Journal of Accounting and Economics 80 (2025) 101805

Contents lists available at ScienceDirect

Journal of Accounting and Economics

journal homepage: www.journals.elsevier.com/journal-of-accounting-and-economics

Estimating profitability decomposition frameworks via machine learning: Implications for earnings forecasting and financial statement analysis☆

Oliver Binz[a,*], Katherine Schipper[b], Kevin R. Standridge[c]

[a] ESMT Berlin, Germany
[b] Duke University, Fuqua School of Business, USA
[c] University of Utah, David Eccles School of Business, USA

ARTICLE INFO

JEL classification:
C53
G10
M41

Keywords:
Financial statement analysis
Machine learning
Earnings forecasting

ABSTRACT

We find that nonlinear estimation of profitability decomposition frameworks yields more accurate out-of-sample profitability forecasts than forecasts from both a random walk and linear estimation. The improvements derive from nonlinear estimation and synergies between nonlinear estimation and profitability decomposition frameworks. We analyze three essential financial statement analysis design choices to provide insights for the practice of fundamental analysis and find robust evidence that higher levels of profitability decomposition, focusing on core items, and using up to three years of historical information improve forecast accuracy. We find that our forecasts predict returns and profitability changes before and after controlling for analyst forecasts and common asset pricing factors.

1. Introduction

We examine whether and how nonlinear, machine learning estimation of hierarchical nonlinear profitability decomposition frameworks improves profitability forecast accuracy. The hierarchical-decomposition approach to analyzing and predicting profitability is foundational in financial statement analysis and valuation (e.g., Palepu and Healy, 2012; Penman, 2012; Sommers et al., 2021; Wahlen et al., 2018; Yohn, 2020). We use machine learning to estimate Nissim and Penman's (2001, 2003) (hereafter NP) profitability decomposition framework, use the estimates to forecast profitability out of sample, benchmark the forecasts against those obtained from a random walk and linear estimation, and examine whether the forecasts embody information investors and analysts could use to improve their trading and profitability forecasting. To provide insights for the practice of financial statement analysis, we

☆ We thank Dan Bens, Qi Chen, Wilbur Chen (discussant), Stephanie Cheng, Scott Dyreng, Mory Elsaifv, John Graham, Luzi Hail, Michelle Hanlon, Robert Hills, Xu Jiang, Peter Joos, Thomas Keusch, Matt Kubic, Wayne Landsman, Charles Lee, Bill Mayew, Suresh Nallareddy, Gans Narayanamoorthy, Stephen Penman, James Pinnington, Peter Pope, Ayung Tseng, Teri Yohn, Haifeng You (discussant), two anonymous referees, and workshop participants at Duke University, INSEAD, Tulane University, the Institute for Quantitative Investment Research (INQUIRE), the 2018 AAA/Deloitte/J. Michael Cook Doctoral Consortium, the 2019 EAA Annual Meeting, the 2021 Hawaii Accounting Research Conference, the Accounting Design Project and the 2021 University of Miami Winter Research Conference on Machine Learning and Business for helpful comments and suggestions. A previous version of this paper was titled "What can Analysts Learn from Artificial Intelligence about Fundamental Analysis?"

* Corresponding author.
E-mail addresses: oliver.binz@esmt.org (O. Binz), katherine.schipper@duke.edu (K. Schipper), kevin.standridge@duke.edu (K.R. Standridge).

https://doi.org/10.1016/j.jacceco.2025.101805
Received 8 February 2023; Received in revised form 28 May 2025; Accepted 16 June 2025
Available online 21 June 2025

analyze several design choices discussed but not analyzed in NP and provide descriptive analyses of how forecast accuracy varies with firm characteristics.

While profitability can be decomposed in many ways, we analyze NP's framework because it is an arithmetically sound (tautological) structure of accounting equalities that disciplines financial statement information to reveal the hierarchy of relations between profitability and its accounting drivers. We apply gradient-boosted-regression-tree estimation, a flexible and popular machine learning algorithm, to find the functional form that uses this information most effectively. A structured approach can outperform, in out-of-sample tests, data mining approaches applied to very large sets of predictors with few or no restrictions beyond data availability (Bertomeu et al., 2025; Liu, 2022). Regardless of whether structured approaches do, or do not, perform out-of-sample as well as or better than data-mining approaches, the advantages of frameworks that incorporate the firm's underlying profitability structure are numerous, including a better understanding of how one should adjust the forecasts for changes in conditions not present in the training dataset (such as a surge in inflation; Nissim, 2024), a reduced risk of overlooking value-relevant information in distorted accounting numbers (Sloan, 2019), and a degree of protection from fluctuations in accounting numbers induced by the reporting process (Penman, 2010, Chapters 4 & 5).[1] Furthermore, a hierarchical profitability decomposition facilitates systematic analysis of the effects of variation in design choices, specifically, the granularity of profitability driver disaggregation, the treatment of transitory and non-core items, and the amount of historical information to use.

We first confirm many of NP's findings for their sample period (1963–1999) for our longer sample period (1963–2023). Because the relations in NP's framework are nonlinear and because both NP's analyses and ours show that past ratios are nonlinearly and interactively linked to future profitability, estimating NP's framework necessitates the use of methods that accommodate complex nonlinear associations. We use gradient-boosted regression trees, a machine learning algorithm that approximates nonlinear and interactive relations among variables, to estimate the nonlinear relations of NP's framework and forecast future profitability. Building on Gerakos and Gramacy (2013) and Li and Mohanram (2014), we benchmark one-year-ahead out-of-sample profitability forecasts against forecasts from a random walk and linear (OLS) estimation, and generally find the machine-learning-based predictions are more accurate. The improvements largely derive from the more extreme portions of the absolute forecast error distribution, that is, from observations for which forecasting is more difficult.

To shed light on whether our results derive from nonlinear estimation, from NP's framework, or from synergies between them, we proceed in two steps. First, we benchmark one-year-ahead profitability forecasts obtained from linear and nonlinear estimation of an autoregressive model against random walk forecasts. Linear estimation of the autoregressive model does not produce significantly lower forecast errors than the random walk. However, relative to a random walk, nonlinear estimation lowers absolute forecast errors by 6.977%, suggesting nonlinear estimation is important for profitability forecasting. This improvement, large at face value, appears even more substantial once one recognizes that forecast errors of the benchmark models largely reflect random (hence unpredictable) variation. Thus, the model improvements we document constitute an even larger share of the (unobservable) variation in the predictable portion of the benchmark models' forecast errors (i.e., the portion of the forecast errors that would be predictable if one had access to the true model connecting future profitability to all information available at the time the forecast is made). This finding is also significant given Gerakos and Gramacy's (2013) and Li and Mohanram's (2014) finding that a simple random walk outperforms linear models including multiple predictors and Monahan's (2018) related analysis.[2]

Second, we test whether analyzing NP's framework, as opposed to exploiting profitability's autoregressive process, further improves performance. Consistent with NP's (p. 168) untabulated findings and inference that "linear [estimation of their framework is] not likely to work well," we find that using the ratios identified by NP's framework as predictors within a linear model (i.e., linear estimation of NP's framework) does not improve forecasting performance; in fact, performance deteriorates. We also find that using nonlinear estimation (that is, using the ratios identified by NP's framework as predictors in a nonlinear model) instead of linear estimation of NP's framework significantly improves forecast accuracy and that this improvement is monotonically increasing in the granularity of the disaggregation. The economic magnitude of these improvements is large relative to those documented in prior research. Relative to nonlinear estimation of the autoregressive model, nonlinear estimation of NP's framework decreases absolute forecast errors by 3.392%, suggesting meaningful synergies between nonlinear estimation and the use of profitability decomposition frameworks.

Our inferences are robust to employing alternative machine learning algorithms. Further, including firms' industry membership, variables capturing the state of the macroeconomy, or firm-level profitability predictors proposed in prior research as alternative or additional predictors does not improve forecast accuracy.

Extending the profitability decomposition results, which shed light on an essential financial statement design choice (the level of disaggregation of profitability drivers), we examine how two other essential design choices that must be made on empirical grounds in applying NP's framework affect forecast accuracy. Specifically, we find that (1) focusing on core items (i.e., items more likely to recur)

[1] For example, recognizing a current-period impairment charge reduces the firm's depreciable asset base and thereby future depreciation expense, which systematically increases future profitability. An understanding of the firm's profitability structure allows users to appreciate these relationships and thereby provides protection from certain unpleasant surprises in profitability realizations.

[2] Monahan (2018) describes the finding that a simple random walk outperforms linear models including multiple predictors as "a provocative result because it leads to the seemingly absurd conclusion that, within the context of forecasting earnings, there is no value to peer analysis, trend analysis and using conditioning information" (p. 146). Further, "the random-walk model is inconsistent with standard economic assumptions, accounting practice and the manner in which financial statement analysis is practiced and taught … if the random-walk model is the best academics can do, the relevance of the entire literature on forecasting and financial statement analysis is called into question" (p. 205).

and (2) using more historical information (up to three years) improves forecast accuracy.

In our final tests, we examine whether forecasts from nonlinear estimation of profitability decomposition frameworks could be helpful to investors and analysts. Specifically, after controlling for the asset pricing factors in Lee et al. (2024) and the consensus analyst forecast, we find that a 1% increase in the forecast obtained from nonlinear estimation of NP's framework is associated with a 0.247% increase in year-ahead stock returns and a 1.045% change in profitability, suggesting that investors and analysts could use the approach outlined in this paper to enhance trading profits and profitability forecast accuracy.

Our research contributes to two literatures. The first analyzes structural profitability decomposition frameworks, in particular, NP's. We extend this literature in two distinct and related ways. First, although NP both argue and demonstrate that evaluating their model requires accommodating its essential nonlinear structure, prior research, for example, Fairfield and Yohn (2001), Soliman (2008), and Esplin et al. (2014) uses OLS to approximate the nonlinearities in NP's model, likely because of technological constraints that we relax by using machine learning. We find that accommodating nonlinearities improves forecast accuracy. These machine-learning-derived improvements are large not only relative to those obtained from estimating NP's model using OLS in our sample but also relative to results reported in Fairfield and Yohn (2001) and Esplin et al. (2014) in their original samples.[3] These results highlight the importance of accommodating nonlinearities in profitability decomposition frameworks. Second, addressing practical concerns, we provide evidence that focusing on core (i.e., likely recurring) items and using historical information improve forecast accuracy.

