 across surrogates. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings.

Table 8: Variable importance by financial statement type per size tercile

Firm size	Financial statement	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Small	B/S	0.2181	0.3016	0.3508	0.3791	0.4213
	CF/S	0.1703	0.1653	0.1579	0.1614	0.1598
	I/S	0.6115	0.5331	0.4913	0.4595	0.4189
Medium	B/S	0.2015	0.2659	0.3196	0.3333	0.3992
	CF/S	0.1545	0.1588	0.1543	0.1420	0.1421
	I/S	0.6440	0.5753	0.5262	0.5246	0.4587
Large	B/S	0.2151	0.2818	0.3265	0.3528	0.3915
	CF/S	0.1536	0.1493	0.1494	0.1503	0.1426
	I/S	0.6313	0.5690	0.5241	0.4968	0.4659

This table reports the relative variable importance per financial statement and per size tercile. Size is defined as the market capitalization as of estimation date. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. B/S, CF/S and I/S denote balance sheet, cash flow statement and income statement respectively. The variables are grouped according to Table B.2 in the Appendix. Importance per financial statement type is defined as the fraction that the respective group contributes to total importance, measured as the sum of absolute SHAP values.

Table 9: Surrogate R^2 per size tercile

Firm size	Interactions	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Small	No	0.8927	0.8686	0.8489	0.8529	0.8496
	Yes	0.9769	0.9611	0.9478	0.9527	0.9456
Medium	No	0.9361	0.9016	0.8947	0.8925	0.8778
	Yes	0.9763	0.9608	0.9494	0.9472	0.9427
Large	No	0.9155	0.8913	0.8971	0.8910	0.8817
	Yes	0.9755	0.9616	0.9530	0.9510	0.9449

This table reports the average adjusted R^2 s of the surrogate models per industry. Size is defined as the market capitalization as of estimation date. The surrogate model without (with) interactions is a linear model which regresses the ENML predictions on the twenty most important variables (plus all possible two-way interactions) as of estimation date. Importance is defined as the fraction that the respective variable contributes to total importance, measured as the sum of absolute SHAP values. For each forecast horizon and industry, we repeat this regression per ENML out-of-sample prediction and report the average adjusted R^2 across surrogates. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings.

Table 10: Variable importance by financial statement type per industry

Industry	Financial statement	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
1	B/S	0.2189	0.2922	0.3427	0.3702	0.4115
	CF/S	0.1592	0.1649	0.1511	0.1507	0.1449
	I/S	0.6219	0.5429	0.5062	0.4792	0.4436
2	B/S	0.2070	0.2804	0.3385	0.3568	0.4089
	CF/S	0.1648	0.1736	0.1702	0.1510	0.1461
	I/S	0.6282	0.5460	0.4912	0.4922	0.4450
3	B/S	0.2284	0.2974	0.3441	0.3654	0.4027
	CF/S	0.1535	0.1578	0.1493	0.1363	0.1435
	I/S	0.6181	0.5448	0.5067	0.4983	0.4538
4	B/S	0.2165	0.2870	0.3258	0.3470	0.4020
	CF/S	0.1565	0.1713	0.1583	0.1521	0.1448
	I/S	0.6270	0.5418	0.5159	0.5008	0.4533
5	B/S	0.2132	0.3024	0.3478	0.4010	0.4388
	CF/S	0.1742	0.1911	0.1878	0.1578	0.1484
	I/S	0.6126	0.5065	0.4644	0.4412	0.4127

This table reports the relative variable importance per financial statement and per industry. Industries are defined following the Fama-French 5 industry portfolios classification, available on Kenneth French's website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html), and as of estimation date. An overview is given in D.2 in the Appendix. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. B/S, CF/S and I/S denote balance sheet, cash flow statement and income statement respectively. The variables are grouped according to Table B.2 in the Appendix. Importance per financial statement type is defined as the fraction that the respective group contributes to total importance, measured as the sum of absolute SHAP values.

Table 11: Surrogate R^2 per industry

Industry	Interactions	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
1	No	0.9439	0.9226	0.9161	0.9110	0.8984
	Yes	0.9849	0.9759	0.9697	0.9652	0.9632
2	No	0.9484	0.9177	0.9202	0.9180	0.9048
	Yes	0.9803	0.9689	0.9627	0.9594	0.9535
3	No	0.9443	0.9205	0.9107	0.9156	0.9161
	Yes	0.9830	0.9717	0.9600	0.9636	0.9620
4	No	0.9593	0.9479	0.9386	0.9263	0.9209
	Yes	0.9910	0.9823	0.9753	0.9731	0.9757
5	No	0.9362	0.9117	0.9036	0.9035	0.9001
	Yes	0.9792	0.9659	0.9552	0.9556	0.9511

This table reports the average adjusted R^2 s of the surrogate models per industry. Industries are defined following the Fama-French 5 industry portfolios classification, available on Kenneth French’s website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html), and as of estimation date. An overview is given in D.2 in the Appendix. The surrogate model without (with) interactions is a linear model which regresses the ENML predictions on the twenty most important variables (plus all possible two-way interactions) as of estimation date. Importance is defined as the fraction that the respective variable contributes to total importance, measured as the sum of absolute SHAP values. For each forecast horizon and industry, we repeat this regression per ENML out-of-sample prediction and report the average adjusted R^2 across surrogates. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings.

