Financial Analysts Journal

ISSN: 0015-198X (Print) 1938-3312 (Online) Journal homepage: www.tandfonline.com/journals/ufaj20

Fundamental Analysis via Machine Learning

Kai Cao & Haifeng You

To cite this article: Kai Cao & Haifeng You (2024) Fundamental Analysis via Machine Learning, Financial Analysts Journal, 80:2, 74-98, DOI: 10.1080/0015198X.2024.2313692

To link to this article: https://doi.org/10.1080/0015198X.2024.2313692

Research

Financial Analysts Journal | A Publication of CFA Institute
https://doi.org/10.1080/0015198X.2024.2313692

Fundamental Analysis via Machine Learning

Kai Cao and Haifeng You, CFA

Kai Cao is an independent researcher. Haifeng You, CFA, is a chair professor of accounting in the School of Economics and Management and the Shenzhen International Graduate School at Tsinghua University. Send correspondence to Haifeng You at hyou@sem.tsinghua.edu.cn.

We examine the efficacy of machine learning in a central task of fundamental analysis: forecasting corporate earnings. We find that machine learning models not only generate significantly more accurate and informative *out-of-sample* forecasts than the state-of-the-art models in the literature but also perform better compared to analysts' consensus forecasts. This superior performance appears attributable to the ability of machine learning to uncover new information through identifying economically important predictors and capturing nonlinear relationships. The new information uncovered by machine learning models is of considerable economic value to investors. It has significant predictive power with respect to future stock returns, with stocks in the most favorable new information quintile outperforming those in the least favorable quintile by approximately 34 to 77 bps per month on a risk-adjusted basis.

Keywords: earnings forecasts; equity valuation; fundamental analysis; machine learning; market efficiency

Disclosure: No potential conflict of interest was reported by the author(s).

PL Credits: 2.0

Introduction

Fundamental analysis is a cornerstone of capital market operations. It requires the assimilation and processing of vast and varied data, often incurring considerable costs. The primary objective of this paper is to assess the potential of machine learning (ML) to enhance a central task of fundamental analysis: the forecasting of corporate earnings (e.g., Penman 1992; Lee 1999; Richardson, Tuna, and Wysocki 2010; Monahan 2018). The literature has laid a substantial foundation with various earnings prediction models; however, their performance often does not surpass that of the rudimentary random walk (RW) approach (Easton, Kelly, and Neuhierl 2018; Monahan 2018). We posit that the shortcomings of these models, particularly their methodological constraints, could be mitigated by leveraging ML technologies.

Corporate earnings are the cumulative result of a myriad of transactions, each reflected within various financial statement items that can have disparate impacts on future earnings. The intricate nature of these transactions, supported by both economic theory and empirical findings, suggests that the relationships between financial statement items and subsequent earnings are often nonlinear.[1] ML algorithms, with their inherent design to process high-dimensional data and discern complex nonlinear and interaction effects (e.g., Gu, Kelly, and Xiu 2020), are potentially well suited to capture these nuanced effects and intricate patterns. However, the intricacy of ML models can also lead to overfitting. Therefore, the true efficacy of ML in providing superior earnings forecasts remains an empirical question.

To answer this research question, we develop an ML earnings forecasting model that combines three popular ML algorithms: two algorithms based on decision trees (i.e., random forest [RF] and gradient boosting regression [GBR]) and one based on artificial neural networks (ANNs). We supply these algorithms with both the level and first-order difference

Research funding support from Hong Kong RGC project no. T31-604/18-N is highly appreciated.

of a comprehensive set of financial statement items,[2] let the algorithms "learn" from the historical data to determine the underlying relationships, and generate *out-of-sample* forecasts for 134,154 firm-year observations from 1975 to 2019. We compare the accuracy, information content, and investment value of the *out-of-sample* ML forecasts to the forecasts obtained from the RW model and five other models developed in the extant literature: the (first-order) autoregressive model (AR), two models (HVZ and SO) developed by Hou, van Dijk, and Zhang (2012) and So (2013), respectively; and the earnings persistence (EP) and residual income (RI) models proposed by Li and Mohanram (2014).

We find that the ML forecasts are significantly more accurate than not only the RW model but also all extant models. The mean absolute forecast errors of the extant models are approximately 7.79% to 26.53% greater than that of the ML model. Cross-sectional analyses indicate that the ML model leads to even greater accuracy improvements among firms with more difficult-to-forecast earnings. The ML forecast also has greater *information content* as measured by its predictive power with respect to future actual earnings changes (ECH). Forecasted earnings changes (FECH) based on the ML forecast explain 18.61% of the variation in ECH, whereas FECH based on the extant models only explain about 8.07% to 12.22%.

We then test whether the new information uncovered by the ML model can lead to significant improvements in investment decision-making. To this end, we orthogonalize the ML forecast against the contemporaneous forecasts from the extant models and use the residuals to measure the new information uncovered by the ML forecasts (beyond the extant models). The results show that the new information component has significant predictive power with respect to future stock returns. The top quintile of stocks with the most favorable new information significantly outperforms the least favorable new information quintile by approximately 34 to 77 bps per month on a risk-adjusted basis.

We also find that the ML forecasts perform well against analyst consensus forecasts, even though analysts have access to much more information than the financial statements. First, the ML forecast has significantly lower *mean* absolute forecast errors than analyst consensus earnings forecasts for all three-year forecast horizons. Second, the ML forecast has greater relative information content than analyst forecasts in predicting future earnings changes. Third, the ML forecast contains a significant amount of information that analysts fail to consider even though our ML forecasts are based on financial statement data only.

Our paper contributes to the burgeoning literature on the application of ML in finance and accounting research (e.g., Chen et al. 2022; Gu, Kelly, and Xiu 2020; Bao et al. 2020; Bertomeu et al. 2021; Ding et al. 2020) by demonstrating the decisional usefulness of ML technology in one of the most important tasks in fundamental analysis: corporate earnings forecasting. We show that by identifying economically important predictors and capturing subtle nonlinear relationships, ML can better utilize financial statement information and produce significantly more accurate and informative earnings forecasts.

Our paper is also of interest to investment professionals. First, our model can be used to generate earnings forecasts at a low cost even for firms with a short history and those without analyst coverage. It not only offers a less biased alternative to analyst forecasts as a valuation input by stock pickers (e.g., Ohlson 1995; Ohlson and Juettner-Nauroth 2005) but also can be directly used to value many firms without analyst coverage. Second, investors may also easily modify our models to forecast other fundamental variables (e.g., sales, gross profit) that are of considerable importance for investors. Third, our paper also offers a potential systematic trading strategy to quantitative investors, who can build on our study and refine the model by, for example, forecasting future quarterly earnings and other fundamentals or by incorporating non–financial statement information to further improve returns to the strategy.

Related Literature and Extant Earnings Forecasting Models

As Monahan (2018) summarizes, early research mostly adopted the time-series approach to forecast future earnings (e.g., Ball and Watts 1972; Watts and Leftwich 1977). Overall, their results suggest that the simple RW model, which predicts expected future earnings to equal current earnings, generates more accurate out-of-sample forecasts than the more sophisticated autoregressive integrated moving average (ARIMA) models (e.g., Brown 1993; Kothari 2001).[3] The superiority and simplicity of the RW model make it a natural benchmark in evaluating other earnings forecasting models.

There are several potential reasons for the poor performance associated with the ARIMA models. First, these models require a long time series to yield reliable parameter estimates, but the earnings process may not be stationary over a long period. Second, these firm-specific time-series models ignore the rich information in other financial statement line items. To overcome these limitations, subsequent studies turn to cross-sectional or panel-data approaches, which use a pooled cross-section of firms to estimate forecasting models. These models are considered as state-of-the-art earnings prediction models in the literature (e.g., Gerakos and Gramacy 2013; Call et al. 2016).[4] We therefore employ them as alternative benchmarks. We discuss these models below and summarize them with detailed variable definitions in Appendix 1 Panel A.

The first cross-sectional model that we examine is the first-order AR model:

$$E_{i,t+1} = \alpha_0 + \alpha_1 E_{i,t} + \varepsilon_{i,t+1} \tag{1}$$

where $E_{i,t}$ is firm i's earnings in year t. Gerakos and Gramacy (2013) show that the AR model performs well relative to the RW model and is more accurate than many sophisticated models.

The second model, the HVZ model, proposed by Hou, van Dijk, and Zhang (2012), extends the Fama and French (2000, 2006) model and forecasts future earnings as follows:

$$\begin{aligned} E_{i,t+1} = {} & \alpha_0 + \alpha_1 A_{i,t} + \alpha_2 D_{i,t} + \alpha_3 DD_{i,t} + \alpha_4 E_{i,t} \\ & + \alpha_5 NegE_{i,t} + \alpha_6 AC_{i,t} + \varepsilon_{i,t+1} \end{aligned} \tag{2}$$

This model forecasts future earnings with total assets ($A_{i,t}$), the dividend payment ($D_{i,t}$), the dividend-paying dummy ($DD_{i,t}$), historical earnings ($E_{i,t}$), and an indicator variable for negative earnings ($NegE_{i,t}$) and accruals ($AC_{i,t}$).

The third model, the SO model, is developed by So (2013). So modifies the Fama and French (2006) model and forecasts future earnings per share (EPS) using the following model:

$$\begin{aligned} EPS_{i,t+1} = {} & \alpha_0 + \alpha_1 EPS^{+}_{i,t} + \alpha_2 NegE_{i,t} + \alpha_3 AC^{-}_{i,t} + \alpha_4 AC^{+}_{i,t} \\ & + \alpha_5 AG_{i,t} + \alpha_6 NDD_{i,t} + \alpha_7 DIV_{i,t} + \alpha_8 BTM_{i,t} \\ & + \alpha_9 Price_{i,t} + \varepsilon_{i,t+1} \end{aligned} \tag{3}$$

where $EPS^{+}_{i,t}$ is the EPS for positive earnings and is zero otherwise; $NegE_{i,t}$ is an indicator variable for negative earnings; $AC^{-}_{i,t}$ is accruals per share for negative accruals and is zero otherwise; $AC^{+}_{i,t}$ is accruals per share for positive accruals and is zero otherwise; $AG_{i,t}$ is the percentage change in total assets; $NDD_{i,t}$ indicates a zero dividend; $DIV_{i,t}$ is dividends per share; $BTM_{i,t}$ is the book-to-market ratio; and $Price_{i,t}$ is the stock price at the end of the third month after the end of fiscal year t. In addition to financial statement items, the SO model uses the stock price and book-to-market ratio, allowing it to utilize more forward-looking information.

The final two cross-sectional models are proposed by Li and Mohanram (2014):

$$E_{i,t+1} = \alpha_0 + \alpha_1 NegE_{i,t} + \alpha_2 E_{i,t} + \alpha_3 NegE_{i,t} * E_{i,t} + \varepsilon_{i,t+1} \tag{4}$$

$$\begin{aligned} E_{i,t+1} = {} & \alpha_0 + \alpha_1 NegE_{i,t} + \alpha_2 E_{i,t} + \alpha_3 NegE_{i,t} * E_{i,t} \\ & + \alpha_4 BVE_{i,t} + \alpha_5 TACC_{i,t} + \varepsilon_{i,t+1} \end{aligned} \tag{5}$$

Equation (4) allows loss firms to have different earnings persistence (the EP model) from profitable firms. Equation (5) is based on the residual income model (the RI model) proposed by Feltham and Ohlson (1996), which further augments the EP model with the book value of equity ($BVE_{i,t}$) and total accruals ($TACC_{i,t}$) from Richardson et al. (2005).

Although the above models are the state-of-the-art models in the literature, they fail to consistently outperform the RW model (e.g., Monahan 2018; Easton, Kelly, and Neuhierl 2018). Taken at face value, these results seem to suggest that there is not much incremental information in financial statements beyond the bottom-line earnings, which contradicts both the conventional wisdom and the fundamental tenet of financial statement analysis. Given that "*the question of whether historical accounting numbers are useful for forecasting earnings is central to accounting research*" (p. 183, Monahan 2018), both Monahan (2018) and Easton, Kelly, and Neuhierl (2018) call for further research in this area. We posit that, rather than indicating the lack of the decisional usefulness of financial statements or fundamental analysis, the above results might be driven by the methodological limitations of the extant models, which can be overcome by ML algorithms.