The second literature to which we contribute tests whether machine learning techniques can be used to increase earnings forecast accuracy, and if so, how best to exploit these techniques in varying contexts. While Callen et al. (1996) fail to find evidence that machine learning improves firm-level time-series earnings forecasting models,[4] more recent research, for example, Gerakos and Gramacy (2013), Anand et al. (2020), Chen et al. (2022), Hunt et al. (2022), Cao and You (2023), and van Binsbergen et al. (2023), documents improvements for panel data models. These papers do not, by design, estimate a profitability decomposition framework such as NP's and instead take a purely statistical approach to predictor selection, an approach NP describe as "trawling through the data without structure" (p. 125). These papers exploit the strengths of machine learning for processing high-dimensionality data sets and arbitrary nonlinearities among very large predictor sets while applying regularization to mitigate overfitting.

From the perspective of providing insights for practical applications, a statistical approach has a key weakness that our approach, grounded in NP's structured profitability decomposition framework, addresses. Specifically, in accounting settings, where double-entry bookkeeping creates collinear variables, unstructured data-mining approaches can produce good predictive performance at the cost of counterintuitive and even uninterpretable results (e.g., Bertomeu, 2020). While predictive power is in and of itself highly desirable in practice, teaching and future research must rest on an understanding of the variables that drive the predictive power.[5] A statistical approach to understanding these effects evaluates how removing one variable at a time affects prediction accuracy (e.g., Chen et al., 2022). However, the results of removing one variable at a time from a set of collinear predictor variables are often uninterpretable.[6] Because NP's hierarchical decomposition creates well-defined reference groups based on the level of disaggregation, our approach to iterative variable removal does not suffer from this interpretability problem.[7]

Our paper proceeds as follows. Section 2 discusses profitability decomposition frameworks and gradient-boosted regression trees. Section 3 describes the data and provides evidence on nonlinear, interactive relations between future profitability and past profitability drivers identified in NP's profitability decomposition framework. Section 4 discusses our main results that nonlinear estimation and NP's framework jointly improve forecast accuracy. Sections 5 and 6 summarize robustness tests and additional analyses. Section 7 concludes. Appendix A contains variable definitions. Online Appendices O1 to O3 discuss additional analyses and the hyperparameter choices that underlie our gradient-boosted regression tree algorithms.

2. Research design

2.1. Profitability decomposition frameworks and earnings forecasting

Profitability decomposition frameworks disaggregate accounting profitability into a set of ratios to inform users about the factors driving economic performance. The first known profitability decomposition framework was introduced in the early 20th century at the Dupont Powder Company by its CFO F. Donaldson Brown to decompose return on assets into profit margins and turnovers (Johnson and Kaplan, 1987, pp. 10–12). Subsequent authors extended the original Dupont framework to derive additional insights about firms' operations and used those insights to forecast profitability. We focus on NP's version of the profitability decomposition framework

[3] Soliman (2008) does not report out-of-sample forecast errors.

[4] Callen et al. (1996) analyze a small sample of 296 New York Stock Exchange firms, likely limiting the generalizability of their inferences.

[5] We do not aim to shed light on whether the approach we take, applying nonlinear machine learning estimation to a structured profitability decomposition framework, does or does not have greater (pure) predictive power than a statistical approach such as that in Chen et al. (2022).

[6] For example, predictions will hardly change if common equity is dropped from a predictor set that also includes assets and liabilities. While it is easy to see the problem in a simple example, combining a statistical approach with a nonlinear algorithm like a gradient-boosted regression tree can make it impossible for a researcher to determine the incremental value of any one variable because the actual degree of collinearity is unknown.

[7] That is, as shown in Fig. 1 and discussed in Section 2.1, variables in the Level 2 disaggregation are incremental to those in Level 1, variables in Level 3 are incremental to those in Level 2, and so on. This clear reference group structure supports sharp insights for practice as to how best to use structural profitability decomposition frameworks.

since it is widely used and, in the words of Monahan (2018, p. 168), has become the authoritative work on the issue.

NP's profitability decomposition framework disaggregates return on common equity (ROCE = CNI/CSE, where CSE denotes book value of shareholders' equity and CNI comprehensive income) into four levels of cumulatively increasing disaggregation (Fig. 1)[8]

Level 1. ROCE = ROTCE × MSR: ROTCE denotes return on total common equity (=(CNI + MII)/(CSE + MI)), MSR minority sharing ratio ($=\frac{CNI/(CNI+MII)}{CSE/(CSE+MI)}$), MII minority (noncontrolling) interest income, and MI minority (noncontrolling) interest.
Level 2. ROTCE = RNOA + FLEV × SPREAD: RNOA denotes return on net operating assets (=OI/NOA), FLEV financial leverage (=NFO/CSE), SPREAD the spread between RNOA and net borrowing cost (= RNOA − NBC), OI operating income, NOA net operating assets (= OA − OL), NFO net financial obligations (= FO − FA), OA operating assets, OL operating liabilities, FO financial obligations, FA financial assets, NBC net borrowing cost (=NFE/NFO), and NFE net financial expense.
Level 3. RNOA = Sales PM × ATO + Other items/NOA: Sales PM denotes sales profit margin (= OI from Sales/Sales) and ATO asset turnover (=Sales/NOA).
Level 4. Sales PM × ATO = Sales PM* × ATO* + OLLEV × OLSPREAD: Sales PM* denotes modified profit margin after considering implicit charges on supplier credit (=(Core OI from Sales + io)/Sales), ATO* modified asset turnover (=Sales/OA), OLLEV operating liability leverage (=OL/NOA), OLSPREAD the spread between return on operating assets and the implicit interest on operating liabilities (=(OI + io)/OA − io/OL), and io the implicit interest charge on operating liabilities.

The analysis reveals nine drivers of ROCE, as shown in Equation (1):

$$\text{ROCE} = \text{MSR} \times \left[\text{Sales PM}^* \times \text{ATO}^* + \frac{\text{Other Items}}{\text{OA}} + \text{OLLEV} \times \text{OLSPREAD} + \text{FLEV} \times (\text{RNOA} - \text{NBC})\right]. \tag{1}$$

Equation (1) illustrates that the predictors in NP's framework are nonlinearly linked to ROCE. Analyses reported by NP and in our Section 3.2 illustrate that past realizations of ratios are nonlinearly and interactively related to future realizations, suggesting the importance of using methods that can accommodate complex nonlinear associations when estimating NP's framework. The next section describes such a tool: gradient-boosted regression trees, a machine learning algorithm.

2.2. *Algorithm selection and forecasting procedure*

To estimate the non-linear relations in the NP framework, we must first identify an estimator that can non-parametrically model the relation between a set of independent variables and a dependent variable. We are aware of three primary candidates: gradient-boosted regression tree (GBRT), random forest, and neural network algorithms.[9] While it is impossible to know ex-ante which of these will have the best performance, prior literature suggests that GBRTs tend to be easier to tune and faster than random forests and neural networks and perform particularly well in tabular data, often outperforming other algorithms (Shwartz-Ziv and Armon, 2022). Therefore, we use the GBRT as our primary algorithm.[10]

We use GBRT estimation to predict out-of-sample ROCE using the ratios from each level of the NP framework as independent variables. To generate an annual forecast, we train the GBRT on ten years of lagged data and generate forecasts for the eleventh year. Because our training data begin in 1964 (i.e., 1963 is the first year in which Compustat is free from survivorship bias and the computation of some ratios identified in NP's framework requires a year of lagged data), the first year for which we generate predictions is 1975. We train the GBRT on data from 1964 to 1973 and use 1974 data to forecast ROCE in 1975. Repeating this process for each 10-year interval, we generate firm-year out-of-sample ROCE forecasts for 1975 to 2023. We select the ideal hyperparameters using a grid search over the number of trees and depth in 1991 as described in the online appendix.[11]

3. Data and descriptive evidence

3.1. *Data*

We obtain annual data for 1963 to 2023 from Compustat. We require all firm-years to have non-missing values for all variables in

[8] Following Penman (2012, p. 147), we use the term "return on common equity" (ROCE) rather than "return on equity" (ROE) to be clear that we are focusing on common shareholders rather than others who might have a stake in the firm's equity.
[9] While machine learning algorithms will yield equivalent estimates to splines or kernels in infinite data, they often perform better in finite data (Michaud et al., 2023; Poulinakis et al., 2023).
[10] Following Easton, Kapons, Kelly, and Neuhierl (2020), we use a mean absolute error loss function.
[11] An anonymous reviewer chose 1991 to prevent researcher biases in selecting the tuning year. Thus, for hyperparameter selection, we use data from 1981 to 1990 to create predictions for ROCE in 1982–1991. We then split these data into an 80% training and a 20% test set. We then choose hyperparameters based on predictive accuracy in the test set.