Appendix A Traditional earnings prediction models

Table A.1: Traditional earnings prediction models

Panel A: Traditional model specifications		
Name	Model	Source
L	$\mathbb{E}_t[E_{i,t+\tau}] = \beta_0 + \beta_1 E_{i,t}$	Gerakos and Gramacy (2012)
HVZ	$\mathbb{E}_t[E_{i,t+\tau}] = \beta_0 + \beta_1 E_{i,t} + \beta_2 A_{i,t} + \beta_3 D_{i,t} + \beta_4 DD_{i,t} + \beta_5 NegE_{i,t} + \beta_6 ACC_{i,t}$	Hou et al. (2012)
EP	$\mathbb{E}_t[E_{i,t+\tau}] = \beta_0 + \beta_1 E_{i,t} + \beta_2 NegE_{i,t} + \beta_3 NegE_{i,t} E_{i,t}$	Li and Mohanram (2014)
RI	$\mathbb{E}_t[E_{i,t+\tau}] = \beta_0 + \beta_1 E_{i,t} + \beta_2 NegE_{i,t} + \beta_3 NegE_{i,t} E_{i,t} + \beta_4 B_{i,t} + \beta_5 ACC_{i,t}$	Li and Mohanram (2014)
Panel B: Variable definitions		
Variable	Definition	
E	Income before extraordinary items (ib) / Common shares outstanding (csho)	
A	Total assets (at) / csho	
D	Dividends total (dvt) / csho	
DD	1 if dvt > 0; 0 else	
$NegE$	1 if ib < 0; 0 else	
ACC	(Income before extraordinary items (ib) - Operating activities - net cash flow (oancf)) / csho	
B	Common/Ordinary equity- total (ceq) / csho	

Panel A reports the traditional earnings models estimated. $\mathbb{E}_t[E_{i,t+\tau}]$ denotes the expectation for earnings E of firm i in period $t + \tau$ as of t . β_0 - β_5 are the model coefficients. Panel B reports the variable definitions for the traditional models. Compustat variable names are provided in parentheses. Note the slight changes as opposed to the original papers. More precisely, we scale all variables by common shares outstanding and use a consistent earnings as well as accruals definition.

Appendix B Machine learning earnings prediction models

Table B.1: Hyperparameters for the machine learning models

RF, GBT & DART		
	Maximum number of trees	512
	Learning rate	$\in [0.001, 0.01, 0.1, 1]$
	Maximum depth	$U^{int}(2, 10)$
	Maximum number of leaves	$U^{int}(2, 512)$
	L1-regularization	$U(0, 0.1)$
	L2-regularization	$U(0, 0.1)$
	Feature fraction	$U(0.25, 1)$
	Bagging fraction	$U(0.25, 1)$
	Bagging frequency	$\in (1, 10, 50)$
DART	Dropout rate	$\in (0.05, 0.1, 0.15)$
DART	Probability of skipping dropout	$\in (0.25, 0.5)$
NN		
	Learning rate	$\in [0.001, 0.01, 0.1, 1]$
	L1-regularization	$U(0, 0.1)$
	Dropout	$U(0, 0.5)$
	Number of hidden layers	$\in [1, 2, 3, 4, 5]$
	First layer size	$\in [32, 64, 128]$
	Batch size	$\in [2^{11}, 2^{12}, 2^{13}, 2^{14}]$

This table gives the hyperparameters that we tune and their respective boundaries. U (U^{int}) means drawing from a uniform (integer-wise uniform) distribution. Our choice of hyperparameters and their respective boundaries is based on Bali et al. (2023). We use the *Ray* Python framework to efficiently optimize the hyperparameters (Liaw et al., 2018).

Table B.2: Predictor variables for the machine learning models

	Variable	Compustat description	Financial statement	Component
1	aco	Current assets - other - total	Balance sheet	Current assets
2	acox	Current assets - other - sundry	Balance sheet	Current assets
3	act	Current assets - total	Balance sheet	Current assets
4	am	Amortization of intangibles	Income statement	Depreciation and amortization
5	ao	Assets - other	Balance sheet	Fixed assets
6	aoloch	Assets and liabilities - other - net change	Cash flow statement	Operating cash flow
7	aox	Assets - other - sundry	Balance sheet	Fixed assets
8	ap	Accounts payable - trade	Balance sheet	Liabilities
9	apalch	Accounts payable and accrued liabilities - increase/(decrease)	Cash flow statement	Operating cash flow
10	aqc	Acquisitions	Cash flow statement	Investing cash flow
11	aqi	Acquisitions - income contribution	Income statement	Interest and other
12	aqs	Acquisitions - sales contribution	Income statement	Sales
13	at	Assets - total	Balance sheet	Total assets
14	caps	Capital surplus/share premium reserve	Balance sheet	Equity
15	capx	Capital expenditures	Cash flow statement	Investing cash flow
16	capxv	Capital expend property, plant and equipment schd v	Cash flow statement	Investing cash flow
17	ceq	Common/ordinary equity - total	Balance sheet	Equity
18	ceql	Common equity - liquidation value	Balance sheet	Supplemental
19	ceqt	Common equity - tangible	Balance sheet	Supplemental
20	ch	Cash	Balance sheet	Current assets
21	che	Cash and short-term investments	Balance sheet	Current assets
22	check	Cash and cash equivalents - increase/(decrease)	Cash flow statement	Total cash flow