ML-Based Earnings Forecasts

The extant models do not make the best use of information in financial statements to forecast future earnings. First, the extant models focus on a small number of aggregate financial statement items, such as bottom-line earnings and total assets, and fail to fully consider many other financial statement line items that could be highly valuable for earnings prediction (e.g., Fairfield, Sweeney, and Yohn 1996; Chen, Miao, and Shevlin 2015). Second, even though economic theories and empirical evidence suggest the prevalence of nonlinear relationships between historical accounting information and future earnings (e.g., Lev 1983; Freeman and Tse 1992; Baginski et al. 1999; Chen and Zhang 2007), these models mostly adopt linear functional forms (or with some simple interactions) and are therefore unlikely to capture these subtle yet important relationships. ML algorithms are designed to handle high-dimensional data and are rather flexible with respect to the functional forms of the underlying relationships. Thus, they can potentially overcome the above limitations and generate better earnings forecasts. Below we describe the development of our ML models.

Financial Statement Line Items as Predictors. Both the financial ratios and raw values of financial statement items can be used in forecasting future earnings.[5] A priori, it is unclear which choice is better. On the one hand, financial ratios are grounded in economic theories, which may help filter out the noise in the raw values of financial items and better capture the key drivers of future earnings. On the other hand, the transformation from raw values to ratios may cause a loss or distortion of information. In a recent study, Bao et al. (2020) compare the predictive power of ML models based on raw accounting numbers versus those based on financial ratios with respect to accounting fraud. They find that raw value–based models perform significantly better. Following their study, we rely on raw accounting values as our primary set of predictors and examine alternative predictor sets of financial ratios in additional analyses. Furthermore, we follow prior literature (e.g., Li and Mohanram 2014) and scale all raw accounting values (including both the input features and the target variable) by the number of common shares outstanding.

As Richardson, Tuna, and Wysocki (2010) suggest, a concern with the early literature of fundamental analysis is the in-sample identification of predictive variables. To mitigate this concern, we select a comprehensive list of key financial statement items and let ML algorithms "learn" from historical data in terms of how to optimally select and combine these items. The resulting models are then used to generate out-of-sample earnings predictions. The detailed list of financial statement items and the rationale for their inclusion are provided in Appendix 1 Panel B. The input predictors (a total of 60) can be broadly categorized as follows:

i. Historical earnings and their major components (8). We include earnings components, as prior literature has shown that disaggregating earnings provides additional information and improves earnings forecasts (Fairfield, Sweeney, and Yohn 1996; Chen, Miao, and Shevlin 2015).
ii. Other important individual income statement items (5). Specifically, we include advertising and R&D expenses, as they tend to generate long-term future benefits (e.g., Lev and Sougiannis 1996; Dou et al. 2021). We also include special and extraordinary items and discontinued operations, as firms may shift core expenses to these items to manipulate core earnings (e.g., Barnea, Ronen, and Sadan 1976; McVay 2006). Finally, we include common dividends, as they may signal firms' future earning power (e.g., Nissim and Ziv 2001).
iii. Summary and individual balance sheet items (16). The balance sheet summarizes resources with potential future economic benefits and may contain incremental information. For example, the literature shows that the book value of equity is an important driver of future earnings (e.g., Feltham and Ohlson 1995; Li and Mohanram 2014), and Ball et al. (2020) argue that retained earnings may measure average earning power better than the book value of equity.
iv. Operating cash flows (1). We include cash flows from operating activities, as prior studies have found that the cash flow component of earnings is more persistent than the accrual component, and separating them helps predict future earnings (e.g., Sloan 1996; Call et al. 2016).
v. The first-order differences of the above items (30). We include them, as prior studies show that changes in financial statement items often contain incremental information beyond the levels (e.g., Kothari 1992; Ohlson and Shroff 1992; Richardson et al. 2005).

ML Algorithms. Following prior studies (e.g., Rasekhschaffe and Jones 2019; Cao et al. 2024), we develop a ML earnings prediction model by ensembling the out-of-sample earnings forecasts generated from several popular ML algorithms. Among them, two algorithms are based on decision trees and one is based on ANNs. Our two decision tree–based algorithms are the standard RF and the GBR algorithms. In implementing our ANN-based algorithm, we adopt the bootstrap aggregating (i.e., bagging) technique by constructing 10 bootstrap samples, with each sample randomly drawing 60% of the observations from the training set. Thereafter, we train an ANN model for each bootstrapped sample and then average the 10 models to generate predictions.[6]

Cross-Validation and Hyperparameter Tuning.

In the implementation of ML, it is imperative to select a model with an appropriate level of complexity because overly simple models tend to underfit the data, while overly complex models tend to have an overfitting problem, and both lead to poor out-of-sample predictability. The level of model complexity is largely determined by the value of certain hyperparameters, which must be set before estimating other parameters such as regression coefficients and neural network weights. For example, values of the hyperparameters such as the maximum depth of the decision trees in the RF and GBR models determine the overall model complexity as well as the number/fraction of input features effectively used in the model.

We search for the "optimal" hyperparameter values through hyperparameter tuning via cross-validation. We use fivefold cross-validation to identify the optimal hyperparameter values that generate the most accurate forecasts on the validation samples. Specifically, for each year in our test sample, we use data from the previous 10 years to train the ML models. We randomly split the 10 years of data into five groups/folds of validation sets, with each fold including 20% of the data. For each fold, we use the remaining 80% of the firm-year observations as the training sample. For each of the ML algorithms, we provide a set of reasonable candidate values for the key hyperparameters (see details provided in Appendix 2). For each combination of the hyperparameters, we compute the mean squared forecast errors of the five validation sets using the model estimated from the remaining 80%, which forms the training set. The mean squared forecast errors on the validation sets are used as the basis to select the optimal hyperparameters, which are then used to train a new model on the training set. We then apply the model to the current year's financial statement data to generate *out-of-sample* earnings forecasts for the following year, which are then compared to the subsequent actual earnings to evaluate the relative performance of various models.

Data, Sample Selection, and Model Estimation Procedure

Our initial sample comprises 267,777 firm-year observations obtained from the intersection of the Compustat fundamentals annual file[7] and the Center for Research in Security Prices (CRSP) data up to fiscal year 2019. We further impose the following data requirements: (1) the following financial statement items must be non-missing: total assets, sales revenue, income before extraordinary items, and common shares outstanding; (2) the stocks must be ordinary common shares listed on the NYSE, AMEX, or NASDAQ; (3) the firms cannot be in the financial (SIC 6000–6999) or regulated utilities (SIC 4900–4999) industries; and (4) the stock prices at the end of the third month after the end of the fiscal year must be greater than US$1. Among the remaining firm-year observations, we replace missing values of the cash flow from operating activities with the corresponding numbers computed from the balance sheet approach (Sloan 1996). We then set the missing values of the remaining line items to zero before computing the first-order differences of the 30 items in Appendix 1. This leaves us with a final sample of 156,256 observations from 1965 to 2019. Because we need data from the past 10 years to estimate the models, our testing sample (i.e., prediction set) starts from 1975 and consists of 142,592 firm-year observations. Table 1 presents the number of firms in the final testing sample by year, where the number of annual observations ranges from 2,299 in 2019 to 4,976 in 1997.

At the third month end after fiscal year t, we generate the out-of-sample forecasts of one-year-forward earnings E_{t+1} for the above testing sample using the aforementioned machine learning algorithms and the 60 predictors. Following prior literature (e.g., Hou, van Dijk, and Zhang 2012; Li and Mohanram 2014), for each year t between 1975 and 2019, we use all observations from the previous 10 years (i.e., year $t-10, t-9, \ldots, t-1$) to estimate the models. As discussed earlier, we use fivefold cross-validation to identify the optimal hyperparameters. Using these optimal values, we retrain the model using the observations from the previous 10 years and then apply the trained models to the predictors of year t to generate earnings forecasts for year $t+1$. For consistency, all extant models are also estimated using the data of the same previous 10 years, and the resulting linear models are applied to their respective predictors in year t to generate earnings forecasts for year $t+1$.[8]

Table 1. Sample Distribution by Year

Year	# obs	Year	# obs	Year	# obs
1975	2,550	1990	3,029	2005	3,303
1976	2,558	1991	3,140	2006	3,259
1977	2,578	1992	3,484	2007	3,166
1978	2,593	1993	3,816	2008	2,945
1979	2,679	1994	4,236	2009	2,583
1980	2,694	1995	4,373	2010	2,747
1981	2,685	1996	4,690	2011	2,673
1982	2,689	1997	4,976	2012	2,538
1983	2,830	1998	4,930	2013	2,499
1984	2,892	1999	4,620	2014	2,522
1985	3,047	2000	4,540	2015	2,497
1986	3,087	2001	3,969	2016	2,419
1987	3,083	2002	3,595	2017	2,383
1988	3,219	2003	3,310	2018	2,349
1989	3,140	2004	3,378	2019	2,299

Notes: This table reports the number of firms with non-missing input features for all models from 1975 to 2019.

Empirical Results

Comparison of Forecast Accuracy. To evaluate the forecast accuracy of the different models, we compare the mean and median absolute forecast errors. We define the forecast error as the difference between the predicted and actual earnings deflated by the market value of equity at the end of three months after the fiscal year end. A larger absolute forecast error indicates less accurate earnings forecasts.[9] Table 2 reports the time-series average of the out-of-sample annual mean and median absolute forecast errors of all models. The ML forecast turns out to be the most accurate forecast, with an average mean absolute forecast error of 0.0687 and an average median absolute forecast error of 0.0291. The benchmark RW model, which prior literature shows to be very difficult to outperform, has an average mean (median) absolute forecast error of 0.0764 (0.0309), approximately 11.20% (6.12%) higher than that of the ML forecast. Consistent with the literature, the extant models are not reliably more accurate than the naïve RW model. More importantly, all extant forecasts are less accurate than the ML forecasts, with mean (median) absolute forecast errors approximately 7.79% to 26.53% (5.87% to 19.45%) higher than that of the ML forecast. Panels B and C of Table 2 present the results for two- and three-year-ahead out-of-sample earnings forecasts, respectively. The results show that ML remains the most accurate forecast. For example, the mean (median) absolute forecast error of RW is approximately 13.06% (6.93%) and 12.54% (6.16%) higher than that of the ML model for two- and three-year-ahead forecasts, respectively.

We also examine whether the ML model predicts the rank and relative level of earnings better than the extant models. For rank prediction, we first rank both the actual earnings and earnings forecasts cross-sectionally and normalize the ranks to [0,1] with the following transformation: $RANK_{norm} = (RANK - RANK_{min})/(RANK_{max} - RANK_{min})$. We then compute the mean absolute difference in the normalized ranks between the actual earnings and the earnings forecast. The results presented in the first five columns of Table 2 Panel D show that the ML forecast still has the lowest mean absolute error of 0.1007, which is about 5% to 14.5% lower than that of the extant model forecasts. For the accuracy of forecasting the relative level, we first remove from both the actual and forecasted earnings their respective cross-sectional median and then compute the mean squared error (MSE) of the median-adjusted numbers. The results presented in the last four columns of Table 2 Panel D show that the ML forecast again has the highest accuracy, with the MSE being about 16.5% to 56.9% lower than that of the extant models.

Cross-Sectional Analysis. The above results suggest that the ML models generate significantly more accurate earnings forecasts than the RW model. For firms with stable performance, historical earnings are quite good indicators for future earnings. The benefit of considering additional financial statement line items and more complex forms of relationships should be of higher importance for firms with more difficult-to-forecast earnings. Table 3 reports the percentage improvement in the forecast accuracy of the ML model relative to the benchmark RW model for subgroups partitioned along the following dimensions: Return on Assets (ROA) volatility, magnitude of accruals, R&D expense, and an indicator variable of loss firms.