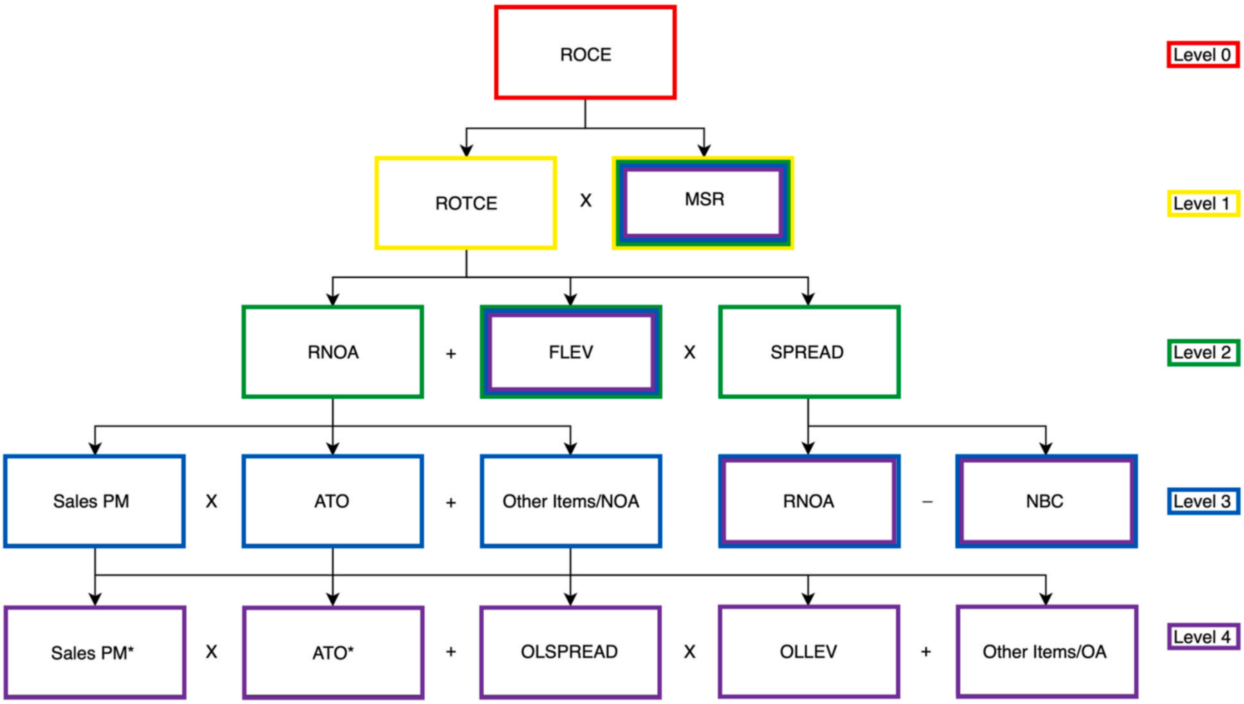

Fig. 1. Nissim and Penman (2001, 2003) Analysis of ROCE This figure depicts the decomposition analysis of profitability in Nissim and Penman (2001, 2003). The ratios are color-coded to illustrate which ratio is included in each level of the decomposition. All variables are defined in Appendix A. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

NP's framework in the current year and for ROCE one year in the future. To ensure sufficient liquidity for the returns analysis, we follow Hou et al. (2020) and require firms to have a market capitalization above the 20th percentile of NYSE firms in a given fiscal year, an SIC code, and assets exceeding \$10 million. Following NP, we require non-negative values for CSE, NOA, OA, MIB, and OL and winsorize all ratios in the NP framework by year at the 1st and 99th percentiles.[12] Table 1 presents descriptive statistics. Table 2 presents correlations for selected ratios, with Pearson (Spearman) correlations below (above) the diagonal.

3.2. *Visual evidence of nonlinearities and interactive relations*

Panels A to H of Fig. 2 present visual evidence of the relation between future profitability and contemporaneous ratios, by plotting median portfolio *ROCE* in period t+1 by *SPREAD* and *ATO*, *ROCE* and *ATO*, *SPREAD* and *OLSPREAD*, and *NBC* and *Sales PM* decile in period t. Except for *ATO*, the plots suggest nonlinear relations. Plotted relative to future *ROCE*, current *ROCE*, *OLSPREAD*, and *RNOA* have an S-shaped association, FLEV and ATO have a U-shaped association, and *SPREAD* and *Sales PM* have a concave association. Panels A to D of Fig. 3 present examples of interactive relations across ratios in predicting future *ROCE*. The surfaces obtained from plotting *FLEV* and *OLSPREAD*, *OLLEV* and *RNOA*, *Sales PM* and *NBC*, and *SPREAD* and *ATO* decile on the X and Y axes and year-ahead *ROCE* on the Z axis exhibit curvatures that are visually different from the straight plane observed under linear, non-interactive relations.

In sum, the visual evidence in Figs. 2 and 3 suggests nonlinearities in the dynamic relations across several fundamental ratios and subsequent profitability. It would be difficult or even infeasible to specify these nonlinear functional forms in a linear model based on accounting or financial statement analysis intuition, which makes flexible machine learning algorithms such as GBRTs the appropriate estimation tool.

4. Main results

We first analyze whether linear OLS and nonlinear GBRT estimation of a simple autoregressive process yield more accurate out-of-

[12] We follow prior research and winsorize the data by year to make our results comparable to results from that research. A drawback of this research design choice is that it does not allow us to speak to outlier prediction, since outliers are, by definition, removed by the winsorization. Appendix O1, Table O1, shows that without winsorization NP's model continues to outperform the random walk when estimated using GBRTs but not when estimated using OLS.

Table 1
Descriptive statistics.

	N	Mean	SD	P1	P25	Median	P50	P99
ROCE	117,186	0.10	0.21	−0.70	0.06	0.12	0.18	0.69
ROTCE	117,186	0.11	0.22	−0.70	0.06	0.12	0.18	0.70
MSR	117,186	0.99	0.05	0.75	1.00	1.00	1.00	1.08
RNOA	117,186	0.14	0.48	−1.19	0.06	0.10	0.18	1.87
FLEV	117,186	0.61	1.46	−0.96	−0.12	0.31	0.89	7.61
SPREAD	117,186	0.10	0.59	−1.66	0.00	0.05	0.16	2.48
Sales PM	117,186	0.06	0.44	−1.10	0.04	0.07	0.14	0.61
ATO	117,186	2.19	2.57	0.12	0.79	1.56	2.55	14.08
Other Items NOA	117,186	0.00	0.01	−0.02	0.00	0.00	0.00	0.07
NBC	117,186	0.04	0.13	−0.48	0.01	0.04	0.06	0.55
*Sales PM**	117,186	0.07	0.44	−1.08	0.04	0.08	0.15	0.67
*ATO**	117,186	1.22	0.92	0.06	0.55	1.06	1.61	4.61
OLSPREAD	117,186	0.05	0.18	−0.58	0.01	0.04	0.09	0.63
OLLEV	117,186	0.88	2.33	0.05	0.26	0.40	0.66	12.30
Other Items OA	117,186	0.00	0.01	−0.01	0.00	0.00	0.00	0.04

This table presents descriptive statistics. The sample period is 1963–2023. The descriptive statistics are based on data from 1964 to 2022 because calculating the variables in the Nissim and Penman (2001, 2003) framework requires a one-year lag and calculating ROCE Lead 1 requires a year-lead. All variables are winsorized by year and defined in Appendix A.

sample year-ahead profitability forecasts than a random walk. Gerakos and Gramacy (2013) and Li and Mohanram (2014) find that a random walk tends to yield more accurate predictions than linear estimation of more complex earnings forecasting models, justifying random walk forecasts as a benchmark in our setting.

Table 3 presents the results. The first row ("All") tests whether linear and nonlinear estimation of ROCE's autoregressive process yield smaller mean out-of-sample absolute forecast errors than a random walk for the full sample. While linear estimation does not yield significantly lower absolute forecast errors than a random walk, nonlinear estimation does (at the 0.01 level), indicating that GBRT estimation effectively captures the nonlinear autoregressive relations visible in Fig. 2 Panel A. In terms of economic magnitude, relative to a random walk, GBRT estimation decreases mean ROCE absolute forecast errors by 0.858 percentage points (=11.439%–12.297%) of ROCE in absolute terms and by 6.977% (=11.439%/12.297%–1) in relative terms.[13] Fig. 4 illustrates this result. The improvement appears even more substantial once one recognizes that benchmark model forecast errors largely reflect random, unpredictable variation. Thus, the model improvements we document constitute an even larger (but unobservable) share of the variation in the predictable portion of the benchmark models' forecast errors (i.e., the portion of the forecast error that would be predictable if one had access to the hypothetical true but unknown model connecting future profitability to all known information at the time the forecast is made).

The remaining rows of the table analyze the distribution of absolute forecast errors. We group the absolute forecast errors in deciles and compute each decile's mean forecast error. While linear estimation of ROCE's autoregressive process yields significantly (at the 0.01 level) smaller absolute forecast errors than a random walk for deciles 8, 9, and 10 of the absolute forecast error distribution, absolute forecast errors in deciles 1 through 7 are larger (but not significantly so, at conventional levels). This finding is consistent with the intuition that the OLS objective function, minimizing squared residuals, tends to place more weight on avoiding large forecast errors than on frequently making precise forecasts. In contrast, except for the small absolute forecast errors in deciles 1 to 3 of the absolute forecast error distribution, nonlinear estimation yields significantly (at the 0.01 level) lower forecast errors than a random walk.[14] Fig. 5 illustrates these results. While linear estimation (the random walk) performs well in the upper (lower) deciles of the absolute forecast error distribution but underperforms in the lower (upper) deciles, nonlinear estimation performs well throughout.

Next, we test whether including the variables identified by the four levels of NP's profitability decomposition further increases forecast accuracy. Table 4 presents the results for OLS estimation (Panel A) and nonlinear GBRT estimation (Panel B). We do not find evidence that increasing the level of disaggregation improves the performance of linear models. With the exception of deciles 3 to 7 of the Level 1 disaggregation (splitting ROCE into the portions accruing to common equity holders and minority shareholders), absolute forecast errors of Level 1 to 4 models are not significantly (at conventional levels) smaller than those of the autoregressive Level 0 model on average and across the absolute forecast error distribution. In contrast, mean absolute forecast errors of forecasts for the full sample from nonlinear estimation monotonically decrease in the level of disaggregation, starting with Level 2 (splitting ROTCE into its operating and financing components), and continuing to Level 3 (splitting RNOA into margins and turnovers and splitting SPREAD into

[13] The mean book-to-market ratio in our sample is 60.5%, indicating that these effects amount to 0.519% and 4.221% of the average firm's market capitalization.

[14] Adding features that help to ex ante identify firms whose profitability will remain approximately unchanged might allow machine learning algorithms to ex post outperform the random walk not only in the higher (3 to 10) deciles but also the lower (1 to 2) deciles of the absolute forecast error distribution. We leave this analysis to future research.

Table 2
Correlations.