Continued on next page

Table B.2: Predictor variables for the machine learning models

	Variable	Compustat description	Financial statement	Component
23	cld2	Capitalized leases - due in 2nd year	Balance sheet	Supplemental
24	cld3	Capitalized leases - due in 3rd year	Balance sheet	Supplemental
25	cld4	Capitalized leases - due in 4th year	Balance sheet	Supplemental
26	cld5	Capitalized leases - due in 5th year	Balance sheet	Supplemental
27	cogs	Cost of goods sold	Income statement	Operating expenses
28	cstk	Common/ordinary stock (capital)	Balance sheet	Equity
29	cstkcv	Common stock-carrying value	Balance sheet	Supplemental
30	cstke	Common stock equivalents - dollar savings	Income statement	Interest and other
31	dc	Deferred charges	Balance sheet	Fixed assets
32	dcl	Debt - capitalized lease obligations	Balance sheet	Liabilities
33	dcpstk	Convertible debt and preferred stock	Balance sheet	Supplemental
34	dcvsr	Debt - senior convertible	Balance sheet	Liabilities
35	dcvsub	Debt - subordinated convertible	Balance sheet	Liabilities
36	dcvt	Debt - convertible	Balance sheet	Liabilities
37	dd	Debt - debentures	Balance sheet	Liabilities
38	dd1	Long-term debt due in one year	Balance sheet	Liabilities
39	dd2	Debt - due in 2nd year	Balance sheet	Liabilities
40	dd3	Debt - due in 3rd year	Balance sheet	Liabilities
41	dd4	Debt - due in 4th year	Balance sheet	Liabilities
42	dd5	Debt - due in 5th year	Balance sheet	Liabilities
43	dlc	Debt in current liabilities - total	Balance sheet	Liabilities
44	dltis	Long-term debt - issuance	Cash flow statement	Financing cash flow

Continued on next page

Table B.2: Predictor variables for the machine learning models

	Variable	Compustat description	Financial statement	Component
45	dlto	Other long-term debt	Balance sheet	Liabilities
46	dltp	Long-term debt - tied to prime	Balance sheet	Liabilities
47	dltr	Long-term debt - reduction	Cash flow statement	Financing cash flow
48	dltt	Long-term debt - total	Balance sheet	Liabilities
49	dm	Debt - mortgages and other secured	Balance sheet	Liabilities
50	dn	Debt - notes	Balance sheet	Liabilities
51	do	Discontinued operations	Income statement	Interest and other
52	dp	Depreciation and amortization	Income statement	Depreciation and amortization
53	dpact	Depreciation, depletion and amortization (accumulated)	Balance sheet	Fixed assets
54	dpc	Depreciation and amortization (cash flow)	Cash flow statement	Operating cash flow
55	dpvieb	Depreciation (accumulated) - ending balance (schedule vi)	Balance sheet	Supplemental
56	ds	Debt-subordinated	Balance sheet	Liabilities
57	dudd	Debt - unamortized debt discount and other	Balance sheet	Liabilities
58	dv	Cash dividends (cash flow)	Cash flow statement	Financing cash flow
59	dvc	Dividends common/ordinary	Income statement	Dividends
60	dvp	Dividends - preferred/preference	Income statement	Dividends
61	dvpa	Preferred dividends in arrears	Balance sheet	Supplemental
62	dvt	Dividends - total	Income statement	Dividends
63	dxd2	Debt (excl capitalized leases) - due in 2nd year	Balance sheet	Supplemental
64	dxd3	Debt (excl capitalized leases) - due in 3rd year	Balance sheet	Supplemental
65	dxd4	Debt (excl capitalized leases) - due in 4th year	Balance sheet	Supplemental
66	dxd5	Debt (excl capitalized leases) - due in 5th year	Balance sheet	Supplemental