The results in Panel A of Table 3 convey the following key messages: first, the ML forecast is significantly more accurate than the RW forecast across the board for all subgroups. Second, ML models lead to significantly greater accuracy improvement among firms with more difficult-to-forecast earnings. For example, relative to the RW model, the ML forecast leads to an accuracy improvement of 15.42%, 17.80%, 11.46%, and 12.99% among firms in the highest quintiles of the ROA volatility, magnitude of total accruals, R&D expense, and loss makers, respectively. In Panel B of Table 3, we conduct a regression analysis of the difference in the forecast accuracy between RW and ML forecasts on the above determinants. Columns (1) through (5) report the univariate regression results, confirming the results in Panel A that ML brings greater improvement in accuracy among firms with higher earnings volatility, a larger magnitude of accruals, more R&D expense, and loss makers. The multivariate regression results in column (6) are largely consistent, except that the

Table 2. Comparison of Forecast Accuracy

A: Accuracy of one-year-ahead earnings forecasts (N = 134,154 firm-years)

	Mean				Median			
		Comparison with ML				Comparison with ML		
	Average*100	DIFF	*t* stat.	%DIFF	Average*100	DIFF	*t* stat.	%DIFF
ML	6.87				2.91			
RW	7.64	0.77	7.85	11.20%	3.09	0.18	5.73	6.12%
AR	7.55	0.68	8.96	9.93%	3.08	0.17	6.16	5.87%
HVZ	7.43	0.55	8.31	8.07%	3.11	0.20	7.95	6.93%
EP	7.42	0.55	8.06	8.03%	3.13	0.22	7.89	7.63%
RI	7.41	0.54	7.61	7.79%	3.11	0.20	7.82	6.90%
SO	8.70	1.82	12.72	26.53%	3.47	0.57	14.81	19.45%

B: Accuracy of two-year-ahead earnings forecasts (N = 123,576 firm-years)

	Mean				Median			
		Comparison with ML				Comparison with ML		
	Average*100	DIFF	*t* stat.	%DIFF	Average*100	DIFF	*t* stat.	%DIFF
ML	9.09				4.42			
RW	10.28	1.19	8.07	13.06%	4.73	0.31	5.25	6.93%
AR	10.18	1.09	9.50	11.99%	4.70	0.28	9.76	6.41%
HVZ	9.71	0.62	7.71	6.84%	4.62	0.19	6.68	4.41%
EP	9.64	0.55	6.03	6.06%	4.66	0.23	6.56	5.31%
RI	9.56	0.47	5.29	5.22%	4.60	0.18	6.01	4.14%
SO	10.31	1.22	11.87	13.42%	4.91	0.49	12.80	11.01%

C: Accuracy of three-year-ahead earnings forecasts (N = 113,601 firm-years)

	Mean				Median			
		Comparison with ML				Comparison with ML		
	Average*100	DIFF	*t* stat.	%DIFF	Average*100	DIFF	*t* stat.	%DIFF
ML	10.88				5.58			
RW	12.25	1.37	7.28	12.54%	5.92	0.34	3.38	6.16%
AR	12.27	1.38	9.86	12.69%	5.93	0.35	8.02	6.28%
HVZ	11.42	0.53	6.96	4.88%	5.73	0.15	4.07	2.64%
EP	11.38	0.50	5.04	4.56%	5.79	0.21	5.03	3.73%
RI	11.21	0.33	3.73	3.03%	5.70	0.12	3.37	2.20%
SO	12.03	1.14	9.62	10.50%	6.11	0.53	11.41	9.47%

continued

Table 2. (continued)

D: Alternative measures of forecast accuracy (N = 134,154 firm-years)

	MAE of Ranks				MSE of De-medianed Forecasts			
		Comparison with ML				Comparison with ML		
	Average*100	DIFF	*t* stat.	%DIFF	Average*100	DIFF	*t* stat.	%DIFF
ML	10.07				1.76			
RW	10.58	0.51	10.92	5.03%	2.21	0.45	3.52	25.67%
AR	10.63	0.56	12.46	5.55%	2.13	0.37	3.52	20.87%
HVZ	10.71	0.63	15.04	6.30%	2.06	0.29	3.32	16.66%
EP	10.58	0.51	10.95	5.04%	2.07	0.31	3.72	17.51%
RI	10.63	0.56	12.26	5.53%	2.05	0.29	3.34	16.54%
SO	11.54	1.47	11.79	14.54%	2.77	1.00	8.77	56.90%

Notes: This table compares the accuracy between the machine learning (ML) forecast and the extant models over the sample period of 1975–2019. Panels A, B, and C report the time-series average of the mean and median absolute forecast errors of one-, two-, and three-year-ahead earnings forecasts, respectively. The absolute forecast error is calculated as the absolute value of the difference between the actual future earnings and the earnings forecasts, scaled by the market equity at the end of three months after the end of the last fiscal year. Panel D reports alternative measures of forecast accuracy, i.e., the time-series average of the mean absolute errors of the scaled rank of forecasts difference and the mean squared errors of the de-medianed forecasts difference. DIFF is the time-series average of the difference, calculated as the mean (median) absolute forecast error of each model minus that of the ML model. The *t* statistic of DIFF time-series is reported accordingly. The percentage difference (%DIFF) is DIFF divided by the time-series average of the annual mean (median) absolute forecast error of the ML model.

Table 3. Cross-Sectional Analysis of Improvement in Forecast Accuracy

*A: The percentage improvement in accuracy of the ML forecast relative to the RW forecasts, i.e., (|RW forecast errors| − |ML forecast errors|)/|RW forecast errors| *100% (N = 129,310, with an average of 2,874 firms per year over the 45-year sample period from 1975 to 2019)*

Partitioning variable	Low	2	3	4	High
ROA Volatility	4.86	5.77	5.88	9.15	15.42
|Total Accruals|/Total Assets	3.51	4.45	6.37	10.12	17.80
|Working Capital Accruals|/Total Assets	6.67	7.67	7.81	8.68	13.81
	MISSING	Low	2	3	High
R&D Expense/Total Assets	9.11	8.54	10.24	10.78	11.46
	Non-Loss	Loss			
Loss	6.64	12.99			

B: Multivariate regression analysis of the improvement in the accuracy of ML forecasts relative to the RW forecasts (N = 129,310)

	Y = |RW forecast errors| − |ML forecast errors|					
	(1)	(2)	(3)	(4)	(5)	(6)
ROA Volatility	0.050					0.006
	(6.25)					(1.45)
|Total Accruals|/Total Assets		0.055				0.038
		(7.35)				(6.44)
|Working Capital Accruals|/Total Assets			0.095			0.053
			(7.92)			(7.32)
R&D Expense/Total Assets				0.012		−0.037
				(2.73)		(−4.74)
Loss					0.020	0.016
					(6.39)	(5.29)

continued

Table 3. (continued)

B: Multivariate regression analysis of the improvement in the accuracy of ML forecasts relative to the RW forecasts (N = 129,310)

	Y = |RW forecast errors| − |ML forecast errors|					
	(1)	(2)	(3)	(4)	(5)	(6)
Const	0.005	0.002	0.001	0.007	0.003	−0.003
	(6.90)	(3.51)	(1.01)	(7.36)	(5.70)	(−3.50)
# years	45	45	45	45	45	45
Avg. # firms per year	2,874	2,874	2,874	2,874	2,874	2,874
Avg adj. R^2	0.89%	2.07%	1.83%	0.13%	2.81%	5.12%

Notes: This table presents a cross-sectional analysis of the improvement in the forecast accuracy of the machine learning (ML) model, relative to that of the random walk (RW) model. The percentage improvement in Panel A is defined as the time-series average of the annual difference in the mean absolute forecast errors (MAFE) between the ML and RW models, divided by the MAFE of the RW model. A positive number indicates improved accuracy of the ML model. In Panel A, we sort all firms into quintiles for each year based on the magnitude of the partition variable (ROA volatility, absolute value of total accruals divided by total assets, and absolute value of working capital accruals divided by total assets, respectively). For R&D expense, we classify all firms with missing R&D expense into a separate group and sort the remaining firms into quartiles for each year based on their R&D expense divided by total assets. We also divide all firms into two groups for each year, depending on whether their earnings are negative. In Panel B, we regress the difference in the forecast accuracy on the determinants each year and report the mean coefficients of the annual regressions, as well as the corresponding Fama–MacBeth *t* statistics.

ROA volatility becomes statistically insignificant and the coefficient on R&D expense flips its sign in the presence of other determinants.

Information Content Analysis. Forecast accuracy is not the sole determinant of the decision usefulness of earnings forecasts. For example, although the RW forecast is more accurate than other forecasts, it provides no information with respect to future earnings changes. In this section, we evaluate the information content of various models by investigating their (out-of-sample) predictive power with respect to the future earnings change, ECH. ECH is computed as the difference between earnings in year $t+1$ and those in year t, scaled by market capitalization at the end of the third month after the end of fiscal year t. We calculate the forecasted earnings change, or FECH, as the predicted earnings for year $t+1$ minus the actual earnings for year t, scaled by market capitalization at the third month end after the end of fiscal year t.

We first compare the mean correlation coefficients between the ECH and FECH calculated from different models. The left two panels of Table 4 show that the Pearson (Spearman) correlation coefficients between FECH and ECH range from 0.199 to 0.321 (0.117 to 0.179) for the extant models, which are all lower than the correlations between the ECH and FECH of the ML forecast, 0.413 (0.3). We then estimate the univariate Fama–MacBeth regression of ECH on the FECH of different models. To facilitate the comparison of the coefficients, we follow the literature to standardize FECH so that it has a zero mean and unit variance each year. The three columns in the middle panel of Table 4 show that the coefficients on FECH for the extant models range from 0.0304 to 0.0480, explaining between 8.07% and 12.22% of the cross-sectional variation in realized earnings changes. In contrast, the FECH based on the ML forecast has a regression coefficient of 0.0606 and an explanatory power of 18.61%. Next, we estimate a multivariate regression of ECH on FECH based on the ML model by controlling for the FECH of all extant models. The right panel of Table 4 shows that the coefficient on $FECH^{ML}$ is significantly positive, with a *t* statistic of 17.98. In contrast, most of the FECH coefficients based on the extant models become statistically insignificant (or have the wrong sign), except for that of the SO model, which also uses forward-looking nonfinancial statement predictors such as the stock price and book-to-market ratio.

Economic Significance Analysis. The above results suggest that ML technology helps generate more accurate earnings forecasts of (*statistically*) significant incremental information beyond the extant models. However, it is unclear whether such new information is economically significant. To shed light on the economic significance of the results, we test

Table 4. Information Content Analysis

	Correlation with *ECH*		Univariate regression *Depvar: ECH*			Multivariate regression *Depvar: ECH*	
	Pearson	Spearman	Coeff.	*t* stat.	Avg. R^2	Coeff.	*t* stat.
FECH^{ML}	0.413	0.300	0.0606	12.01	18.61%	0.0589	17.98
FECH^{AR}	0.199	0.117	0.0304	4.81	8.07%	0.0078	1.59
FECH^{HVZ}	0.283	0.179	0.0422	8.93	9.98%	–0.009	–2.48
FECH^{EP}	0.321	0.154	0.0480	9.96	12.22%	0.0067	0.52
FECH^{RI}	0.313	0.148	0.0467	9.95	11.68%	–0.0159	–1.54
FECH^{SO}	0.291	0.153	0.0440	10.47	9.66%	0.0132	4.46
Avg. R^2						20.87%	

Notes: This table performs the information content analysis of the machine learning (ML) forecast against the extant models. The left panel reports the average annual cross-sectional Pearson (Spearman) correlation coefficients between the forecasted earnings changes calculated using various models and the actual earnings changes for a total of 134,154 firm-year observations, with an average of 2,981 firms per year over the 45-year sample period from 1975 to 2019. The middle panel reports the univariate Fama–MacBeth regression results. In the regression, all forecasted earnings changes are standardized to have a zero mean and unit variance each year. The right panel reports the multivariate Fama–MacBeth regression results. Specifically, we regress earnings changes (ECH) on the forecasted earnings changes (FECH) of the ML forecasts and control for all earnings changes predicted using the extant models. All independent variables are standardized to have a zero mean and unit variance each year. All earnings changes are scaled by the market equity at the end of three months after the end of the last fiscal year. The table presents the average coefficients, along with the Fama–MacBeth *t* statistics and the average adjusted R^2. The subscripts are omitted for brevity.

whether the new information in the ML forecasts has (statistically and economically) significant predictive power with respect to future stock returns.[10] To capture the new information uncovered by ML models, we orthogonalize the ML-based forecasts against the forecasts generated using the RW and extant models. Specifically, we run an annual *cross-sectional* regression of the ML forecasts on the RW forecasts and the forecasts of the five extant models each year and use the residual to measure the new information uncovered by the ML models. Then, we estimate the following models to test whether the residual forecasts predict future stock returns:

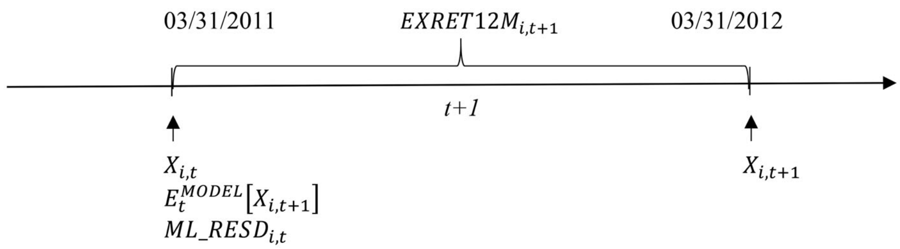
$$EXRET12M_{i,t+1} = \beta_0 + \beta_1 ML\_RESD_{i,t} + \sum\nolimits_{s=1}^{S} \gamma_s X_{i,s,t} + IndustryFE + \varepsilon_{i,t+1} \quad (6)$$

where $EXRET12M_{i,t+1}$ is the one-year-ahead cumulative excess return starting from the fourth month of fiscal year $t+1$ for firm *i*. $ML\_RESD_{i,t}$ is the residual from the regression that orthogonalizes the ML forecast against the RW model and the five extant models. $X_{i,s,t}$ is the end-of-year *t* value of firm *i*'s control characteristics. We follow Bartram and Grinblatt (2018) to select the list of control variables and also control for industry fixed effects. Furthermore, we add ROE (return on equity) and INV (growth rate of total assets) in light of Fama and French (2015). The timeline for the computation of $EXRET12M_{i,t+1}$ and $ML\_RESD_{i,t}$ is provided in Figure 1, and the detailed variable definitions are provided in Panel C of Appendix 1. If the new information component has already been fully priced (when the year *t* financial statements are announced), the coefficient on $ML\_RESD_{i,t}$, i.e., β_1 would be statistically insignificant.