Variable		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
ROCE Lead 1	1	1.00	0.46	0.63	0.03	0.49	0.03	0.41	0.29	0.22	0.05	0.00	0.28	0.22	0.47	0.13	0.05
ROCE	2	0.46	1.00	0.99	0.05	0.78	−0.01	0.65	0.53	0.25	0.06	−0.01	0.51	0.26	0.74	0.12	0.06
ROTCE	3	0.46	0.99	1.00	−0.04	0.78	0.00	0.65	0.53	0.25	0.07	−0.01	0.51	0.26	0.75	0.13	0.06
MSR	4	0.00	0.02	−0.04	1.00	−0.01	0.01	−0.01	−0.03	0.03	−0.07	0.03	−0.01	0.06	−0.05	−0.06	−0.06
RNOA	5	0.23	0.40	0.40	−0.01	1.00	−0.10	0.83	0.56	0.45	0.06	−0.17	0.54	0.40	0.90	0.30	0.06
FLEV	6	0.06	0.04	0.04	−0.04	−0.10	1.00	−0.11	−0.02	−0.44	0.06	0.43	−0.01	−0.35	−0.35	−0.33	0.06
SPREAD	7	0.19	0.34	0.34	−0.01	0.89	−0.11	1.00	0.45	0.35	0.03	−0.56	0.49	0.31	0.78	0.23	0.03
Sales PM	8	0.29	0.38	0.38	−0.02	0.51	0.06	0.45	1.00	0.01	0.01	−0.12	0.98	−0.34	0.63	−0.14	0.01
ATO	9	0.05	0.06	0.06	0.02	0.34	−0.25	0.31	0.01	1.00	0.05	−0.10	−0.33	0.92	0.29	0.57	−0.01
Other Items/NOA	10	0.05	0.06	0.07	−0.05	0.15	−0.05	0.14	0.05	0.05	1.00	−0.03	0.01	−0.01	0.06	0.03	1.00
NBC	11	0.02	0.01	0.01	0.00	−0.05	0.09	−0.39	0.01	−0.06	−0.03	1.00	0.00	−0.04	−0.20	−0.12	0.05
*Sales PM**	12	0.29	0.38	0.38	−0.02	0.51	0.07	0.45	0.99	0.01	0.05	0.00	1.00	0.00	0.56	−0.08	0.01
*ATO**	13	0.11	0.14	0.14	0.06	0.22	−0.24	0.20	0.02	0.74	−0.01	−0.02	0.00	1.00	0.22	0.32	−0.02
OLSPREAD	14	0.31	0.53	0.53	−0.01	0.85	−0.10	0.75	0.65	0.20	0.15	−0.05	0.64	0.22	1.00	0.01	0.06
OLLEV	15	−0.01	−0.03	−0.02	−0.01	0.18	−0.14	0.17	0.00	0.42	0.04	−0.07	0.04	−0.01	0.01	1.00	−0.01
Other Items/OA	16	0.05	0.06	0.07	−0.04	0.11	−0.04	0.10	0.05	−0.01	0.95	−0.02	0.05	−0.03	0.13	−0.01	1.00

This table presents Pearson (Spearman) correlations below (above) the diagonal. The sample period is 1963–2023. The correlations are based on data from 1964 to 2022 because calculating the variables in the Nissim and Penman (2001, 2003) framework requires a one-year lag and calculating ROCE Lead 1 requires a year-lead. All variables are defined in Appendix A.

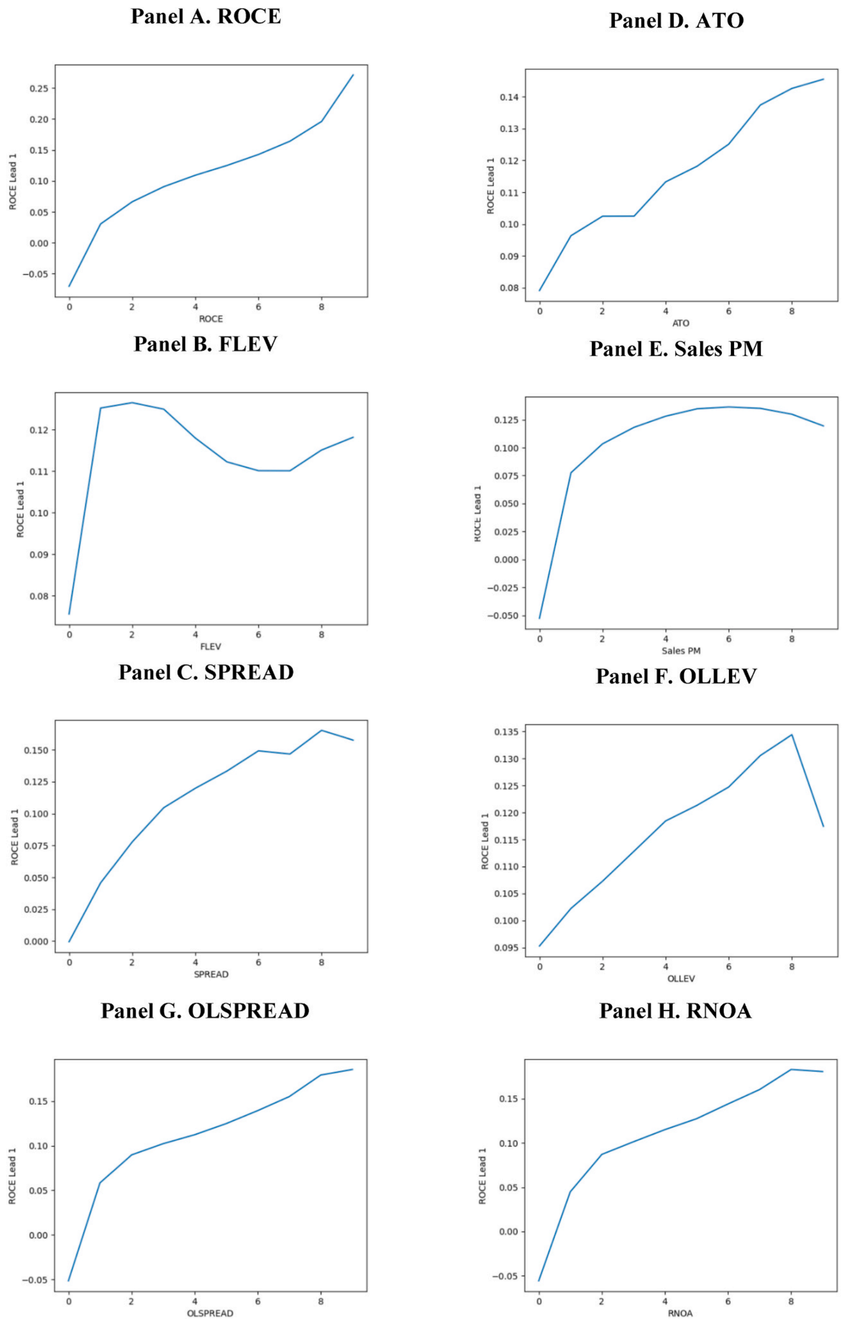

Fig. 2. Univariate Time-Series Plots Panels A to H plot median portfolio ROCE in periods t + 1, t + 3, and t + 5 by ROCE, FLEV, SPREAD, ATO, Sales PM, OLLEV, OLSPREAD, and RNOA decile in period t. All variables are defined in Appendix A.

RNOA and Net Borrowing Cost), and Level 4 (splitting out operating leverage).[15] These findings are consistent with NP's untabulated finding that linear estimation of their ratio decomposition framework does not improve out-of-sample prediction, and their conjecture that the reason is the nonlinear relation between current and future ratios, making accommodating nonlinearities imperative in

[15] An additional advantage of GBRTs over OLS is that GBRTs allow us to include all predictors jointly. In contrast, when using OLS, higher-level predictors, such as RNOA and NBC in Level 3, are collinear with lower-level predictors, such as SPREAD, which precludes including them jointly. Therefore, in the GBRT when analyzing Level 4, we also include variables from Level 1, 2, and 3 as independent variables, whereas with OLS we can include only the ratios from Level 4 to avoid collinearity. As a result, the incremental improvements we observe when moving from Level 0 to 4 in Table 4 Panel B (but not in Panel A) are interpretable as the incremental value of having access to the additional disaggregated information.