Continued on next page

Table B.2: Predictor variables for the machine learning models

	Variable	Compustat description	Financial statement	Component
67	ebit	Earnings before interest and taxes	Income statement	EBIT
68	ebitda	Earnings before interest	Income statement	EBITDA
69	esub	Equity in earnings - unconsolidated subsidiaries	Income statement	Interest and other
70	esubc	Equity in net loss - earnings	Cash flow statement	Operating cash flow
71	exre	Exchange rate effect	Cash flow statement	Total cash flow
72	fatb	Property, plant, and equipment - buildings at cost	Balance sheet	Supplemental
73	fatc	Property, plant, and equipment - construction in progress at cost	Balance sheet	Supplemental
74	fate	Property, plant, and equipment - machinery and equipment at cost	Balance sheet	Supplemental
75	fatl	Property, plant, and equipment - leases at cost	Balance sheet	Supplemental
76	fatn	Property, plant, and equipment - natural resources at cost	Balance sheet	Supplemental
77	fato	Property, plant, and equipment - other at cost	Balance sheet	Supplemental
78	fatp	Property, plant, and equipment - land and improvements at cost	Balance sheet	Supplemental
79	fiao	Financing activities - other	Cash flow statement	Financing cash flow
80	fincf	Financing activities - net cash flow	Cash flow statement	Financing cash flow
81	fopo	Funds from operations - other	Cash flow statement	Operating cash flow
82	gp	Gross profit	Income statement	Operating expenses
83	ib	Income before extraordinary items	Income statement	Net income
84	ibadj	Income before extraordinary items - adjusted for common stock equivalents	Income statement	Net income
85	ibc	Income before extraordinary items (cash flow)	Cash flow statement	Operating cash flow
86	ibcom	Income before extraordinary items - available for common	Income statement	Net income
87	icapt	Invested capital - total	Balance sheet	Supplemental
88	idit	Interest and related income - total	Income statement	Interest and other

Continued on next page

Table B.2: Predictor variables for the machine learning models

Variable	Compustat description	Financial statement	Component
89 intan	Intangible assets - total	Balance sheet	Fixed assets
90 intc	Interest capitalized	Income statement	Interest and other
91 intpn	Interest paid - net	Cash flow statement	Operating cash flow
92 invch	Inventory - decrease (increase)	Cash flow statement	Operating cash flow
93 invfg	Inventories - finished goods	Balance sheet	Current assets
94 invo	Inventories - other	Balance sheet	Current assets
95 invrm	Inventories - raw materials	Balance sheet	Current assets
96 invt	Inventories - total	Balance sheet	Current assets
97 invwip	Inventories - work in process	Balance sheet	Current assets
98 itcb	Investment tax credit (balance sheet)	Balance sheet	Liabilities
99 itci	Investment tax credit (income account)	Income statement	Taxes
100 ivaco	Investing activities - other	Cash flow statement	Investing cash flow
101 ivaeq	Investment and advances - equity	Balance sheet	Fixed assets
102 ivao	Investment and advances - other	Balance sheet	Fixed assets
103 ivch	Increase in investments	Cash flow statement	Investing cash flow
104 ivncf	Investing activities - net cash flow	Cash flow statement	Investing cash flow
105 ivst	Short-term investments - total	Balance sheet	Current assets
106 ivstch	Short-term investments - change	Cash flow statement	Investing cash flow
107 lco	Current liabilities - other - total	Balance sheet	Liabilities
108 lcox	Current liabilities - other - sundry	Balance sheet	Liabilities
109 lct	Current liabilities - total	Balance sheet	Liabilities
110 lifr	Lifo reserve	Balance sheet	Supplemental

Continued on next page

Table B.2: Predictor variables for the machine learning models

	Variable	Compustat description	Financial statement	Component
111	lo	Liabilities - other - total	Balance sheet	Liabilities
112	lse	Liabilities and stockholders equity - total	Balance sheet	Total liabilities and equity
113	lt	Liabilities - total	Balance sheet	Liabilities
114	mib	Noncontrolling interest (balance sheet)	Balance sheet	Liabilities
115	mii	Noncontrolling interest (income account)	Income statement	Interest and other
116	mrc1	Rental commitments - minimum - 1st year	Balance sheet	Supplemental
117	mrc2	Rental commitments - minimum - 2nd year	Balance sheet	Supplemental
118	mrc3	Rental commitments - minimum - 3rd year	Balance sheet	Supplemental
119	mrc4	Rental commitments - minimum - 4th year	Balance sheet	Supplemental
120	mrc5	Rental commitments - minimum - 5th year	Balance sheet	Supplemental
121	mrct	Rental commitments - minimum - 5 year total	Balance sheet	Supplemental
122	msa	Marketable securities adjustment	Balance sheet	Supplemental
123	ni	Net income (loss)	Income statement	Net income
124	niadj	Net income adjusted for common/ordinary stock (capital) equivalents	Income statement	Net income
125	nopi	Nonoperating income (expense)	Income statement	Interest and other
126	nopio	Nonoperating income (expense) - other	Income statement	Interest and other
127	np	Notes payable - short-term borrowings	Balance sheet	Liabilities
128	oancf	Operating activities - net cash flow	Cash flow statement	Operating cash flow
129	oiadp	Operating income after depreciation	Income statement	EBIT
130	oibdp	Operating income before depreciation	Income statement	EBITDA
131	pi	Pretax income	Income statement	EBIT
132	ppeg	Property, plant and equipment - total (gross)	Balance sheet	Fixed assets