Table 5 presents the Fama–MacBeth regression results of model (6). Column (1) reports the regression results using the same set of control variables as in Bartram and Grinblatt (2018). The results show that the new information component of the ML forecast, ML_RESD, is significantly associated with future 12-month excess returns, even after controlling for various return-predictive factors. In column (2), we further control for ROE and INV (Fama and French 2015). The results remain robust, showing that the new information component still exhibits a positive and statistically significant association with future stock returns.[11]

We also conduct a portfolio analysis. Specifically, at the beginning of each month, we estimate the new information component as the residual from the regression of the ML forecasts on the contemporaneous forecasts generated from the RW model and the five extant

Figure 1. Timeline of Return Prediction Analysis



Notes: This figure presents the timeline for the variables used in the future return prediction analysis. Assume that fiscal year $t+1$ (2011) of firm *i* ends on 12/31/2011. In the future return prediction analysis, we regress $EXRET12M_{i,t+1}$, which is the excess cumulative return over the period of 04/01/2011 to 03/31/2012 on $ML\_RESD_{i,t}$, which is a proxy for the new information component of a machine learning forecast, $E_t^{MODEL}[X_{i,t+1}]$, estimated on 03/31/2011.

models. We then sort all stocks into quintiles based on the residuals for each three-digit SIC industry. We construct a hedge portfolio that takes long positions in quintiles with the most favorable new information and short positions in quintiles with the least favorable new information. Table 6 reports the mean monthly return, Sharpe ratio (annualized), CAPM alpha, Fama–French three-factor alpha, Carhart four-factor alpha, Fama–French five-factor alpha (i.e., the three-factor model plus the Conservative Minus Aggressive (CMA) and Robust Minus Weak (RMW) factors), and the alpha after controlling for all factors in the Fama and French dataset, for the five new information component quintiles, as well as the hedge portfolios that take long positions in the top quintiles and short positions in the bottom quintiles.

Panel A of Table 6 reports the results for the equally weighted portfolios, showing that the mean monthly excess returns increase monotonically from 0.50% for the lowest quintile to 1.23% for the highest quintile. The hedge portfolio generates a monthly mean return of 0.73%. The Sharpe ratio also increases monotonically from 0.25 to 0.69 for the five quintiles. The hedge portfolio generates a Sharpe ratio of 1.29. Furthermore, we find that the risk-adjusted returns (or alphas) increase consistently with the quintile rank. Finally, even after controlling for all factors in the Fama and French dataset, the hedge portfolio still earns a monthly alpha of 51 bps.

The results for the value-weighted portfolios appearing in Panel B of Table 6 are slightly weaker but still significant both statistically and economically. The hedge portfolio generates a monthly mean return of 0.48%, with a Sharpe ratio of 0.6. Furthermore, all Fama–French factor alphas are still significant for the hedge portfolio. For example, the hedge portfolio yields a Fama–French five-factor alpha of approximately 45 bps per month. In Figure 2, we plot the cumulative log returns (value-weighted) for the five quintiles and the hedge portfolio. The plot shows that the cumulative returns increase monotonically with the quintile rank, and the hedge portfolio returns are reasonably consistent over time.[12]

Additional Analyses

The Importance of Nonlinear Effects. We conduct several additional analyses to better understand the underlying reasons for the superior performance of the ML forecasts. We first plot the feature (Gini) importance charts of the RF model to check whether they use economically sensible features to generate predictions. Figure 3 shows that past earnings and operating cash flows are important predictors for future earnings, ranked first and third, respectively. Interestingly, total income tax and its first-order difference are the second and fourth most important features, respectively. These findings are consistent with recent literature on the important role of tax income or expenses in capturing the quality of earnings and predicting future fundamentals and stock returns (e.g., Lev and Nissim 2004; Hanlon 2005; Hanlon, Laplante, and Shevlin 2005; Thomas and Zhang 2011, 2014). Panels A through C of Figure 4 present the accumulated local effects (ALE) plots (Apley and Zhu 2020) of the top five most important features of the RF models[13] for 1975, 1995, and 2015, respectively. The figures show obvious nonlinear relationships between these input features and future earnings.[14]

Table 5. Regression Analysis of Future 12-Month Cumulative Excess Returns on the New Information Component of Machine Learning Forecasts

	Dep = EXRET12M	
	(1)	(2)
ML_RESD	0.807	0.789
	(5.55)	(5.28)
Beta	−0.005	−0.006
	(−1.08)	(−1.28)
SIZE	0.063	0.061
	(5.87)	(5.66)
BM	−0.044	−0.041
	(−2.35)	(−2.29)
MOM	−0.246	−0.216
	(−5.09)	(−4.57)
ACC	0.006	0.007
	(0.79)	(0.94)
ST_Reversal	−0.064	−0.067
	(−1.37)	(−1.43)
LT_Reversal	−0.013	−0.011
	(−2.58)	(−2.16)
SUE	0.113	0.117
	(2.62)	(2.64)
Gross Profitability	0.130	0.113
	(7.55)	(6.41)
Earnings yield	0.208	0.174
	(5.39)	(5.13)
ROE		0.042
		(2.85)
INV		−0.042
		(−3.07)
Const	YES	YES
Industry FE	YES	YES
Number of years	44	44
Avg # firms per year	2,009	2,009
Avg. R^2	5.59%	4.32%

Notes: This table reports the Fama–MacBeth regression results that regress future one-year-ahead cumulative excess returns (*EXRET12M)* on the new information component of the machine learning (ML) forecast (*ML_RESD*) and various known return-predicting factors and industry fixed effects (three-digit SIC). EXRET12M is the difference between cumulative stock returns starting from the fourth month of the next fiscal year minus the cumulative risk-free rate (from the Fama and French database) over the same period. *ML_RESD* is the residual from the annual cross-sectional regression of the ML forecast on forecasts of the random walk (RW) model and the five extant models. The definitions of the control variables are given in Appendix 1. All independent variables are winsorized at 1% and 99% each year. The table presents the average coefficients with the corresponding Fama–MacBeth *t* statistics, along with the average adjusted *R*-squares. The subscripts are omitted for brevity.

In order to provide further empirical evidence on the importance of nonlinear and interaction effects, we develop a linear forecast (LR) using the same set of input features and combine the out-of-sample predictions of the OLS, LASSO, and Ridge algorithms. Untabulated results suggest that on a standalone basis, the LR forecast is statistically more accurate and informative (about future ECH) than the extant models but is significantly less accurate and informative than the ML forecast. Furthermore, Panel A of Table 7 shows that in the multivariate regression of ECH on FECH based on both the LR and ML forecasts, both the coefficients and the *t* statistics are greater for the ML forecast than the linear forecasts. Finally, to better understand the collective importance of nonlinear effects in the ML forecast, we regress the ML forecast on the 60 inputs using OLS to filter out the linear effects. The fitted value of the regression (FITTED) captures the linear effect of the predictors, while the residuals capture the nonlinear and interaction effects in the ML forecasts (denoted as NONLR). We then decompose $FECH^{ML}$ into $FECH^{FITTED}$ (=FITTED − current earnings) and NONLR and examine their predictive information content. The results reported in Panel B of Table 7 show that NONLR has significant predictive power with respect to ECH in the presence of $FECH^{FITTED}$ and the FECHs of the extant models, again suggesting that the ability to accommodate nonlinearity and interaction effects allows the ML forecast to uncover a significant amount of new information.

Comparison between ML Forecasts and Analyst Consensus Forecasts.

We compare the ML forecasts to the consensus analyst forecasts (Analyst) issued around the same time, that is, the third month after the fiscal year end. Because Bradshaw et al. (2012) find that the superiority of analyst forecasts over RW varies over forecast horizons, we conduct a comparison for horizons ranging from one to three years. Table 8 Panel A shows that the mean absolute forecast errors of the ML forecast are significantly lower than those of the analyst consensus forecasts for all three forecast horizons (0.0541, 0.0679, and 0.0776 for ML vs. 0.0588, 0.0742, and 0.0922 for analyst forecasts). The median results are largely similar, except for the one-year horizon, where the ML forecast has a slightly higher median absolute forecast error (0.0219 vs. 0.0202). Panel B of Table 8 further shows that the ML forecast not only has slightly greater *relative* information content than analyst forecasts (see

Table 6. Portfolio Analysis of the New Information Component of Machine Learning Forecasts

A: Monthly return/alpha of the equal-weighted portfolios (N = 1,630,370, with an average of 3,088 stocks per month in the 528-month sample period)

	Lowest	2	3	4	Highest	Hedge
Mean Return	0.50	0.70	0.84	1.01	1.23	0.73
	(1.64)	(2.64)	(3.36)	(4.19)	(4.56)	(8.53)
Shape Ratio	0.25	0.40	0.51	0.63	0.69	1.29
CAPM Alpha	−0.33	−0.09	0.06	0.25	0.44	0.77
	(−1.73)	(−0.67)	(0.52)	(2.34)	(2.99)	(9.14)
FF3 Alpha	−0.46	−0.19	−0.02	0.16	0.28	0.74
	(−3.2)	(−2.34)	(−0.4)	(2.85)	(2.92)	(8.97)
Carhart4 Alpha	−0.18	−0.03	0.10	0.27	0.47	0.65
	(−1.13)	(−0.34)	(1.73)	(4.23)	(4.63)	(7.05)
FF5 Alpha	−0.18	−0.04	0.04	0.18	0.35	0.53
	(−1.11)	(−0.39)	(0.59)	(2.55)	(3)	(6.35)
Alpha (all factors)	−0.06	0.02	0.10	0.23	0.45	0.51
	(−0.38)	(0.2)	(1.74)	(3.75)	(4.33)	(5.89)

B: Monthly return/alpha of the value-weighted portfolios (N = 1,630,370, with an average of 3,088 stocks per month in the 528-month sample period)

	Lowest	2	3	4	Highest	Hedge
Mean Return	0.48	0.55	0.69	0.77	0.96	0.48
	(1.98)	(2.68)	(3.3)	(3.81)	(4.68)	(3.98)
Shape Ratio	0.30	0.40	0.50	0.57	0.71	0.60
CAPM Alpha	−0.30	−0.15	−0.03	0.07	0.28	0.59
	(−3.09)	(−2.9)	(−0.52)	(1.81)	(3.9)	(4.79)
FF3 Alpha	−0.37	−0.15	0.02	0.10	0.28	0.65
	(−3.88)	(−3.13)	(0.47)	(2.65)	(4.11)	(5.45)
Carhart4 Alpha	−0.20	−0.13	0.03	0.12	0.30	0.50
	(−2.2)	(−2.85)	(0.8)	(3.03)	(4.14)	(4.15)
FF5 Alpha	−0.26	−0.14	0.04	0.06	0.18	0.45
	(−2.63)	(−2.98)	(0.85)	(1.3)	(2.68)	(3.78)
Alpha (all factors)	−0.14	−0.12	0.06	0.06	0.20	0.34
	(−1.48)	(−2.69)	(1.25)	(1.38)	(2.99)	(2.88)