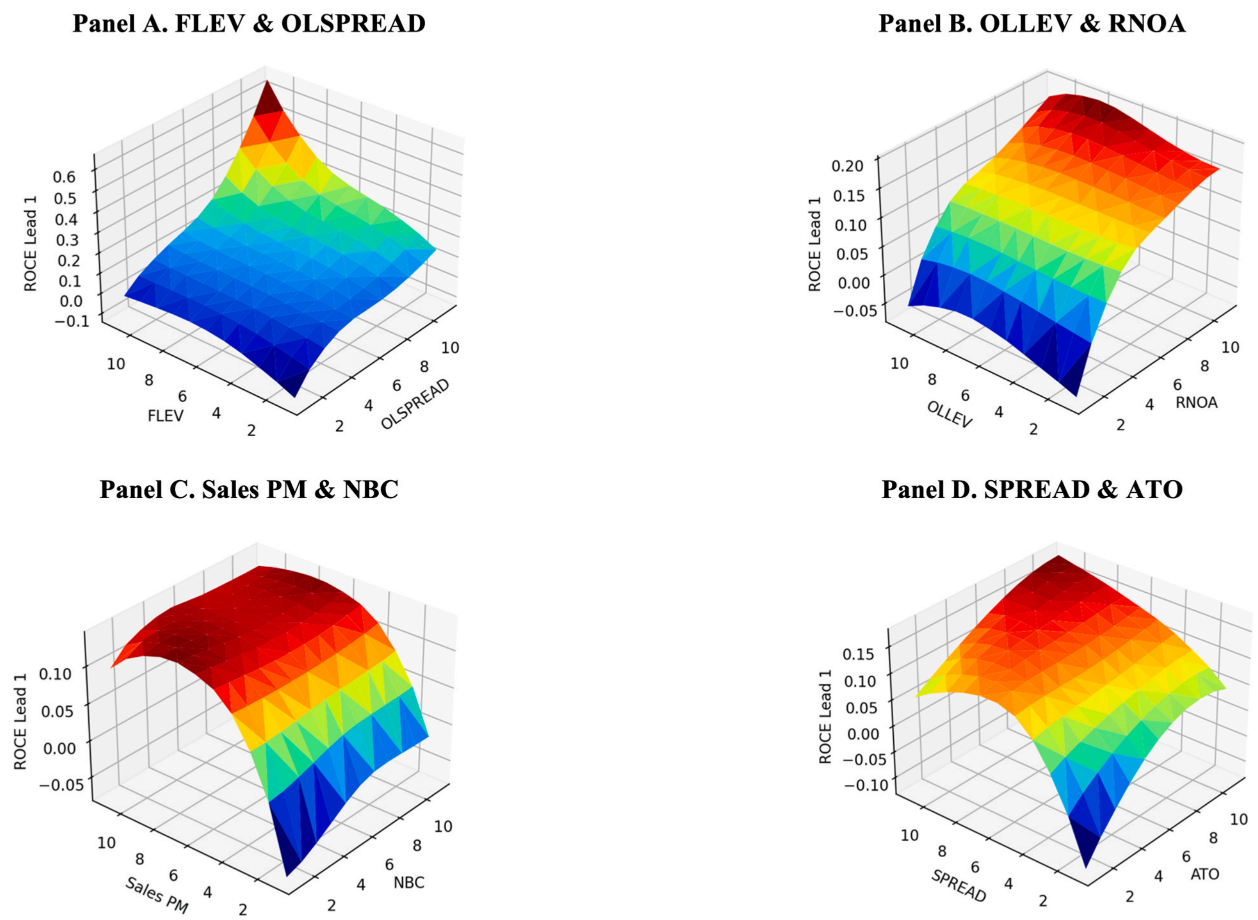

Fig. 3. Interactive Relations across Variables in ROCE Prediction Panels A to D plot median portfolio ROCE in period t + 1 by FLEV and OLSPREAD, OLLEV and RNOA, Sales PM and NBC, and SPREAD and ATO decile in period t. All variables are defined in Appendix A.

Table 3
Absolute forecast errors from random walk, linear (OLS) estimation, and nonlinear (GBRT) estimation of year-ahead profitability in an autoregressive model.

Error Decile	RW	OLS	GBRT
All	**12.297**	**12.203**	**11.439*****
1	0.294	0.514	0.302
2	0.930	1.521	0.946
3	1.696	2.561	1.692
4	2.638	3.675	2.584***
5	3.867	4.933	3.716***
6	5.533	6.476	5.230***
7	8.042	8.516	7.422***
8	12.055	11.799***	10.982***
9	20.318	18.884***	18.280***
10	67.593	63.154***	63.230***

This table shows mean absolute year-ahead ROCE forecast errors of models based on a random walk (RW), linear (OLS) estimation and nonlinear (GBRT) estimation of a model including lagged ROCE as the sole predictor. The row labeled "All" lists the mean absolute forecast error for the full sample. The rows labeled "1" through "10" list the mean absolute forecast errors within the 1st to 10th deciles of the absolute forecast error distribution. *, **, and *** indicate that the model's mean absolute forecast error is significantly smaller at the 10%, 5%, and 1% levels, respectively, than that of the random walk model in the second column of the same row.

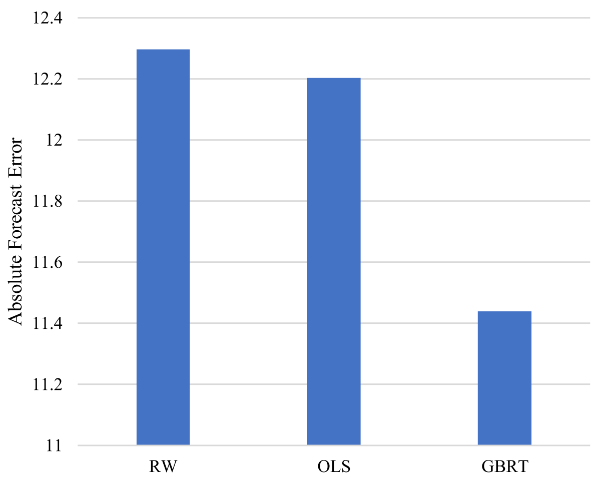

Fig. 4. Mean Absolute Forecast Error Comparison: Random Walk, OLS, and GBRT This figure plots mean absolute forecast errors for the full sample for the random walk (RW), OLS, and GBRT (gradient-boosted regression tree) models shown in Table 3.

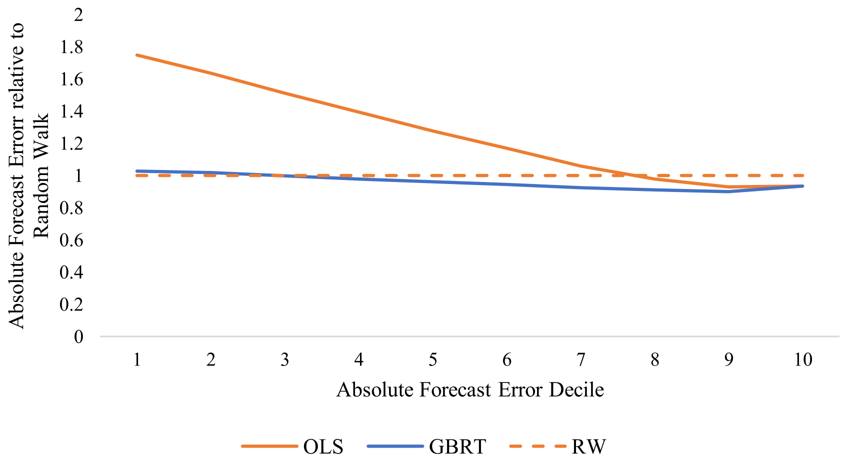

Fig. 5. OLS and GBRT relative to Random Walk Absolute Forecast Errors across the Absolute Forecast Error Distribution This figure plots mean absolute forecast errors across the absolute forecast error deciles for OLS and gradient-boosted regression tree (GBRT) models scaled by random walk forecast errors shown in Table 3.

estimating their framework (p. 128).

In terms of economic magnitude, the improvement in accuracy is substantial relative to results reported in Fairfield and Yohn (2001) and Esplin et al. (2014) who examine whether linear estimation of NP's Level 2 and Level 3 disaggregations improves out-of-sample profitability forecast accuracy. Relative to the simple autoregressive Level 0 model, employing the Level 4 disaggregation in NP's framework incrementally decreases mean ROCE absolute forecast errors by 0.388 percentage points [= 11.051–11.439] of ROCE in absolute terms and by 3.392% [= 11.051/11.439–1] in relative terms. Fig. 6 illustrates this result. The improvements largely derive from deciles 5 through 10 of the absolute forecast error distribution, providing evidence that NP's profitability decomposition framework is especially useful when it comes to avoiding large forecast errors.

To summarize, results in Tables 3 and 4 indicate that nonlinear estimation of even a simple autoregressive model improves on a random walk and that increasing the level of disaggregation identified by NP's ratio decomposition framework combined with nonlinear estimation further increases forecasting performance, in particular, by avoiding large forecast errors. The finding that NP's framework improves performance only after accounting for its inherent nonlinearities suggests synergies between nonlinear estimation and the use of nonlinear profitability decomposition frameworks, such as NP's.

5. Robustness tests

5.1. Alternative machine learning approaches

Panels A and B of Table 5 test the robustness of our inferences to three alternative machine learning approaches. First, we estimate

Table 4
Absolute One-Year Ahead Profitability Forecast Errors from Linear (OLS) and Nonlinear (GBRT) Estimation of NP's Framework
This table shows mean absolute year-ahead ROCE forecast errors of models based on linear (OLS) and nonlinear (GBRT) estimation of Level 0 to Level 4 of NP's profitability decomposition framework in Panels A and B, respectively. The row labeled "All" lists the mean absolute forecast error for the full sample. The rows labeled "1" through "10" list the mean absolute forecast errors within the 1st to 10th deciles of the absolute forecast error distribution. *, **, and *** indicate that the model's mean absolute forecast error is significantly smaller at the 10%, 5%, and 1% levels, respectively, than that of the Level 0 model in the second column of the same row.

Panel A. Linear Estimation (OLS)

Error Decile	Level 0	Level 1	Level 2	Level 3	Level 4
All	**12.203**	**12.198**	**14.279**	**14.101**	**13.648**
1	0.514	0.513	0.637	0.628	0.603
2	1.521	1.516	1.916	1.892	1.813
3	2.561	2.553**	3.252	3.220	3.064
4	3.675	3.662***	4.691	4.647	4.411
5	4.933	4.910***	6.345	6.265	5.960
6	6.476	6.454***	8.426	8.256	7.848
7	8.516	8.501*	11.118	10.880	10.348
8	11.799	11.804	15.076	14.736	14.157
9	18.884	18.894	22.635	22.089	21.542
10	63.154	63.173	68.690	68.397	66.730

Panel B. Nonlinear estimation (GBRT)

Error decile	Level 0	Level 1	Level 2	Level 3	Level 4
All	**11.439**	**11.437**	**11.314**	**11.192****	**11.051*****
1	0.302	0.307	0.311	0.312	0.308
2	0.946	0.948	0.954	0.970	0.958
3	1.692	1.693	1.694	1.713	1.695
4	2.584	2.585	2.578*	2.596	2.583
5	3.716	3.717	3.698***	3.699***	3.675***
6	5.230	5.230	5.202***	5.195***	5.141***
7	7.422	7.421	7.365***	7.317***	7.241***
8	10.982	10.973	10.857***	10.719***	10.604***
9	18.280	18.275	17.957***	17.666***	17.402***
10	63.230	63.214	62.521	61.733**	60.903***

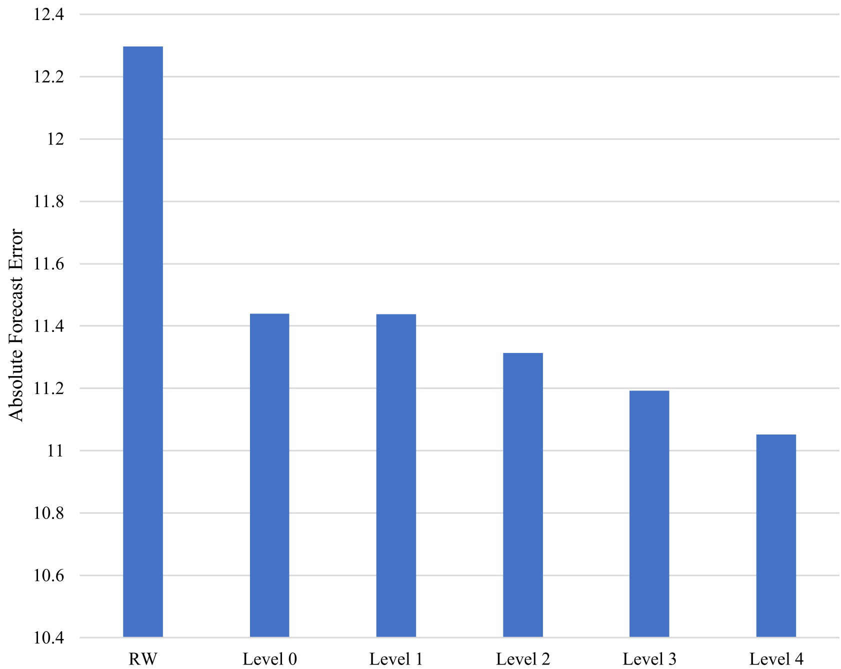

Fig. 6. Mean Absolute Forecast Errors across Decomposition Levels This figure plots mean absolute forecast errors for the full sample for the random walk (RW) and across Level 0 to Level 4 of NP's profitability decomposition framework shown in Fig. 1 and GBRT models shown in Table 4 Panel B.