Continued on next page

Table B.2: Predictor variables for the machine learning models

	Variable	Compustat description	Financial statement	Component
133	ppent	Property, plant and equipment - total (net)	Balance sheet	Fixed assets
134	ppeveb	Property, plant, and equipment - ending balance (schedule v)	Balance sheet	Supplemental
135	prstk	Purchase of common and preferred stock	Cash flow statement	Financing cash flow
136	pstk	Preferred/preference stock (capital) - total	Balance sheet	Equity
137	pstkc	Preferred stock - convertible	Balance sheet	Equity
138	pstkl	Preferred stock - liquidating value	Balance sheet	Supplemental
139	pstkn	Preferred/preference stock - nonredeemable	Balance sheet	Equity
140	pstkr	Preferred/preference stock - redeemable	Balance sheet	Equity
141	pstkrv	Preferred stock - redemption value	Balance sheet	Supplemental
142	re	Retained earnings	Balance sheet	Equity
143	rea	Retained earnings - restatement	Balance sheet	Supplemental
144	reajo	Retained earnings - other adjustments	Balance sheet	Supplemental
145	recch	Accounts receivable - decrease (increase)	Cash flow statement	Operating cash flow
146	recco	Receivables - current - other	Balance sheet	Current assets
147	recd	Receivables - estimated doubtful	Balance sheet	Current assets
148	rect	Receivables - tota	Balance sheet	Current assets
149	recta	Retained earnings - cumulative translation adjustment	Balance sheet	Supplemental
150	rectr	Receivables - trade	Balance sheet	Current assets
151	reuna	Retained earnings - unadjusted	Balance sheet	Equity
152	revt	Revenue - total	Income statement	Sales
153	sale	Sales/turnover (net)	Income statement	Sales
154	seq	Stockholders equity - parent	Balance sheet	Equity

Continued on next page

Table B.2: Predictor variables for the machine learning models

	Variable	Compustat description	Financial statement	Component
155	siv	Sale of investments	Cash flow statement	Investing cash flow
156	spi	Special items	Income statement	Interest and other
157	sppe	Sale of property	Cash flow statement	Operating cash flow
158	sppiv	Sale of property, plant and equipment and investments - gain (loss)	Cash flow statement	Operating cash flow
159	sstk	Sale of common and preferred stock	Cash flow statement	Financing cash flow
160	tlcf	Tax loss carry forward	Balance sheet	Supplemental
161	tstk	Treasury stock - total (all capital)	Balance sheet	Equity
162	tstkc	Treasury stock - common	Balance sheet	Equity
163	tstkp	Treasury stock - preferred	Balance sheet	Equity
164	txach	Income taxes - accrued - increase/(decrease)	Cash flow statement	Operating cash flow
165	txc	Income taxes - current	Income statement	Taxes
166	txdb	Deferred taxes (balance sheet)	Balance sheet	Liabilities
167	txdc	Deferred taxes (cash flow)	Income statement	Operating cash flow
168	txdfed	Deferred taxes-federal	Income statement	Taxes
169	txdfo	Deferred taxes-foreign	Income statement	Taxes
170	txdi	Income taxes - deferred	Income statement	Taxes
171	txditc	Deferred taxes and investment tax credit	Balance sheet	Liabilities
172	txds	Deferred taxes-state	Income statement	Taxes
173	txfed	Income taxes - federal	Income statement	Taxes
174	txfo	Income taxes - foreign	Income statement	Taxes
175	txo	Income taxes - other	Income statement	Taxes
176	txp	Income taxes payable	Balance sheet	Liabilities

Continued on next page

Table B.2: Predictor variables for the machine learning models

	Variable	Compustat description	Financial statement	Component
177	txpd	Income taxes paid	Cash flow statement	Operating cash flow
178	txr	Income tax refund	Balance sheet	Current assets
179	txs	Income taxes - state	Income statement	Taxes
180	txt	Income taxes - total	Income statement	Taxes
181	txw	Excise taxes	Income statement	Taxes
182	wcap	Working capital (balance sheet)	Balance sheet	Supplemental
183	xacc	Accrued expenses	Balance sheet	Liabilities
184	xi	Extraordinary items	Income statement	Interest and other
185	xido	Extraordinary items and discontinued operations	Income statement	Interest and other
186	xidoc	Extraordinary items and discontinued operations (cash flow)	Cash flow statement	Operating cash flow
187	xint	Interest and related expense - total	Income statement	Interest and other
188	xopr	Operating expenses - total	Income statement	Operating expenses
189	xpp	Prepaid expenses	Balance sheet	Current assets
190	xpr	Pension and retirement expense	Income statement	Operating expenses
191	xrd	Research and development expense	Income statement	Operating expenses
192	xrent	Rental expense	Income statement	Operating expenses
193	xsga	Selling, general and administrative expense	Income statement	Operating expenses

This table reports the input variables used in our machine learning models. We also report the Compustat description, the financial statement group as well as the financial statement component group we assign to the respective variable. EBITDA denotes earnings before interest, taxes, depreciation and amortization. EBIT denotes earnings before interest and taxes. EBT denotes earnings before taxes. We scale all variables by common shares outstanding.