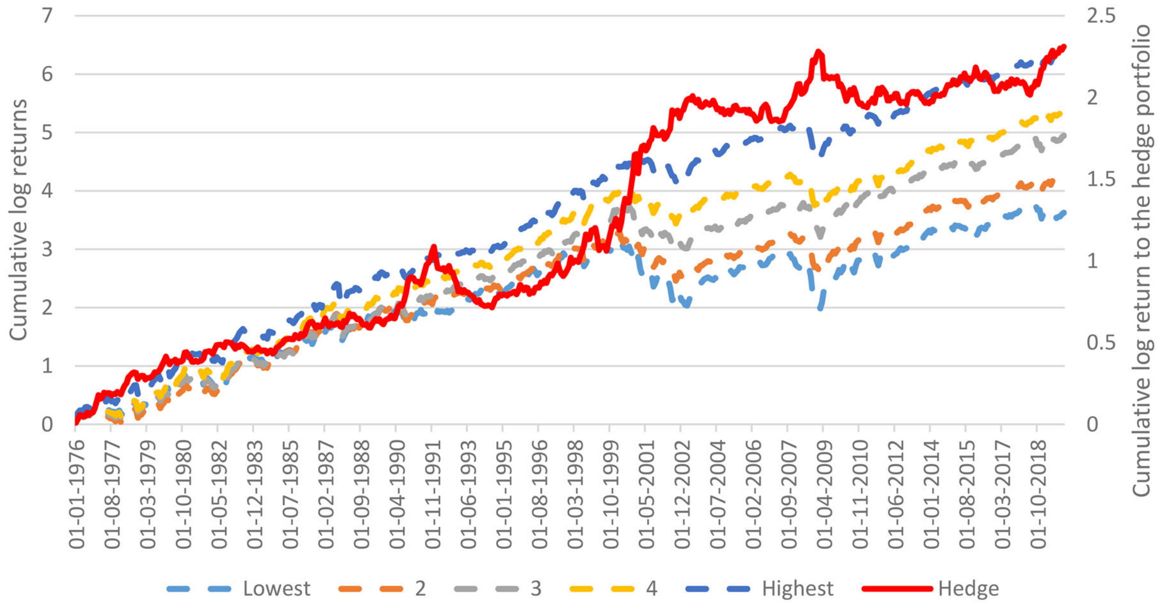
Notes: This table summarizes the monthly return/alpha to quintiles sorted based on the new information uncovered using the machine learning (ML) models. At the beginning of each month, we estimate the new information component as the residual from the cross-sectional regression of ML forecasts on the forecasts generated from the random walk (RW) model and the five extant models. We sort the stocks into quintiles based on the resulting residuals for each three-digit SIC industry and report the return performance of the hedge portfolio, which takes long positions in quintiles with the most favorable new information and short positions in quintiles with the least favorable new information. Panel A reports the results for the equal-weighted portfolios. Panel B reports the results for the value-weighted hedge portfolios. We report the mean monthly returns and risk-adjusted returns (alpha) to the portfolios with the corresponding t statistics in the parentheses.

models 1 and 2) but also has significant *incremental* information beyond them (see models 3 and 4 of the panel). These results suggest that (i) analysts do not *fully* incorporate financial statement information into their forecasts and (ii) investors can benefit from the ML forecasts even if they already have access to analysts' forecasts, as the ML forecasts contain a considerable amount incremental information beyond analyst forecasts.[15]

Transaction Costs and Other Implementation Considerations.

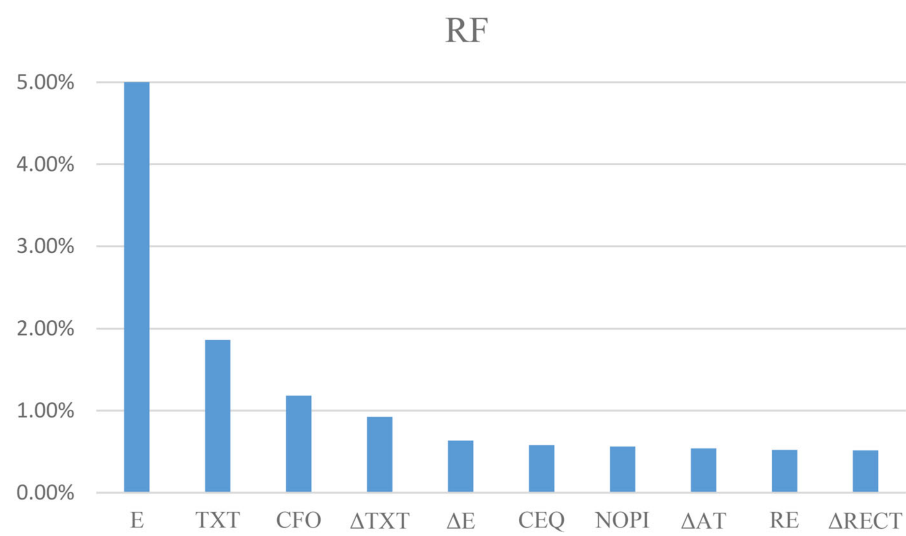
In this section, we discuss the effect of transaction costs and other implementation issues. Our investment strategy is based on the ML earnings forecasts derived from annual financial statement data and has relatively low turnover. Untabulated analyses show that for the value-weighted portfolio, the monthly portfolio turnover is less than 30% most of the time and averages

Figure 2. Cumulative Value-Weighted Returns to Quantiles of the New Information Component of the Machine Learning (ML) Forecasts



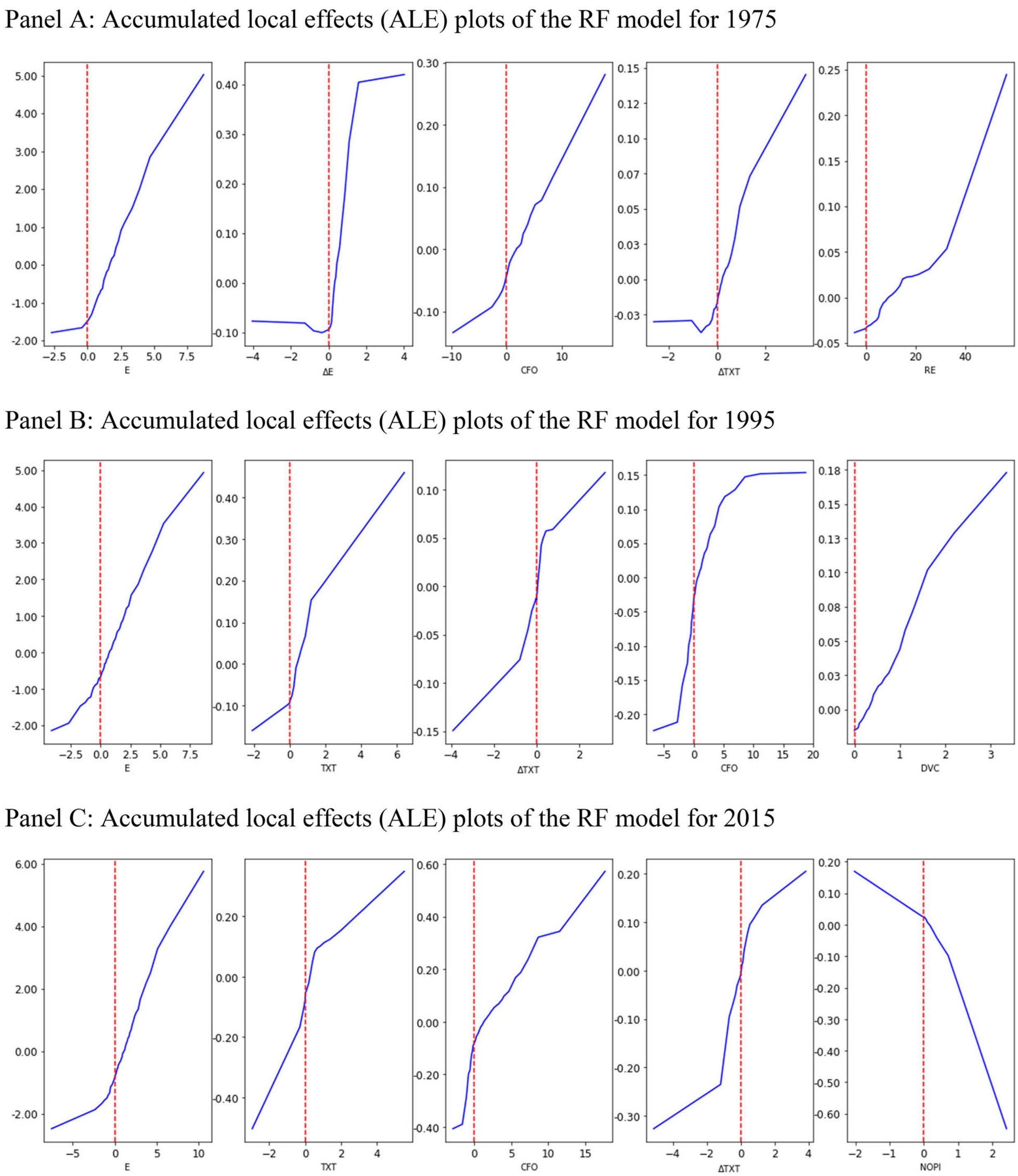
Notes: This figure plots the cumulative (log) returns to the value-weighted quintile portfolios sorted on the new information component of the ML forecasts (i.e., regression residual of the ML forecast on the extant forecasts) and the hedge portfolio, which is the difference between the two extreme quintiles. The left vertical axis is for the cumulative log returns to the five quintile portfolios while the right vertical axis is for the cumulative log return to the hedge portfolio (the solid red line).

Figure 3. Top 10 Influential Features of the Random Forest Model



Notes: This figure plots the average feature importance extracted from the fitted models of the random forest (RF) that we train with data from the 1975–2019 period. The higher the importance score, the more important the feature is. To facilitate representation, we set the maximum of the y axis at 0.05, while the average feature importance values for earnings ("E") is 0.7932.

Figure 4. Accumulated Local Effects (ALE) of the Top Five Most Influential Features



Notes: (A) Accumulated local effects (ALE) plots of the random forest (RF) model for 1975. (B) Accumulated local effects (ALE) plots of the random forest (RF) model for 1995. (C) Accumulated local effects (ALE) plots of the random forest (RF) model for 2015.

Table 7. The Importance of Nonlinear and Interaction Effects

A. Information content comparison between the ML and linear forecasts (N = 134,154, with an average of 2,982 observation per year over the 45-year sample period from 1975 to 2019)

Model: $ECH = \beta_0+\beta_1 FECH^{ML}+\beta_2 FECH^{LR}+\beta_3 FECH^{AR}+\beta_4 FECH^{HVZ}+\beta_5 FECH^{EP}+ \beta_6 FECH^{RI}+\beta_7 FECH^{SO}+\varepsilon$

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	Avg. R^2
Model 1	0.0016	0.0515	0.011						0.1913
	(0.67)	(10.91)	(3.62)						
Model 2	0.0016	0.0466	0.0163	0.0071	−0.0097	0.0081	−0.02	0.0149	0.2149
	(0.67)	(12.19)	(5.11)	(1.48)	(−2.66)	(0.62)	(−1.89)	(4.97)	

B. Information content analysis on the linear (FITTED) vs. nonlinear (NONLR) components of the ML forecasts (N = 134,154, with an average of 2,982 observation per year over the 45-year sample period from 1975 to 2019)

Model: $ECH = \beta_0+\beta_1 NONLR + \beta_2 FECH^{FITTED}+\beta_3 FECH^{AR}+\beta_4 FECH^{HVZ}+\beta_5 FECH^{EP}+ \beta_6 FECH^{RI}+\beta_7 FECH^{SO}+\varepsilon$

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	Avg. R^2
Model 1	0.0016	0.0368	0.0451						0.1765
	(0.67)	(6.21)	(9.29)						
Model 2	0.0016	0.0844	0.0525	−0.018	−0.013	−0.0056	−0.0384	0.0153	0.2119
	(0.67)	(11.95)	(13.35)	(−3.7)	(−3.4)	(−0.54)	(−4.01)	(5.08)	

Notes: This table examines the importance of nonlinear and interaction effects in explaining the superior information content in the machine learning (ML) forecasts. Panel A presents the Fama–MacBeth regression of earnings changes (ECH) on the forecasted earnings changes (FECH) of both the linear and nonlinear composite forecasts, without (Model 1) and with (Model 2), controlling for the FECHs of the extant models. Panel B presents the Fama–MacBeth regression of ECH on the nonlinear component of the ML forecasts (i.e., the residual from the annual cross-sectional regression of the ML forecasts on the 60 input features) and the FECH of the fitted value from the regression, without (Model 1) and with (Model 2), controlling for the FECHs of the extant models. All independent variables are standardized to have a zero mean and unit variance each year. All earnings changes are scaled by the market equity at the end of three months after the end of the last fiscal year. The table presents the average coefficients, along with the corresponding Fama–MacBeth *t* statistics in brackets and the average adjusted *R*-squares. The subscripts are omitted for brevity.

approximately 28.5%. Using an aggressive estimate of the round-trip cost of 50 bps, the annualized return would be reduced by approximately 7 bps per month (= 50 bps / 2 * 28.5%).[16] However, for retail investors who trade aggressively with market orders and buy (sell) stocks at the ask (bid) prices, the average monthly return to the value-weighted strategies would drop substantially to about 20 bps, with a Sharpe ratio of 0.25.