Table 5
Robustness Test: Alternative Machine Learning Approaches and Holdout Sample
This table shows mean absolute year-ahead ROCE forecast errors of models based on nonlinear estimation of Level 0 and Level 4 of NP's profitability decomposition framework estimated using a holdout sample (Holdout), random forest (Random Forest) estimation, and neural network (Neural Network) estimation. Panel A compares the Level 0 forecast against a random walk and Panel B compares the Level 4 forecast against the corresponding Level 0 forecast. The row labeled "All" lists the mean absolute forecast error for the full sample. The rows labeled "1" through "10" list the mean absolute forecast errors within the 1st to 10th deciles of the absolute forecast error distribution. *, **, and *** indicate that the model's mean absolute forecast error is significantly smaller at the 10%, 5%, and 1% levels, respectively, than that of the random walk (Panel A) and that of the corresponding Level 0 model (Panel B) in the adjacent column of the same row.

Panel A. Comparison to RW

Error Decile	RW	Neural Network	Random Forest	Holdout
All	**12.297**	**11.437*****	**11.428*****	**11.457*****
1	0.294	0.306	0.301	0.303
2	0.930	0.954	0.940	0.942
3	1.696	1.695	1.683***	1.686***
4	2.638	2.585***	2.576***	2.579***
5	3.867	3.710***	3.703***	3.708***
6	5.533	5.211***	5.220***	5.233***
7	8.042	7.393***	7.412***	7.432***
8	12.055	10.921***	10.983***	10.996***
9	20.318	18.195***	18.298***	18.352***
10	67.593	63.393***	63.157***	63.332***

Panel B. NP Variables

	Neural Network		Random Forest		Holdout	
Error Decile	Level 0	Level 4	Level 0	Level 4	Level 0	Level 4
All	**11.437**	**11.187****	**11.428**	**11.143*****	**11.457**	**11.168*****
1	0.306	0.328	0.301	0.304	0.303	0.310
2	0.954	1.018	0.940	0.946	0.942	0.971
3	1.695	1.793	1.683	1.692	1.686	1.715
4	2.585	2.692	2.576	2.573	2.579	2.607
5	3.710	3.804	3.703	3.692**	3.708	3.714
6	5.211	5.264	5.220	5.185***	5.233	5.182***
7	7.393	7.369**	7.412	7.317***	7.432	7.272***
8	10.921	10.739***	10.983	10.738***	10.996	10.638***
9	18.195	17.627***	18.298	17.632***	18.352	17.541***
10	63.393	61.237***	63.157	61.349**	63.332	61.725**

our models using neural networks and random forests, two commonly used machine learning algorithms. Second, to address the concern that observations deriving from the same firm may be autocorrelated, we repeat our analysis for a holdout sample of 20% of firms.[16] For each of the three alternative machine learning algorithms, we find that nonlinear estimation yields lower absolute ROCE forecast errors than a random walk and that nonlinear estimation of NP's framework further improves performance, particularly in higher deciles of the forecast error distribution.

5.2. *Additional predictors*

Table 6 compares the performance of NP's model (Baseline) to that of: (1) a Dupont analysis (Dupont); (2) Hou, Van Dijk, and Zhang's (2012) profitability forecasting model (Hou); (3) adding Fama and French (1997) 48 industry fixed effects to NP's model (Industry); (4) adding real GDP growth, inflation, and unemployment rate to NP's model (Macro); and (5) adding both industry fixed effects and real GDP growth, inflation, and unemployment rate to NP's model (Industry & Macro).[17] To ensure that the performance differences we document are due to the models rather than estimation, we estimate all models via GBRTs and optimize hyperparameters for each model separately. None of the alternative models yields significantly (at conventional levels) lower absolute forecast errors, as compared to the Baseline (NP) model.

[16] To obtain a prediction for every observation without using the same firms in both the training and test samples, we perform cross-validation when estimating prediction accuracy on the holdout sample. Specifically, we split our sample of firms into five equally sized groups. We then use four groups to train and the fifth group to evaluate the model. To generate predictions for each of the five groups, we repeat this process five times, each time with a different evaluation group. As a result, the holdout sample on which we test the out-of-sample prediction accuracy contains neither years nor firms in the training sample, which addresses concerns that the accuracy improvements we document are driven by within-year or within-firm correlations. To choose hyperparameters, we randomly select one of the five groups and use it to choose hyperparameters for all five groups.

[17] In the Dupont and Hou et al. (2012) models, we include only the variables listed in the Appendix. The other three models include all ratios in Levels 1, 2, 3, and 4 of the NP framework and the Industry and Macro variables.

Table 6
Robustness test: Additional predictors.

Error Decile	Baseline	Dupont	Hou	Industry	Macro	Industry & Macro
All	**11.376**	**11.722**	**12.340**	**11.464**	**11.405**	**11.388**
1	0.317	0.350	0.416	0.329	0.330	0.336
2	0.980	1.076	1.276	1.016	1.014	1.031
3	1.742	1.893	2.188	1.792	1.783	1.801
4	2.645	2.851	3.229	2.706	2.696	2.716
5	3.768	4.042	4.477	3.857	3.834	3.855
6	5.275	5.617	6.091	5.375	5.339	5.366
7	7.432	7.826	8.337	7.548	7.483	7.505
8	10.892	11.333	11.899	11.047	10.929	10.947
9	17.907	18.440	19.336	18.140	17.918	17.898
10	62.803	63.792	66.147	62.828	62.724	62.426

This table shows mean absolute year-ahead ROCE forecast errors of models based on nonlinear estimation of Level 4 of NP's ratio decomposition framework (Baseline); the ratios from the Dupont decomposition (Dupont); the predictors in Hou et al. (2012) (Hou); Level 4 of NP's ratio decomposition plus indicators for each Fama-French 48-industry (Industry); Level 4 of NP's ratio decomposition plus annual real GDP growth, inflation, and unemployment rate (Macro); and Level 4 of NP's ratio decomposition plus both the Industry and Macro variables. The row labeled "All" lists the mean absolute forecast error for the full sample. The rows labeled "1" through "10" list the mean absolute forecast errors within the 1st to 10th deciles of the absolute forecast error distribution. *, **, and *** indicate that the model's mean absolute forecast error is significantly smaller at the 10%, 5%, and 1% levels, respectively, than that of the Baseline model in the second column of the same row.

6. Additional tests

6.1. Core items

While some of the eight ROCE drivers in NP's framework, such as ATO, are persistent, others, such as RNOA deriving from unusual operating income, are mean reverting (transitory). Excluding transitory components could enhance forecasting performance by decreasing prediction-irrelevant noise or impair performance because of information loss; that is, the treatment of non-core/transitory items is a distinct financial statement analysis design choice involving a tradeoff between information loss and noise. Acknowledging this, NP adjust their decomposition of ROCE as follows:

$$\begin{aligned} \text{ROCE} \;=\; & \text{MSR} \times \left[\text{Core Sales PM}^{*} \times \text{ATO}^{*} + \frac{\text{Core Other Items}}{\text{OA}} + \frac{\text{UOI}}{\text{OA}} + \text{OLLEV} \times \text{OLSPREAD} + \text{FLEV} \right. \\ & \left. \times \left(\text{Core RNOA} - \text{Core NBC} + \frac{\text{UOI}}{\text{NOA}} - \frac{\text{UFE}}{\text{NFO}} \right) \right], \end{aligned} \tag{2}$$

where Core Sales PM* denotes modified profit margin from core sales (=[Core OI from Sales + io]/Sales), UOI unusual operating income, Core RNOA core return on net operating assets (= Core OI from Sales/NOA + Core Other Items/NOA), Core NBC core net borrowing cost (= Core NFE/NFO), and UFE unusual financial expense. Equation (2) identifies eight relatively more persistent drivers of ROCE: MSR, FLEV, Core NBC, ATO*, Core Sales PM*, Core Other Items/OA, OLLEV, and OLSPREAD.

There are at least three considerations as to whether including versus excluding items labeled transitory/non-core (i.e., UOI/OA, UFE/NFO) improves forecasting. First, accounting requirements may produce transitory income items with predictive ability. Penman and Zhang (2002) argue that conservative accounting rules can generate future-period (accounting) benefits while decreasing current-period income; for example, recording a current-period impairment loss implies an increase in future accounting performance. The impairment loss, classified as transitory/non-core, would be relevant for predicting future earnings. Second, while models that include persistent operating items and exclude transitory non-operating items should (theoretically) produce better forecasts, both theory (Dye, 2002) and empirical research (Barnea et al., 1976; Kinney and Trezevant, 1997; Givoly, Hayn, and D'Souza, 2000; McVay, 2006) suggest managers sometimes manipulate income statement presentation to blur the core/non-core distinction. Third, the distinction between transitory/non-core and persistent/core income items arises at least partly from firms' business models. The empirical measures used by NP and in this paper are based on Compustat data definitions applied to all entities, which may imperfectly separate core from non-core items for some firms. Thus, whether a financial statement analysis design choice to focus on core items improves forecasts is an empirical question.