Appendix C Implied Cost of Capital models

Table C.1: Implied cost of capital models

Name	Model/Description
GLS	$P_t = B_t + \sum_{\tau=1}^{11} \frac{\mathbb{E}_t[(ROE_{t+\tau} - ICC_{GLS}) \cdot B_{t+\tau-1}]}{(1 + ICC_{GLS})^\tau} + \frac{\mathbb{E}_t[(ROE_{t+12} - ICC_{GLS}) \cdot B_{t+11}]}{(1 + ICC_{GLS})^{11} \cdot ICC_{GLS}}$ <p>This model is given by Gebhardt et al. (2001). P_t denotes the stock price as of the estimation date in t, ICC_{GLS} denotes the implied cost of capital (ICC), B_t denotes the book value of equity per share in t and ROE_t is the return on equity in t. B_t is calculated using the clean surplus relation following Hou et al. (2012). ROE_t is calculated by dividing earnings per share (forecasts) E_t by B_{t-1}. For $ROE_{t+\tau}$ up to $\tau = 3$ we use the respective models earnings per share forecast. Afterwards, we assume $ROE_{t+\tau}$ to revert to the historical industry median by $\tau = 11$ (e.g., Hou et al., 2012). The industry median of ROE is derived using 10 years of data while excluding loss firms (e.g., Gebhardt et al., 2001). We expect ROE to be constant after $\tau = 11$.</p>
CT	$P_t = B_t + \sum_{\tau=1}^5 \frac{\mathbb{E}_t[(ROE_{t+\tau} - ICC_{CT}) \cdot B_{t+\tau-1}]}{(1 + ICC_{CT})^\tau} + \frac{\mathbb{E}_t[(ROE_{t+5} - ICC_{CT}) \cdot B_{t+4}] \cdot (1+g)}{(1 + ICC_{CT})^5 \cdot (ICC_{CT} - g)}$ <p>This model is given by Claus and Thomas (2001). P_t denotes the stock price as of the estimation date in t, ICC_{CT} denotes the implied cost of capital (ICC), B_t denotes the book value of equity per share in t, ROE_t is the return on equity in t and g is the perpetuity growth rate. B_t is calculated using the clean surplus relation following Hou et al. (2012). ROE_t is calculated by dividing earnings per share (forecasts) E_t by B_{t-1}. g is calculated as the current risk-free rate minus 3% (e.g., Hou et al., 2012).</p>
OJ	$P_t = \frac{\mathbb{E}_t[E_{t+1}] \cdot (g_{st} - (\gamma - 1))}{(R - A) - A^2}, \quad \text{with}$ $A = 0.5((\gamma - 1) \frac{\mathbb{E}_t[E_{t+1}] \cdot payout}{P_t}), \quad g_{st} = 0.5(\frac{\mathbb{E}_t[E_{t+3}] - \mathbb{E}_t[E_{t+2}]}{\mathbb{E}_t[E_{t+2}]} - \frac{\mathbb{E}_t[E_{t+5}] - \mathbb{E}_t[E_{t+4}]}{\mathbb{E}_t[E_{t+4}]})$ <p>This model is given by Ohlson and Juettner-Nauroth (2005). P_t denotes the stock price as of the estimation date in t, ICC_{OJ} denotes the implied cost of capital (ICC), $E_{t+\tau}$ is the earnings forecast for $t + \tau$, g_{st} is the short-term growth rate, γ is the perpetual growth rate and $payout$ is the current payout ratio. g_{st} is calculated as the mean of forecasted earnings growth in $\tau = 3$ and $\tau = 5$ (e.g., Hou et al., 2012). γ is the current risk-free rate minus 3% (e.g., Hou et al., 2012). $payout$ is calculated as dividends divided by earnings for profit firms and as dividends divided by $0.06 \cdot \text{total assets}$ for loss firms (e.g., Hou et al. (2012)).</p>
MPEG	$P_t = \frac{\mathbb{E}_t[E_{t+2}] + (ICC_{MPEG} \cdot payout - 1) \cdot \mathbb{E}_t[E_{t+1}]}{ICC_{MPEG}^2}$ <p>This model is given by Easton (2004). P_t denotes the stock price as of the estimation date in t, ICC_{MPEG} denotes the implied cost of capital (ICC) and $E_{t+\tau}$ denotes the earnings per share forecast for $t + \tau$. $payout$ is derived as dividends divided by earnings for profit firms and as dividends divided by $0.06 \cdot \text{total assets}$ for loss firms (e.g., Hou et al., 2012).</p>
GG	$P_t = \frac{\mathbb{E}_t[E_{t+1}]}{ICC_{GG}}$ <p>This model is given by Gordon and Gordon (1997). P_t denotes the stock price as of the estimation date in t, ICC_{GG} denotes the implied cost of capital (ICC) and E_{t+1} denotes the earnings per share forecast for $t + 1$.</p>

This table gives implied cost of capital (ICC) models that we base our composite ICC on. For simplicity, we drop the firm index i . The composite ICC that we use is derived as the average of these ICC.

Appendix D Subsample analysis

Table D.1: Firm life cycle stages

Cash flows from	Intro	Growth	Mature	Shake-out	Shake-out	Shake-out	Decline	Decline
operating activities	-	+	+	+	+	-	-	-
investing activities	-	-	-	+	+	+	-	-
financing activities	+	+	-	+	-	+	+	-

This table provides the firm life cycle stages based on the respective firm's cash flows. A plus sign indicates a positive cash flow and a minus sign indicates a negative cash flow. The classification follows Dickinson (2011).