However, it is worth pointing out that simple decile-/quintile-ranked portfolios are far from an efficient way to implement a quantitative strategy. Sophisticated investors mostly utilize risk and transaction cost models together with portfolio optimizers to construct more (risk and *t* cost) efficient portfolios, which tend to yield much better performance (e.g., Sivaramakrishnan, Brown, and Kasturi 2018). Furthermore, investors may also be able to improve their strategy performance by using more advanced ML technology, incorporating non-accounting information, and forecasting quarterly earnings, etc. We leave these further refinements to interested readers.

Conclusions

Our exploration of ML in the context of fundamental analysis, particularly in forecasting corporate earnings, yields enlightening results. ML models outstrip contemporary earnings prediction models from the literature in terms of forecast accuracy and the richness of information provided. This enhanced performance of ML models is primarily due to their ability to unearth new economically significant information from the publicly available financial statement data. They achieve this by identifying the key predictors and capturing complex nonlinear relationships that traditional models and methodologies might overlook or be unable to process effectively.

Table 8. Comparison between the Machine Learning Forecasts and Analyst Forecasts

A. Forecast accuracy comparison

	Mean absolute forecast errors*100			Median absolute forecast errors*100		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$
ML	5.41	6.79	7.76	2.19	3.14	3.68
Analyst	5.88	7.42	9.22	2.02	3.30	4.41
ML-Analyst	−0.47	−0.63	−1.46	0.17	−0.16	−0.73
(t-stat)	(−4.75)	(−5.77)	(−7.77)	(5.46)	(−2.99)	(−5.74)
# years	35	34	33	35	34	33
Avg # firms per year	2,279	1,839	696	2,279	1,839	696
N	79,766	62,531	22,956	79,766	62,531	22,956

B. Information content comparison (N = 79,766, with an average of 2,279 observations per year in the 35-year period from 1985 to 2019)

Model: $ECH = \beta_0+\beta_1 FECH^{ML}+\beta_2 FECH^{Analyst}+\beta_3 FECH^{AR}+\beta_4 FECH^{HVZ}+\beta_5 FECH^{EP}+ \beta_6 FECH^{RI}+\beta_7 FECH^{SO}+\varepsilon$

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	Avg. R^2
Model 1	−0.0019	0.051							19.36%
	(−0.80)	(8.99)							
Model 2	−0.0019		0.0492						18.92%
	(−0.80)		(7.53)						
Model 3	−0.0019	0.031	0.0289						23.61%
	(−0.80)	(11.12)	(5.14)						
Model 4	−0.0019	0.0352	0.0307	0.018	−0.0135	0.0077	−0.0293	0.0124	26.85%
	(−0.80)	(13.78)	(7.10)	(2.67)	(−2.83)	(0.84)	(−3.86)	(5.94)	

Notes: This table compares the composite forecasts based on the machine learning model (ML) and analyst consensus forecasts. Panel A reports the time-series average of the mean and median absolute forecast errors of the one- to three-year-ahead earnings forecasts for the ML model and the analyst consensus forecasts issued in the third month after the last fiscal year end. Panel B reports the Fama–MacBeth regression results. All independent variables are standardized to have a zero mean and unit variance each year. The panel presents the average coefficients, along with the Fama–MacBeth *t* statistics in brackets and the average adjusted *R*-square.

The new information that ML models bring to light is not merely statistically significant; it bears considerable economic value for investors. The residuals from ML forecasts—representing information not captured by existing models—show a robust predictive relationship with future stock returns. This finding indicates that the market has not fully priced in the information revealed by ML models, thus allowing investors using ML-derived insights to potentially achieve superior returns.

Additionally, the comparison between ML forecasts and consensus analyst forecasts reveals noteworthy findings: (i) ML forecasts are as accurate as consensus analyst forecasts over a one-year forecast horizon and are significantly more accurate than them over longer forecast horizons and (ii) ML forecasts contain significant incremental information beyond analyst consensus forecasts even if analysts have access to all the financial statements used in ML models (and much more), suggesting that analysts fail to fully incorporate the information in key financial statement items into their forecasts.

The overarching conclusion is that ML technology holds considerable promise in refining investment decision-making processes. By more effectively extracting and utilizing value-relevant information from financial statements, ML can play a transformative role in enhancing the accuracy and efficacy of earnings forecasts and, by extension, in the broader domain of financial analysis and investment.

Appendix 1: Models and Variable Definitions

A: Summary of extant models and variable definitions

Extant Models

RW: $E_{i,t+1} = E_{i,t} + \varepsilon_{i,t+1}$

AR: $E_{i,t+1} = \alpha_0 + \alpha_1 E_{i,t} + \varepsilon_{i,t+1}$

HVZ: $E_{i,t+1} = \alpha_0 + \alpha_1 A_{i,t} + \alpha_2 D_{i,t} + \alpha_3 DD_{i,t} + \alpha_4 E_{i,t} + \alpha_5 NegE_{i,t} + \alpha_6 AC_{i,t} + \varepsilon_{i,t+1}$

SO: $EPS_{i,t+1} = \alpha_0 + \alpha_1 EPS^{+}_{i,t} + \alpha_2 NegE_{i,t} + \alpha_3 AC^{-}_{i,t} + \alpha_4 AC^{+}_{i,t} + \alpha_5 AG_{i,t} + \alpha_6 NDD_{i,t}$
$+\alpha_7 DIV_{i,t} + \alpha_8 BTM_{i,t} + \alpha_9 Price_{i,t} + \varepsilon_{i,t+1}$

EP: $E_{i,t+1} = \alpha_0 + \alpha_1 NegE_{i,t} + \alpha_2 E_{i,t} + \alpha_3 NegE_{i,t} * E_{i,t} + \varepsilon_{i,t+1}$

RI: $E_{i,t+1} = \alpha_0 + \alpha_1 NegE_{i,t} + \alpha_2 E_{i,t} + \alpha_3 NegE_{i,t} * E_{i,t} + \alpha_4 BVE_{i,t} + \alpha_5 TACC_{i,t} + \varepsilon_{i,t+1}$

Variable	Definition
E_{t+1}	Earnings (ib – spi) in year $t+1$
EPS_{t+1}	Earnings (ib – spi) in year $t+1$ scaled by shares outstanding (csho)
E_t	Earnings (ib – spi) in year t
A_t	Total assets (at)
D_t	Dividend payment (dvc)
DD_t	Dummy variable indicating dividend payment
$NegE_t$	Dummy variable indicating negative earnings
AC_t	Accruals calculated as the change in non-cash current assets (act – che) minus the change in current liabilities, excluding short-term debt and taxes payable (lct – dlc – txp) minus depreciation and amortization (dp)
EPS^{+}_t	Earnings per share when earnings are positive and zero otherwise
AC^{-}_t	Accruals per share when accruals are negative and zero otherwise
AC^{+}_t	Accruals per share when accruals are positive and zero otherwise
AG_t	Percentage change in total assets
NDD_t	Dummy variable indicating a zero dividend per share
DIV_t	Dividend per share (dvpsx_f)
BTM_t	Book-to-market ratio, calculated as the book value of equity divided by the market equity as of three months after the end of the last fiscal year
$Price_t$	Stock price as of three months after the end of fiscal year t
BVE_t	Book value of equity (ceq)
$TACC_t$	Total accruals defined in Richardson et al. (2005), which is the sum of the change in WC (i.e., (act – che) − (lct – dlc)), change in NCO (i.e., (at – act – ivao) − (lt – lct – dltt)), and change in FIN (i.e., (ivst + ivao) − (dltt + dlc + pstk))

B: Rationales and variables (or predictors) for machine learning models

Variables	Definition
Category I: Historical earnings and their major components	
Rationale: The random walk model is difficult to beat. It suggests that historical earnings are one of the most important determinants (Monahan 2018; Gerakos and Gramacy 2013). The disaggregation of earnings provides additional information about future fundamentals and improves future earnings forecasts (Fairfield, Sweeney, and Yohn 1996; Chen, Miao, and Shevlin 2015). According to the balancing model of Compustat, earnings are computed as: $E = SALE\text{-}COGS\text{-}XSGA\text{-}DP\text{-}XINT + NOPI\text{-}TXT\text{-}MII$. We include all the components except *MII* to avoid perfect multicollinearity.	
E_t	Earnings (ib – spi)
$SALE_t$	Sales (sale)
$COGS_t$	Cost of goods sold (cogs)
$XSGA_t$	Selling, general, and administrative expenses (xsga)
DP_t	Depreciation and amortization (dp)

continued

Appendix 1: Models and Variable Definitions (continued)

Variables	Definition
$XINT_t$	Interest and related expense (xint)
$NOPI_t$	Non-operating income (expense) (nopi)
TXT_t	Income taxes (txt)
Category II: Other important individual items on the income statement *Rationale:* We also include several individual items on income statements that prior literature has shown to have different/important implications with regard to future earnings: (i) advertising expense (*XAD*) and research and development expense (*XRD*) may generate long-term future benefits (e.g., Bublitz and Ettredge 1989; Sougiannis 1994; Lev and Sougiannis 1996; Chan, Lakonishok, and Sougiannis 2001; Vitorino 2014; Dou et al. 2021); (ii) firms may shift core expenses to special items (*SPI*), extraordinary items, and discontinued operations (*XIDO*) to manipulate/smooth core earnings (e.g., Barnea, Ronen, and Sadan 1976; McVay 2006; Barua, Lin, and Sbaraglia 2010; Kaplan, Kenchington, and Wenzel 2020); and (iii) a common dividend (*DVC*) signals a firm's future earning power (e.g., Nissim and Ziv 2001).	
XAD_t	Advertising expense (xad)
XRD_t	Research and development (R&D) expense (xrd)
SPI_t	Special items (spi)
$XIDO_t$	Extraordinary items and discontinued operations (xido)
DVC_t	Common dividend (dvc)
Category III: Summary and individual accounts on the balance sheet *Rationale:* A balance sheet summarizes the resources that have potential future economic benefit and may contain incrementally useful information regarding future earnings. For example, the literature shows that the book value of equity is one of the most important drivers of future earnings (e.g., Feltham and Ohlson 1995; Li and Mohanram 2014). Furthermore, Ball et al. (2020) contend that retained earnings may measure average earning power better than the book value of equity. Finally, the balance sheet is an earnings management constraint that may affect the reversal of accruals (e.g., Baber, Kang, and Li 2011). Thus, we include the summary balance sheet accounts of AT, ACT, LCT, LT, and CEQ as well as other important individual balance sheet items: CHE, INVT, RECT, PPENT, IVO, INTAN, AP, DLC, TXP, DLTT, and RE. We again omit some summary and individual accounts to avoid perfect multicollinearity.	
AT_t	Total assets (at)
ACT_t	Total current assets (act)
LCT_t	Total current liabilities (lct)
LT_t	Total liabilities (lt)
CEQ_t	Common/ordinary equity (ceq)
CHE_t	Cash and short-term investments (che)
$INVT_t$	Inventories (invt)
$RECT_t$	Receivables (rect)
$PPENT_t$	Property, plant, and equipment – net (ppent)
IVA_t	Investment and advances, equity (ivaeq) + investments and advances, other (ivao)
$INTAN_t$	Intangible assets (intan)
AP_t	Accounts payable (ap)
DLC_t	Debt in current liabilities (dlc)
TXP_t	Income taxes payable (txp)
$DLTT_t$	Long-term debt (dltt)
RE_t	Retained earnings (re)
Category IV: Cash flow from operating activities *Rationale:* Cash is king. The cash flow component of earnings is more persistent than the accrual component, and separating them helps predict future earnings (e.g., Sloan 1996; Call et al. 2016). Because cash flow statements were not available until 1989, we only include the cash flow from operating activities (CFO) and estimate it using the indirect approach with balance sheet data when it is missing.	
CFO_t	Cash flow from operating activities (oancf – xidoc); if missing, it is computed using the balance sheet approach (ib – accruals)

continued

Appendix 1: Models and Variable Definitions (continued)