Table 7 reports results of tests as to whether focusing on core items improves model performance relative to our baseline model (Baseline) by including only the core components of each ratio in our models (Core). We find that mean absolute forecast errors are significantly (at the 0.01 level) smaller than those of the baseline model for the full sample. Fig. 7 Panel A illustrates this result. These improvements derive from medium to large absolute forecast errors in deciles 2 through 10 of the absolute forecast error distribution. These findings suggest that focusing on core (i.e., likely recurring) items improves forecast accuracy.

6.2. Amount of past information

An analyst must decide how much historical information to consider. On the one hand, more lags of past data increase the amount

Table 7
Additional test: Core items.

Error Decile	Baseline	Core
All	**11.051**	**10.677*****
1	0.308	0.309
2	0.958	0.949***
3	1.695	1.650***
4	2.583	2.484***
5	3.675	3.523***
6	5.141	4.886***
7	7.241	6.833***
8	10.604	10.037***
9	17.402	16.636***
10	60.903	59.464**

This table shows mean absolute year-ahead ROCE forecast errors of models based on nonlinear estimation of Level 4 of NP's profitability decomposition framework without differentiating between core and transitory items (Baseline) and focusing on core items only (Core). The row labeled "All" lists the mean absolute forecast error for the full sample. The rows labeled "1" through "10" list the mean absolute forecast errors within the 1st to 10th deciles of the absolute forecast error distribution. *, **, and *** indicate that the model's mean absolute forecast error is significantly smaller at the 10%, 5%, and 1% levels, respectively, than that of the Baseline model in the second column of the same row.

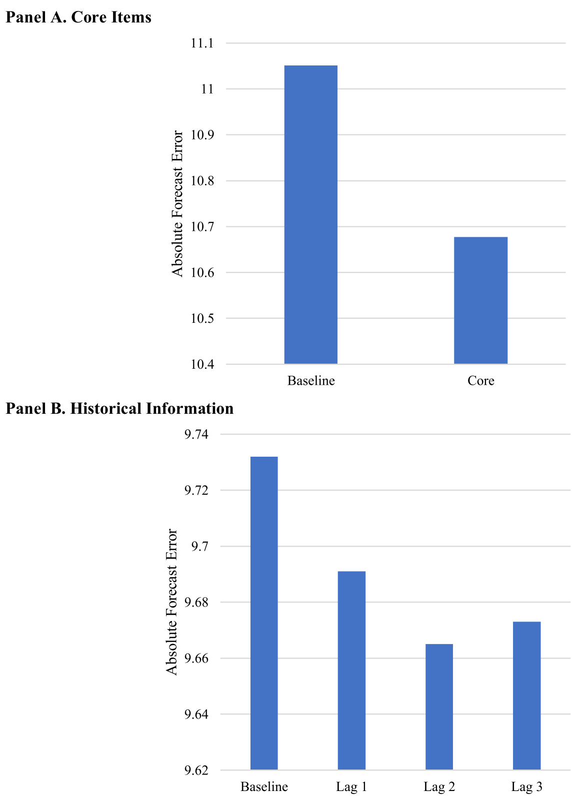

Fig. 7. Mean Absolute Forecast Error Comparison: Core Items (Panel A) and Historical Information (Panel B) Panel A plots mean absolute forecast errors for models including core and transitory items (Baseline) and models focusing on core items (Core) shown in Table 7. Panel B plots mean absolute forecast errors for models using only current financial statement data (Baseline) and models using current data and one to three lags of past data (Lag 1 to Lag 3) shown in Table 8.

Table 8
Additional test: Historical information.

Error Decile	Baseline	Lag 1	Lag 2	Lag 3
All	**9.732**	**9.691**	**9.665**	**9.673**
1	0.294	0.289**	0.284***	0.290*
2	0.905	0.903	0.896***	0.901*
3	1.609	1.595***	1.592***	1.599***
4	2.442	2.433**	2.434**	2.433**
5	3.490	3.473***	3.471***	3.473***
6	4.846	4.836	4.803***	4.813***
7	6.804	6.722***	6.699***	6.704***
8	9.812	9.658***	9.616***	9.613***
9	15.675	15.446***	15.369***	15.350***
10	51.439	51.552	51.482	51.558

This table shows mean absolute year-ahead ROCE forecast errors of models based on nonlinear estimation of Level 4 of NP's profitability decomposition framework using only current data (Baseline) and one (Lag 1), two (Lag 2), and three (Lag 3) additional lags of historical data. The row labeled "All" lists the mean absolute forecast error for the full sample. The rows labeled "1" through "10" list the mean absolute forecast errors within the 1st to 10th deciles of the absolute forecast error distribution. *, **, and *** indicate that the model's mean absolute forecast error is significantly smaller at the 10%, 5%, and 1% levels, respectively, than that of the Baseline model in the second column of the same row.

of information supporting predictions. For example, a longer time series helps capture how the firm's profitability components behave through the business cycle, an important consideration given that business cycle fluctuations explain considerable variation in firms' profitability.[18] On the other hand, using more historical data increases the likelihood a firm's activities have changed sufficiently since the information was recorded that the historical information is no longer useful for prediction. Given these conflicting considerations, we test how the performance of our model that includes only current data (Baseline) changes when we add one (Lag 1), two (Lag 2), and three (Lag 3) lags of historical data. For consistency, we include observations with three lags of non-missing data for all ratios in NP's framework.

Results of these tests, presented in Table 8, show that including two or three lags of historical information significantly (at the 0.10 level or better) improves forecast performance for all but the most extreme forecast error decile. Fig. 7 Panel B illustrates that most of this benefit derives from the first two lags; adding the third lag improves performance less. Viewed as a whole, these results provide suggesting that using up to three years of past data improves forecast performance.

6.3. Can forecasts from nonlinear estimation of NP's profitability framework inform investors and analysts?

This section analyzes whether the forecasts produced via nonlinear estimation of NP's profitability decomposition framework provide information incremental to the information priced by investors or provided by analysts. We proceed in three steps.

In Step 1, we jointly optimize the following hyperparameter and design choices: (1) the GBRT's number of trees ∊ [10, 50, 100, 500, 1000, 1500]; (2) the GBRT's maximum depth ∊ [1, 3, 5, 10, 20]; (3) the disaggregation level ∊ [0, 1, 2, 3, 4]; (4) focus on core items ∊ [0, 1]; (5) number of years of lagged information to include ∊ [0, 1, 2, 3]; (6) whether to include macroeconomic variables ∊ [0, 1]; and (7) whether to include industry fixed effects ∊ [0, 1]. We find that the optimal model features 100 trees, a depth of 3, level 4 disaggregation, a focus on core items, one year of lagged information, macroeconomic variables, and no industry fixed effects.

Second, following Chen et al. (2022), Table 9 Panel A Columns (1) and (2) report results of regressing stock returns cumulated over the 12-months starting three months after the fiscal year end on the difference between the optimal model earnings forecast described in Step 1 and the current earnings realization (*Optimal Model*), firm fixed effects, and fiscal year-end-date fixed effects before and after including Lee et al.'s (2024) six asset pricing factors. We cluster standard errors by firm and year. We find evidence that the forecasts generated through nonlinear estimation of NP's framework predict returns. The slope coefficient on *Optimal Model* is significantly (at the 0.01 level) positive before and after controlling for Lee et al.'s (2024) asset pricing factors. In terms of economic magnitude, the Column (2) estimates suggest that a forecasted one percentage point increase in ROCE is associated with a 0.247% increase in returns, which is large relative to the magnitude of the slope coefficients of Lee et al.'s (2024) factors.

For a sample of observations for which IBES analyst consensus GAAP earnings forecasts are available, we test whether our forecasts are informative beyond analyst forecasts.[19] Specifically, Table 9, Panel A Columns (3) to (6) repeat the analysis in Columns (1) and (2) including the difference between first mean (*Mean Analyst Forecast*) or median (*Median Analyst Forecast*) consensus analyst ROCE forecasts that become available three months after the fiscal year end and the current earnings realization. While consensus analyst forecasts do not significantly (at conventional levels) predict returns, consistent with the notion that these forecasts represent information that has already been priced by the market, results show that our forecasts predict returns after controlling for analyst forecasts.

[18] See, for example, Ball and Brown (1967), Ball et al. (2009), Bonsall et al. (2013), Binz et al. (2022), Binz (2022), and Binz et al. (2023).

[19] We do not analyze IBES Street Earnings forecasts because these forecasts are not directly comparable to GAAP Earnings forecasts, across analysts (since each analyst forecasts his or her own proprietary definition of earnings), or within analyst (since an analyst might change his or her definition of earnings across firms or within firm across time).

Table 9
Informativeness of profitability forecasts from nonlinear estimation of NP's profitability framework for investors (Panel A) and analysts (Panel B).