Table D.2: Industry classifications

No.	Industry Name	SIC Code Ranges
1	Consumer Durables, Nondurables, Wholesale, Retail, and Some Services	0100-0999, 2000-2399, 2700-2749, 2770-2799, 3100-3199, 3940-3989, 2500-2519, 2590-2599, 3630-3659, 3710-3711, 3714-3714, 3716-3716, 3750-3751, 3792-3792, 3900-3939, 3990-3999, 5000-5999, 7200-7299, 7600-7699
2	Manufacturing, Energy, and Utilities	2520-2589, 2600-2699, 2750-2769, 2800-2829, 2840-2899, 3000-3099, 3200-3569, 3580-3621, 3623-3629, 3700-3709, 3712-3713, 3715-3715, 3717-3749, 3752-3791, 3793-3799, 3860-3899, 1200-1399, 2900-2999, 4900-4949
3	Business Equipment, Telephone and Television Transmission	3570-3579, 3622-3622, 3660-3692, 3694-3699, 3810-3839, 7370-7372, 7373-7373, 7374-7374, 7375-7375, 7376-7376, 7377-7377, 7378-7378, 7379-7379, 7391-7391, 8730-8734, 4800-4899
4	Healthcare, Medical Equipment, and Drugs	2830-2839, 3693-3693, 3840-3859, 8000-8099
5	Other – Mines, Construction, Building Materials, Transportation, Hotels, Business Services, Entertainment, Finance	Remaining

This table provides industry classifications based on SIC code ranges. We follow the Fama-French 5 industry portfolios classification, available on Kenneth French's website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

Table D.3: Median PAFE per life cycle stage

Life cycle stage	Model	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Intro	ENTD	0.0573***	0.0685***	0.0737***	0.0832***	0.0915***
	ENML	0.0464***	0.0597***	0.0696***	0.0778***	0.0799***
	ENML - ENTD	-0.0110***	-0.0087***	-0.0041	-0.0054	-0.0116*
Growth	ENTD	0.0228***	0.0308***	0.0358***	0.0413***	0.0465***
	ENML	0.0206***	0.0283***	0.0345***	0.0410***	0.0450***
	ENML - ENTD	-0.0022***	-0.0024***	-0.0013**	-0.0003	-0.0015
Mature	ENTD	0.0222***	0.0289***	0.0331***	0.0369***	0.0417***
	ENML	0.0196***	0.0261***	0.0306***	0.0343***	0.0392***
	ENML - ENTD	-0.0026***	-0.0028***	-0.0026***	-0.0026***	-0.0024***
Shake-out	ENTD	0.0301***	0.0374***	0.0421***	0.0501***	0.0568***
	ENML	0.0267***	0.0339***	0.0397***	0.0459***	0.0511***
	ENML - ENTD	-0.0034***	-0.0035***	-0.0024	-0.0042**	-0.0058***
Decline	ENTD	0.0748***	0.0894***	0.0936***	0.1002***	0.1091***
	ENML	0.0662***	0.0824***	0.0935***	0.0957***	0.1035***
	ENML - ENTD	-0.0086***	-0.0070***	-0.0001	-0.0044	-0.0056

This table reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) per life cycle stage for the two ensemble models. Life cycle stages are defined following Dickinson (2011) and as of estimation date. An overview is given in Table D.1 in the Appendix. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. ENTD denotes the traditional ensemble and ENML denotes the machine learning ensemble. The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year. ***, **, and * denote statistical significance at the 1%, the 5% and the 10% level, respectively. Standard errors used for deriving statistical significance are adjusted following Newey and West (1987) assuming a lag length of three years.

Table D.4: Median PAFE per size tercile

Firm size	Model	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Small	ENTD	0.0594***	0.0695***	0.0773***	0.0851***	0.0958***
	ENML	0.0522***	0.0651***	0.0731***	0.0780***	0.0837***
	ENML - ENTD	-0.0073***	-0.0045**	-0.0042*	-0.0071***	-0.0121***
Medium	ENTD	0.0277***	0.0361***	0.0411***	0.0466***	0.0525***
	ENML	0.0244***	0.0330***	0.0387***	0.0447***	0.0500***
	ENML - ENTD	-0.0033***	-0.0031***	-0.0024***	-0.0020***	-0.0025***
Large	ENTD	0.0167***	0.0224***	0.0260***	0.0294***	0.0330***
	ENML	0.0138***	0.0197***	0.0240***	0.0287***	0.0324***
	ENML - ENTD	-0.0030***	-0.0027***	-0.0020***	-0.0006	-0.0006

This table reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) per size tercile for the two ensemble models. Size is defined as the market capitalization as of estimation date. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. ENTD denotes the traditional ensemble and ENML denotes the machine learning ensemble. The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year. ***, **, and * denote statistical significance at the 1%, the 5% and the 10% level, respectively. Standard errors used for deriving statistical significance are adjusted following Newey and West (1987) assuming a lag length of three years.