Variables	Definition
Category V: The first-order difference of the above variables *Rationale:* The literature shows that changes in these items often contain incremental information beyond the levels of the financial statement items (e.g., Kothari 1992; Ohlson and Shroff 1992; Richardson et al. 2005).	
$\Delta E_t \sim \Delta CFO_t$	Computed as the corresponding item in year t minus the same item in year $t-1$

C: Variable definitions of return prediction analyses

Dependent variable	
$EXRET12M_{t+1}$	One-year-ahead excess return, computed as the 12-month cumulative return less that of the risk-free rate, starting from the fourth month after the end of the last fiscal year
Main independent variables	
ML_RESD_t	The new information uncovered by machine learning models, estimated as the residual by regressing the one-year-ahead machine learning-based forecasts on the one-year-ahead earnings forecasts from the RW model and the five extant models in year t
Controls	
$Beta_t$	Annual market beta using the market model calculated in WRDS
$SIZE_t$	Logarithm of market capitalization at the end of the third month after the end of the last fiscal year
BM_t	Book-to-market ratio, calculated as the book value of equity divided by the market equity at the end of three months after the end of the last fiscal year
MOM_t	Momentum calculated as the cumulative return during the 11-month period starting 12 months ago
ACC_t	Accruals scaled by total assets
$ST\_Reversal_t$	Return in the prior month
$LT\_Reversal_t$	Return in the prior five years, excluding the prior year
SUE_t	Quarterly unexpected earnings surprise based on a rolling seasonal random walk model (Livnat and Mendenhall 2006)
$Gross\ Profitability_t$	(sale (sale) – cost of goods sold (cogs)) / total assets (at)
$Earnings\ yield_t$	Earnings to price (Penman et al. 2015)
ROE_t	Earnings (ib – spi) divided by common equity (ceq)
INV_t	Growth rate of total assets ($at_t/at_{t-1} - 1$)
IndustryFE	Three-digit SIC industry fixed effects

Appendix 2: Tuning of Hyperparameters for the Machine Learning Models

Model	Candidate values	Algorithms in sklearn
LASSO	alphas = np.linspace(1e-3,1e-1,1000)	*LassoCV(alphas = np.linspace(1e-3,1e-1,1000), fit_intercept = False,max_iter = 25000,n_jobs=-1)*
Ridge	alphas = np.linspace(5e1,1e3,500)	*RidgeCV(alphas = np.linspace(5e1,1e3,500), fit_intercept = False,cv = 5)*
RF	parameters = {'max_features':['auto'], 'max_depth':[20,25,30,35], 'min_samples_leaf':[15,20,25,50]}	*GridSearchCV(RandomForestRegressor(n_estimators = 500, criterion='mse',oob_score = True,n_jobs=-1, random_state = 10), parameters, cv = 5, n_jobs=-1, scoring='neg_mean_squared_error')*
GBR	parameters = {'max_features':['auto'], 'max_depth':[1,3,5], 'min_samples_leaf':[75,100,125,150]}	*GridSearchCV(GradientBoostingRegressor(learning_rate = 0.1, n_estimators = 500,loss='huber',alpha = 0.7, subsample = 0.9,random_state = 10), parameters, cv = 5, n_jobs=-1, scoring='neg_mean_squared_error')*
ANN	parameters = {'activation':['relu','tanh'], 'hidden_layer_sizes':[(64,32,16,8),(32,16,8,4), (16,8,4,2),(64,32,16),(32,16,8),(16,8,4),(8,4,2), (64,32),(32,16),(16,8),(8,4),(4,2),(64,), (32,),(16,),(8,),(4,)],'alpha':[1e-4,1e-5]}	*GridSearchCV(MLPRegressor(max_iter = 1000, random_state = 10,early_stopping = True,tol = 1e-6), parameters, cv = 5, n_jobs=-1, scoring='neg_mean_squared_error')*

Acknowledgments

We thank Utpal Bhattacharya, Sean Cao, Kevin Chen, Shuping Chen, Peter Chen, Zhihong Chen, Patty Dechow, Pingyang Gao, Allen Huang, Mingyi Hung, Steve Monahan, Arthur Morris, Kevin Li, Xinlei Li, Mei Luo, Ting Luo, Scott Murray, Stephen Penman, David Reeb, Richard Sloan, Ayung Tseng, Ranik Raaen Wahlstrøm, Kun Wang, Rongfei Wang, Zheng Wang, Jian Xue, Amy Zang, Guochang Zhang, Yue Zheng, and participants at workshops at HKUST, Tsinghua University, the Accounting Design Project of Columbia University, the 4th Shanghai Edinburgh Fintech Conference, the 2021 Winter Research Conference on Machine Learning and Business at University of Miami, the 4th Annual CFA NY/SQA Data Science in Finance Conference and the 3rd Future of Financial Information Conference for their helpful comments and suggestions.

Editor's Note

Submitted 28 September 2022

Accepted 30 January 2024 by William N. Goetzmann

Notes

1. For example, the law of diminishing returns predicts a nonlinear relationship between capital investment and future earnings. Prior literature has also shown that the relationship between current and future earnings is nonlinear and varies across firms (e.g., Lev 1983; Freeman and Tse 1992; Baginski et al. 1999; Chen and Zhang 2007).

2. The financial statement items include (i) earnings and their major components; (ii) other income statement items that prior studies have shown to produce long-term benefits or be associated with earnings manipulation; (iii) balance sheet accounts that are important for earnings prediction; and (iv) operating cash flows, which will be discussed in greater detail in Section 3.

3. Subsequent research finds that the RW model performs well even when compared with analyst forecasts. For example, Bradshaw et al. (2012) find that analysts' earnings forecasts are not economically more accurate than the naïve RW forecasts, and those for longer horizons are even less accurate than the naïve RW forecasts.

4. There was also a flurry of early studies on fundamental analysis using financial ratios (see reviews by Kothari [2001] and Richardson et al. [2010]). However, most of these studies examine in-sample associations and are subject to *concerns of the in-sample identification of predictive variables* (p. 424 of Richardson et al. 2010). They provide little (if any) evidence on the accuracy or informativeness of out-of-

sample forecasts. Furthermore, these studies focus on the sample period before the 1990s. It is unclear whether their conclusions still hold in recent years.

5. We limit the input variables to the financial statement data because we are interested in the decision usefulness of fundamental analysis using financial statement information. Furthermore, we compare the ML models with the extant models, most of which also only use financial statement items (with the exception of the SO model).

6. These ML algorithms have been widely adopted in the literature. For the sake of brevity, we do not discuss the technical details about them. Interested readers can refer to Gu et al. (2020) for these details. We select these models rather than other more complex ones, which incorporate time-series dynamics such as Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), or reinforcement learning because all benchmark extant models are cross-sectional. To the extent that more sophisticated models may be able to produce better forecasts and that other inputs such as industry and macroeconomic variables might contain additional information, our results should be interpreted as a lower bound that ML can attain in earnings forecasting tasks.

7. Ideally, we would like to use the point-in-time data to avoid any subsequent restatements in accounting numbers. However, we do not have access to the data. We therefore follow related studies (e.g., Hou et al. 2012; So 2013; Li and Mohanram 2014) and use the Compustat fundamental annual file.

8. For each round of training, cross-validation (CV) and prediction, we normalize all independent variables to [0,1] for the training set. We also transform the independent variables in the prediction set using the same scaler obtained from the training set. Furthermore, to alleviate the influence of outliers, we winsorize the input features (i.e., per-share accounting numbers) at the top and bottom 1% for the training sample.

9. We also conduct robustness tests and compare the forecast accuracy for earnings scaled by total assets and shares outstanding (i.e., EPS), and the inference remains the same. Furthermore, we partition the sample into subgroups based on the market cap, and the results still show that the ML forecast is significantly more accurate than the RW and extant model forecasts in all subgroups. The results are untabulated but available upon request.

10. Note that this is a test of two joint hypotheses: (i) the new information uncovered by the ML model is economically significant and (ii) the market does not fully understand the new information. This test also sheds light on the potential decision usefulness of fundamental analysis and financial information. As Fama (1965) suggests, fundamental analysis is only of value if it provides *new* information not yet fully priced. Thus, this analysis also sheds light on whether fundamental analysis is useful for investors.

11. It is possible that the results might not be robust if we control for a large number of other characteristics that the literature has discovered. Similarly, although we have controlled for all factors in the Fama–French database in the subsequent time-series test, the risk-adjusted returns may not be significant if we control for other additional factors that are theoretically or empirically related to the new predictor. We thank one of the referees for pointing this out.

12. The left vertical axis is for the cumulative log returns to the five quintile portfolios, while the right vertical axis is for the cumulative log return to the hedge portfolio. The two axes have different scales. It is worth pointing out that the hedge portfolio has a lower volatility than the highest quintile. The annualized volatility of the two portfolios is 9.5% and 16.3%, respectively.

13. The plots for the GBR models are similar but untabulated for the sake of brevity.

14. The nonlinear models also accommodate the interaction effects between predictors. Untabulated analyses show that the interaction effects between the following economically related pairs contribute the most to the explanatory power of the GBR model: the change in sales revenue and change in the cost of goods sold, change in short-term debts and change in total current liabilities, sales revenue and accounts payable, depreciation and amortization expense and net property, plant and equipment, and cost of goods sold and inventories. The finding that the ML models pick up the interaction between income statement items and the corresponding gross accrual items (e.g., depreciation and amortization (DA) and net PP&E, COGS and inventories) resonates remarkably well with the recent call for research by Dichev (2020) on the role of gross accruals in determining earnings quality.

15. In untabulated analyses, we find that ML forecasts are also a useful benchmark to assess the ex-ante bias in analyst forecasts. When analysts' forecasts are substantially higher (lower) than the ML forecasts, these forecasts tend to be overly optimistic (pessimistic) than the actual earnings. Furthermore, these stocks tend to have significantly negative (positive) returns over subsequent periods.

16. The literature has yet to arrive at a consensus on the level of trading costs. While Novy-Marx and Velikov (2016) suggest that "[r]ound-trip transaction costs for typical value-weighted strategies average in excess of 50 basis points (bps)," Frazzini et al. (2018) argue that this estimate is "an order of magnitude larger than [what] our model or live costs suggest." They suggest that for a large institutional trader, "the effective bid-ask spread across all trades averages less than 0.015% per year," and the market impact is just under nine basis points on average for all trades completed within a day. The impact of the transaction cost would be much lower if we use Frazzini et al.'s estimate.

References

Apley, D. W., and J. Zhu. 2020. "Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82 (4): 1059–1086. doi:10.1111/rssb.12377.

Baber, W. R., S. H. Kang, and Y. Li. 2011. "Modeling Discretionary Accrual Reversal and the Balance Sheet as an Earnings Management Constraint." *The Accounting Review* 86 (4): 1189–1212. doi:10.2308/accr-10037.

Baginski, S. P., K. S. Lorek, G. L. Willinger, and B. C. Branson. 1999. "The Relationship between Economic Characteristics and Alternative Annual Earnings Persistence Measures." *The Accounting Review* 74 (1): 105–120. doi:10.2308/accr.1999.74.1.105.

Ball, R., J. Gerakos, J. T. Linnainmaa, and V. Nikolaev. 2020. "Earnings, Retained Earnings, and Book-to-Market in the Cross Section of Expected Returns." *Journal of Financial Economics* 135 (1): 231–254. doi:10.1016/j.jfineco.2019.05.013.

Ball, R., and R. Watts. 1972. "Some Time Series Properties of Accounting Income." *The Journal of Finance* 27 (3): 663–681. doi:10.1111/j.1540-6261.1972.tb00991.x.

Bao, Y., B. Ke, B. Li, Y. J. Yu, and J. Zhang. 2020. "Detecting Accounting Fraud in Publicly Traded US Firms Using a Machine Learning Approach." *Journal of Accounting Research* 58 (1): 199–235. doi:10.1111/1475-679X.12292.

Barnea, A., J. Ronen, and S. Sadan. 1976. "Classificatory Smoothing of Income with Extraordinary Items." *The Accounting Review* 51 (1): 110–122.

Bartram, S. M., and M. Grinblatt. 2018. "Agnostic Fundamental Analysis Works." *Journal of Financial Economics* 128 (1): 125–147. doi:10.1016/j.jfineco.2016.11.008.

Barua, A., S. Lin, and A. M. Sbaraglia. 2010. "Earnings Management Using Discontinued Operations." *The Accounting Review* 85 (5): 1485–1509. doi:10.2308/accr.2010.85.5.1485.

Bertomeu, J., E. Cheynel, E. Floyd, and W. Pan. 2021. "Using Machine Learning to Detect Misstatements." *Review of Accounting Studies* 26 (2): 468–519. doi:10.1007/s11142-020-09563-8.

Bradshaw, M. T., M. S. Drake, J. N. Myers, and L. A. Myers. 2012. "A Re-Examination of Analysts' Superiority over Time-Series Forecasts of Annual Earnings." *Review of Accounting Studies* 17 (4): 944–968. doi:10.1007/s11142-012-9185-8.

Brown, L. D. 1993. "Earnings Forecasting Research: Its Implications for Capital Markets Research." *International Journal of Forecasting* 9 (3): 295–320. doi:10.1016/0169-2070(93)90023-G.

Bublitz, B., and M. Ettredge. 1989. "The Information in Discretionary Outlays: Advertising, Research, and Development." *The Accounting Review* 64 (1): 108–124.

Cao, S., W. Jiang, J. L. Wang, and B. Yang. 2024. "From Man vs. Machine to Man + Machine: The Art and AI of Stock Analyses." *Journal of Financial Economics*. Forthcoming.

Call, A. C., M. Hewitt, T. Shevlin, and T. L. Yohn. 2016. "Firm-Specific Estimates of Differential Persistence and Their Incremental Usefulness for Forecasting and Valuation." *The Accounting Review* 91 (3): 811–833. doi:10.2308/accr-51233.

Chan, L. K., J. Lakonishok, and T. Sougiannis. 2001. "The Stock Market Valuation of Research and Development Expenditures." *The Journal of Finance* 56 (6): 2431–2456. doi:10.1111/0022-1082.00411.

Chen, P. F., and G. Zhang. 2007. "How Do Accounting Variables Explain Stock Price Movements? Theory and Evidence." *Journal of Accounting and Economics* 43 (2-3): 219–244. doi:10.1016/j.jacceco.2007.01.001.

Chen, S., B. Miao, and T. Shevlin. 2015. "A New Measure of Disclosure Quality: The Level of Disaggregation of Accounting Data in Annual Reports." *Journal of Accounting Research* 53 (5): 1017–1054. doi:10.1111/1475-679X.12094.

Chen, X., Y. H. 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. doi:10.1111/1475-679X.12429.

Dichev, I. D. 2020. "Fifty Years of Capital Markets Research in Accounting: Achievements so Far and Opportunities Ahead." *China Journal of Accounting Research* 13 (3): 237–249. doi:10.1016/j.cjar.2020.07.005.

Ding, K., B. Lev, X. Peng, T. Sun, and M. A. Vasarhelyi. 2020. "Machine Learning Improves Accounting Estimates: Evidence from Insurance Payments." *Review of Accounting Studies* 25 (3): 1098–1134. doi:10.1007/s11142-020-09546-9.

Dou, W. W., Y. Ji, D. Reibstein, and W. Wu. 2021. "Inalienable Customer Capital, Corporate Liquidity, and Stock Returns." *The Journal of Finance* 76 (1): 211–265. doi:10.1111/jofi.12960.

Easton, P. D., P. Kelly, and A. Neuhierl. 2018. "Beating a Random Walk." Available at SSRN: https://ssrn.com/abstract=3040354 or doi:10.2139/ssrn.3040354.

Fairfield, P. M., R. J. Sweeney, and T. L. Yohn. 1996. "Accounting Classification and the Predictive Content of Earnings." *The Accounting Review* 71 (3): 337–355.

Fama, E. F. 1965. "Random Walks in Stock Market Prices." *Financial Analysts Journal* 21 (5): 55–59. doi:10.2469/faj.v21.n5.55.

Fama, E. F., and K. R. French. 2000. "Forecasting Profitability and Earnings." *The Journal of Business* 73 (2): 161–175. doi:10.1086/209638.

Fama, E. F., and K. R. French. 2006. "Profitability, Investment and Average Returns." *Journal of Financial Economics* 82 (3): 491–518. doi:10.1016/j.jfineco.2005.09.009.

Fama, E. F., and K. R. French. 2015. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics* 116 (1): 1–22. doi:10.1016/j.jfineco.2014.10.010.

Feltham, G. A., and J. A. Ohlson. 1995. "Valuation and Clean Surplus Accounting for Operating and Financial Activities." *Contemporary Accounting Research* 11 (2): 689–731. doi:10.1111/j.1911-3846.1995.tb00462.x.

Feltham, G. A., and J. A. Ohlson. 1996. "Uncertainty Resolution and the Theory of Depreciation Measurement." *Journal of Accounting Research* 34 (2): 209–234. doi:10.2307/2491500.

Frazzini, A., R. Israel, and T. J. Moskowitz. 2018. "Trading Costs." Available at SSRN 3229719.

Freeman, R. N., and S. Y. Tse. 1992. "A Nonlinear Model of Security Price Responses to Unexpected Earnings." *Journal of Accounting Research* 30 (2): 185–209. doi:10.2307/2491123.

Gerakos, J. J., and R. Gramacy. 2013. "Regression-Based Earnings Forecasts." Chicago Booth Research Paper, 12–26.

Available at SSRN: https://ssrn.com/abstract=2112137 or doi:10.2139/ssrn.2112137.

Gu, S., B. Kelly, and D. Xiu. 2020. "Empirical Asset Pricing via Machine Learning." *The Review of Financial Studies* 33 (5): 2223–2273. doi:10.1093/rfs/hhaa009.

Hanlon, M. 2005. "The Persistence and Pricing of Earnings, Accruals, and Cash Flows When Firms Have Large Book-Tax Differences." *The Accounting Review* 80 (1): 137–166. doi:10.2308/accr.2005.80.1.137.

Hanlon, M., S. Laplante, and T. Shevlin. 2005. "Evidence for the Possible Information Loss of Conforming Book Income and Taxable Income." *The Journal of Law and Economics* 48 (2): 407–442. doi:10.1086/497525.

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. doi:10.1016/j.jacceco.2011.12.001.

Kaplan, S. E., D. G. Kenchington, and B. S. Wenzel. 2020. "The Valuation of Discontinued Operations and Its Effect on Classification Shifting." *The Accounting Review* 95 (4): 291–311. doi:10.2308/tar-2016-0235.

Kothari, S. P. 1992. "Price-Earnings Regressions in the Presence of Prices Leading Earnings: Earnings Level versus Change Specifications and Alternative Deflators." *Journal of Accounting and Economics* 15 (2-3): 173–202. doi:10.1016/0165-4101(92)90017-V.

Kothari, S. P. 2001. "Capital Markets Research in Accounting." *Journal of Accounting and Economics* 31 (1-3): 105–231. doi:10.1016/S0165-4101(01)00030-1.

Lee, C. M. 1999. "Accounting-Based Valuation: Impact on Business Practices and Research." *Accounting Horizons* 13 (4): 413–425. doi:10.2308/acch.1999.13.4.413.

Lev, B. 1983. "Some Economic Determinants of Time-Series Properties of Earnings." *Journal of Accounting and Economics* 5: 31–48. doi:10.1016/0165-4101(83)90004-6.

Lev, B., and D. Nissim. 2004. "Taxable Income, Future Earnings, and Equity Values." *The Accounting Review* 79 (4): 1039–1074. doi:10.2308/accr.2004.79.4.1039.

Lev, B., and T. Sougiannis. 1996. "The Capitalization, Amortization, and Value-Relevance of R&D." *Journal of Accounting and Economics* 21 (1): 107–138. doi:10.1016/0165-4101(95)00410-6.

Li, K. K., and P. Mohanram. 2014. "Evaluating Cross-Sectional Forecasting Models for Implied Cost of Capital." *Review of Accounting Studies* 19 (3): 1152–1185. doi:10.1007/s11142-014-9282-y.

Livnat, J., and R. R. Mendenhall. 2006. "Comparing the Post–Earnings Announcement Drift for Surprises Calculated from Analyst and Time Series Forecasts." *Journal of Accounting Research* 44 (1): 177–205. doi:10.1111/j.1475-679X.2006.00196.x.

McVay, S. E. 2006. "Earnings Management Using Classification Shifting: An Examination of Core Earnings and Special Items." *The Accounting Review* 81 (3): 501–531. doi:10.2308/accr.2006.81.3.501.

Monahan, S. J. 2018. "Financial Statement Analysis and Earnings Forecasting." *Foundations and Trends® in Accounting* 12 (2): 105–215. doi:10.1561/1400000036.

Nissim, D., and A. Ziv. 2001. "Dividend Changes and Future Profitability." *The Journal of Finance* 56 (6): 2111–2133. doi:10.1111/0022-1082.00400.

Novy-Marx, R., and M. Velikov. 2016. "A Taxonomy of Anomalies and Their Trading Costs." *Review of Financial Studies* 29 (1): 104–147. doi:10.1093/rfs/hhv063.

Ohlson, J. A. 1995. "Earnings, Book Values, and Dividends in Equity Valuation." *Contemporary Accounting Research* 11 (2): 661–687. doi:10.1111/j.1911-3846.1995.tb00461.x.

Ohlson, J. A., and B. E. Juettner-Nauroth. 2005. "Expected EPS and EPS Growth as Determinants of Value." *Review of Accounting Studies* 10 (2-3): 349–365. doi:10.1007/s11142-005-1535-3.

Ohlson, J. A., and P. K. Shroff. 1992. "Changes versus Levels in Earnings as Explanatory Variables for Returns: Some Theoretical Considerations." *Journal of Accounting Research* 30 (2): 210–226. doi:10.2307/2491124.

Penman, S. H. 1992. "Return to Fundamentals." *Journal of Accounting, Auditing & Finance* 7 (4): 465–483. doi:10.1177/0148558X9200700403.

Penman, S. H., F. Reggiani, S. A. Richardson, and A. Tuna. 2015. "An Accounting-Based Characteristic Model for Asset Pricing." Available at SSRN 1966566.

Rasekhschaffe, K. C., and R. C. Jones. 2019. "Machine Learning for Stock Selection." *Financial Analysts Journal* 75 (3): 70–88. doi:10.1080/0015198X.2019.1596678.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and I. Tuna. 2005. "Accrual Reliability, Earnings Persistence and Stock Prices." *Journal of Accounting and Economics* 39 (3): 437–485. doi:10.1016/j.jacceco.2005.04.005.

Richardson, S. A., I. Tuna, and P. Wysocki. 2010. "Accounting Anomalies and Fundamental Analysis: A Review of Recent Research Advances." *Journal of Accounting and Economics* 50 (2-3): 410–454. doi:10.1016/j.jacceco.2010.09.008.

Sivaramakrishnan, K., M. Brown, and B. Kasturi. 2018. "Smart Beta: Even Smarter with an Optimizer and a Custom Risk Model." Axioma Research Paper 126.

Sloan, R. G. 1996. "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?" *The Accounting Review* 71 (3): 289–315.

So, E. C. 2013. "A New Approach to Predicting Analyst Forecast Errors: Do Investors Overweight Analyst Forecasts?" *Journal of Financial Economics* 108 (3): 615–640. doi:10.1016/j.jfineco.2013.02.002.

Sougiannis, T. 1994. "The Accounting Based Valuation of Corporate R&D." *The Accounting Review* 69 (1): 44–68.

Thomas, J., and F. X. Zhang. 2011. "Tax Expense Momentum." *Journal of Accounting Research* 49 (3): 791–821. doi:10.1111/j.1475-679X.2011.00409.x.

Thomas, J., and F. X. Zhang. 2014. "Valuation of Tax Expense." *Review of Accounting Studies* 19 (4): 1436–1467. doi:10.1007/s11142-013-9274-3.

Vitorino, M. A. 2014. "Understanding the Effect of Advertising on Stock Returns and Firm Value: Theory and Evidence from a Structural Model." *Management Science* 60 (1): 227–245. doi:10.1287/mnsc.2013.1748.

Watts, R. L., and R. W. Leftwich. 1977. "The Time Series of Annual Accounting Earnings." *Journal of Accounting Research* 15 (2): 253–271. doi:10.2307/2490352.