Table D.5: Median PAFE per industry

Industry	Model	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
1	ENTD	0.0257***	0.0324***	0.0375***	0.0414***	0.0457***
	ENML	0.0225***	0.0290***	0.0342***	0.0387***	0.0431***
	ENML - ENTD	-0.0032***	-0.0034***	-0.0033***	-0.0028***	-0.0026***
2	ENTD	0.0287***	0.0340***	0.0376***	0.0410***	0.0444***
	ENML	0.0257***	0.0317***	0.0348***	0.0382***	0.0415***
	ENML - ENTD	-0.0030***	-0.0023***	-0.0028***	-0.0028***	-0.0028**
3	ENTD	0.0300***	0.0382***	0.0424***	0.0454***	0.0519***
	ENML	0.0258***	0.0333***	0.0371***	0.0406***	0.0461***
	ENML - ENTD	-0.0042***	-0.0049***	-0.0053***	-0.0049***	-0.0059***
4	ENTD	0.0296***	0.0378***	0.0439***	0.0515***	0.0569***
	ENML	0.0246***	0.0351***	0.0400***	0.0471***	0.0504***
	ENML - ENTD	-0.0050***	-0.0026***	-0.0039***	-0.0043*	-0.0065**
5	ENTD	0.0263***	0.0340***	0.0397***	0.0468***	0.0555***
	ENML	0.0233***	0.0321***	0.0401***	0.0474***	0.0539***
	ENML - ENTD	-0.0031***	-0.0019*	0.0004	0.0006	-0.0016

This table reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) per industry for the two ensemble models. Industries are defined following the Fama-French 5 industry portfolios classification, available on Kenneth French's website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html), and as of estimation date. An overview is given in D.2 in the Appendix. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. ENTD denotes the traditional ensemble and ENML denotes the machine learning ensemble. The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year. ***, **, and * denote statistical significance at the 1%, the 5% and the 10% level, respectively. Standard errors used for deriving statistical significance are adjusted following Newey and West (1987) assuming a lag length of three years.

CFR working papers are available for download from www.cfr-cologne.de.

2025

No.	Author(s)	Title
25-06	D. Hess, F. Simon, S. Weibels	Interpretable Machine Learning for Earnings Forecasts: Leveraging High-Dimensional Financial Statement Data
25-05	B. F. Ballensiefen	Collateral Choice
25-04	C. Andres, F. Brochet, P. Limbach, N. Schumacher	Sell-Side Analysts with Accounting Experience
25-03	W. Bazley, G. Cici, J. Liao	Conflicts of Interest among Affiliated Financial Advisors in 401(k) Plans: Implications for Plan Participants
25-02	A. T. Maître, N. Pugachyov, F. Weigert	Twitter-Based Attention and the Cross-Section of Cryptocurrency Returns
25-01	N. Käfer, M. Mörke, F. Weigert, T. Wiest	A Bayesian Stochastic Discount Factor for the Cross-Section of Individual Equity Options

2024

No.	Author(s)	Title
24-06	V. Beyer, T. Bauckloh	Non-Standard Errors in Carbon Premia
24-05	C. Achilles, P. Limbach, M. Wolff, A. Yoon	Inside the Blackbox of Firm Environmental Efforts: Evidence from Emissions Reduction Initiatives
24-04	Ivan T. Ivanov, T. Zimmermann	The “Privatization” of Municipal Debt
24-03	T. Dyer, G. Köchling, P. Limbach	Traditional Investment Research and Social Networks: Evidence from Facebook Connections
24-02	A. Y. Chen, A. Lopez-Lira, T. Zimmermann	Does Peer-Reviewed Research Help Predict Stock Returns?
24-01	G. Cici, P. Schuster, F. Weishaupt	Once a Trader, Always a Trader: The Role of Traders in Fund Management

2023

No.	Author(s)	Title
23-08	A. Braun, J. Braun, F. Weigert	Extreme Weather Risk and the Cost of Equity
23-07	A. G. Huang, R. Wermers, J. Xue	“Buy the Rumor, Sell the News”: Liquidity Provision by Bond Funds Following Corporate News Events
23-06	J. Dörries, O. Korn, G. J. Power	How Should the Long-term Investor Harvest Variance Risk Premiums?
23-05	V. Agarwal, W. Jiang, Y. Luo, H. Zou	The Real Effect of Sociopolitical Racial Animus: Mutual Fund Manager Performance During the AAPI Hate
23-04	V. Agarwal, B. Barber, S. Cheng, A. Hameed, H. Shanker, A. Yasuda	Do Investors Overvalue Startups? Evidence from the Junior Stakes of Mutual Funds
23-03	A. Höck, T. Bauckloh, M. Dumrose, C. Klein	ESG Criteria and the Credit Risk of Corporate Bond Portfolios
23-02	T. Bauckloh, J. Dobrick, A. Höck, S. Utz, M. Wagner	In partnership for the goals? The level of agreement between SDG ratings
23-01	F. Simon, S. Weibels, T. Zimmermann	Deep Parametric Portfolio Policies

this document only covers the most recent cfr working papers. a full list can be found at www.cfr-cologne.de.



centre for financial research
cfr/university of cologne
albertus-magnus-platz
d-50923 cologne
fon +49(0)221-470-6995
fax +49(0)221-470-3992
kempf@cfr-cologne.de
www.cfr-cologne.de