Panel A. Stock Returns

	Returns					
	(1)	(2)	(3)	(4)	(5)	(6)
Optimal Model	0.255***	0.247***	0.256***	0.248***	0.258***	0.251***
	(3.028)	(2.868)	(3.028)	(2.870)	(3.074)	(2.915)
Median Analyst Forecast			−0.001	−0.001		
			(-0.422)	(-0.476)		
Mean Analyst Forecast					−0.005	−0.005
					(-1.322)	(-1.335)
Beta		−0.003		−0.003		−0.003
		(-0.164)		(-0.165)		(-0.166)
SMB		0.009		0.009		0.009
		(0.437)		(0.435)		(0.430)
HML		0.022**		0.022**		0.022**
		(2.177)		(2.177)		(2.178)
RMW		−0.022		−0.022		−0.022
		(-1.270)		(-1.270)		(-1.270)
CMA		0.013		0.013		0.013
		(1.109)		(1.110)		(1.112)
UMD		−0.017		−0.017		−0.017
		(-0.581)		(-0.580)		(-0.580)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,050	19,050	19,050	19,050	19,050	19,050
R-squared	0.464	0.466	0.464	0.466	0.464	0.466

Panel B. Changes in profitability

	ΔROCE					
	(1)	(2)	(3)	(4)	(5)	(6)
Optimal Model	1.058***	1.058***	1.051***	1.050***	1.046***	1.045***
	(23.638)	(23.714)	(22.536)	(22.600)	(23.663)	(23.716)
Median Analyst Forecast			0.010	0.010		
			(1.395)	(1.405)		
Mean Analyst Forecast					0.017*	0.017*
					(1.974)	(1.977)
Beta		0.003		0.003		0.003
		(0.353)		(0.369)		(0.368)
SMB		−0.001		−0.001		−0.001
		(-0.276)		(-0.199)		(-0.158)
HML		−0.002		−0.002		−0.002
		(-0.387)		(-0.414)		(-0.438)
RMW		0.005		0.005		0.005
		(1.105)		(1.104)		(1.103)
CMA		0.000		0.000		−0.000
		(0.029)		(0.004)		(-0.016)
UMD		−0.004		−0.004		−0.005
		(-0.615)		(-0.618)		(-0.622)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,050	19,050	19,050	19,050	19,050	19,050
R-squared	0.443	0.443	0.444	0.444	0.445	0.445

Panel A [Panel B] regresses stock returns (*Returns*) [changes in profitability (*ΔROCE*)] over the subsequent year on the change in profitability predicted by the optimal model described in Section 6.3 estimated via gradient-boosted regression trees (*Optimal Model*), the change in profitability predicted by the mean (*Mean Analyst Forecast*) or median (*Median Analyst Forecast*) analyst consensus forecast, risk factors from Lee et al. (2024), and fixed effects. Standard errors are clustered by firm and year. All variables are defined in Appendix A. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Third, Table 9 Panel B examines a potential mechanism or channel through which our forecasts predict returns incrementally to analyst forecasts, namely, improved profitability forecasting. Specifically, we test whether our forecasts continue to predict year-ahead changes in ROCE (*ΔROCE*) after controlling for the mean or median consensus analyst forecast. We find evidence that this is the case. Results show that after controlling for the median (Column (4)) and mean (Column (6)) consensus analyst forecast a one percent increase in forecasted ROCE predicts a 1.050% (Column (4)) and 1.045% (Column (6)) increase in future ROCE, suggesting that our earnings forecasts provide incremental information beyond those made by analysts.

Viewed as a whole, the results in this section suggest analysts could use nonlinear estimation of profitability decomposition frameworks to enhance their forecasting precision and investors could build on the resulting improvements in forecast accuracy to trade profitably.

7. Conclusion

We examine how nonlinear estimation of profitability decomposition frameworks via machine learning enhances profitability forecasting. We estimate the nonlinear profitability decomposition framework proposed by Nissim and Penman (2001, 2003) but not estimated, because estimation methods available at the time were insufficient for the framework's nonlinear structure. Subsequent studies that apply linear approximations to analyze components of NP's framework are likely vulnerable to incorrect inferences stemming from some combination of the linear approximation and considering piecemeal components, not the framework as a whole. We resolve the issues of nonlinearity and piecemeal analysis by using gradient-boosted regression trees, a widely used machine learning algorithm that can capture arbitrarily complex nonlinear and interactive relations among variables, to forecast firm-specific profitability using NP's framework as a whole.

We aim to analyze NP's profitability decomposition framework, not to search statistically for the best profitability predictors, as might be done, for example, by applying machine learning to a large set of predictors without a framework that specifies, and thereby limits, the information to be considered. Such an approach focuses on prediction, not explanation, as discussed by, for example, Bertomeu (2020), and precludes consideration of tradeoffs in financial statement analysis design choices. To provide insights for the practice of fundamental analysis, we explore the effects of variation in three financial statement analysis design choices that analysts must make on empirical rather than theoretical grounds: the granularity of profitability disaggregation, the inclusion vs exclusion of non-core items, and the number of lags of past data to include.

We replicate NP's finding that linear estimation of their framework does not improve forecast performance. However, we find that nonlinear estimation of a simple autoregressive model via gradient-boosted regression trees produces profitability forecasts that are substantially more accurate than those derived from either a random walk or linear estimation, especially for firm-years for which forecasting is more difficult. Using the variables in NP's framework as inputs rather than an autoregressive model further improves forecasting precision, suggesting substantial synergies in the combination of nonlinear estimation and NP's framework. Our inferences are robust to estimation using alternative machine learning algorithms. We find no evidence that considering firms' industry membership, variables capturing the state of the macroeconomy, or firm-level profitability predictors proposed in prior research as alternative or additional predictors further enhances forecast accuracy.

Analyzing the effects of financial statement analysis design choices, we find that higher levels of disaggregation, a focus on core items, and using more historical information improve forecast accuracy. Finally, we find that our forecasts predict returns and changes in profitability before and after controlling for analyst forecasts and common asset pricing factors, suggesting that analyses similar to those in this paper could be useful to analysts and investors.

While our results indicate that machine learning has the capacity to use the information identified by ratio decomposition frameworks to derive more accurate profitability predictions than are produced by the benchmark models we consider, they do not directly speak to whether or how decomposition frameworks can successfully be implemented by human beings. We leave this question, which would require a different research design, to future research.

Appendix A. Variable Definitions

Panel A lists the ratios in the NP framework, the level (0 through 4) to which they correspond, a description of the variable, the Compustat line items and internally generated variables used to calculate each ratio, and any corresponding core ratio. Panel B lists variables used for constructing the variables in Panel A. Panel C provides definitions for variables that are not part of the NP framework. Compustat line items are in lower case, and internally generated variables are in upper case. Δ indicates the change in the variable over the past year.

Panel A. NP Ratios

Highest NP Level	Description	Variable Name	Definition	Core Variable Name	Core Definition
4	Modified asset turnover	*ATO**	Sale/OA		
4	Operating leverage	*OLLEV*	OL/NOA		
4	Operating leverage spread	*OLSPREAD*	(OI + IO)/OA − IO/OL		
4	Other items divided by operating assets	*Other Items/ OA*	(Other income Items)/OA	*Core Other Items/ OA*	(Other income items − UOI)/OA
4	Modified sales profit margin	*Sales PM**	(OI from sales + IO)/sale	*Core Sales PM**	(Core OI from sales + IO)/sale
3	Sales profit margin	*Sales PM*	OI from sales/sale	*Core Sales PM*	Core OI from sales/sale
3	Asset turnover	*ATO*	Sale/NOA		
3	Other items divided by net operating assets	*Other Items/ NOA*	Other income Items/NOA	*Core Other Items/ NOA*	(Other income items-UOI)/NOA
3	Net borrowing costs	*NBC*	NFE/NFO	*Core NBC*	Core NFE/NFO

(continued on next page)

(*continued*)

Highest NP Level	Description	Variable Name	Definition	Core Variable Name	Core Definition
2	Return on net operating assets	*RNOA*	Sales PM × ATO + other Items/NOA	*Core RNOA*	Core sales PM × ATO + core other Items/NOA
2	Financial leverage	*FLEV*	NFO/CSE		
2	Financial leverage spread	*SPREAD*	RNOA − NBC	*Core SPREAD*	Core RNOA – Core NBC
1	Minority share ratio	*MSR*	(CNI/CSE)/[(CNI + mii)/(CSE + mib)]		
1	Return on total common equity	*ROTCE*	FLEV × SPREAD + RNOA	*Core ROTCE*	FLEV × core SPREAD + core RNOA
0	Return on common equity	*ROCE*	ROTCE × MSR	*Core ROCE*	Core ROTCE × MSR

Panel B. Internally Calculated Variables

Variable Name	Description	Calculation
FA	Financial assets	che + ivao
FO	Financial obligations	dlc + dltt + pstk – tstkp + dvpa
NFO	Net financial obligations	FO – FA
OA	Operating assets	At – FA
CSE	Common shareholders equity	ceq + tstkp – dvpa
NOA	Net operating assets	NFO + CSE + mib
OL	Operating liabilities	OA – NOA
MTR	Marginal Tax rate	Top statutory federal tax rate plus 2 percent average state tax rate. The top federal statutory corporate tax (in percent): 52 (1963), 50 (1964), 48 (1965–1967), 52.8 (1968–1969), 49.2 (1970), 48 (1971–1978), 46 (1979–1986), 40 (1987), 34 (1988–1992), 35 (1993–2017), and 21 (2018–2023).
CORE NFE	Core net financial expense	xint × (1 – MTR) + dvp – idit × (1 – MTR)
UFE	Unusual financial expense	Lag(msa) – msa
NFE	Net financial expense	Core NFE + UFE
CSA	Clean Surplus Adjustment	msa – Lag(msa) + recta – Lag(recta)
CNI	Comprehensive net income	Ni – dvp + CSA
OI	Operating income	NFE + CNI + mii
UOI	Unusual operating income	(1 – MTR) × (nopio + spi – esub) + xido + recta – Lag(recta)
Other income items	Other income items	esub
OI from sales	Operating income from sales	OI – Other income items
Core OI from sales	Core operating income from sales	OI from sales – UOI
RF	Risk-free rate	One-year Treasury bill yield. We use the average yield for the first month of the corresponding calendar year of DGS1.
IO	Implicit interest charge on current liabilities	RF × (OL – txditc)

Panel C. Other Variables

Panel C.1. Hou et al. (2012) Variables

Variables	Definition
Earnings	ib
Negative Earnings	Indicator for ib < 0
Total Assets	at
Dividends	Dvc + dvp
Dividends Indicator	Indicator for Dividends >0
Accruals	1988 onwards: ib – oancf
	Before 1988: Δ (act – ch) – Δ(lct – dlc – txp + dp)

Panel C.2. Returns Variables

Variables	Definition
Returns	The 12-month stock return starting three months after the end of the period in which a forecast is made. For example, if 12/31/2018 data are used to forecast ROCE as at 12/31/2019, the 12-month return would be from 4/1/2019 to 3/31/2020.
SMB, HML, RMW, CMA, UMD	Calculated as a firm-year variable by regressing a firm's monthly returns on excess market returns and five additional monthly factors (SMB, HML, RMW, CMA, UMD) over the previous five years. For example, to calculate the 2019 Beta and factors for a given firm, one

(*continued on next page*)

(*continued*)

Variables	Definition
	would regress the monthly return of the firm from January 1, 2015 to 12/31/2019 on excess market returns and the given monthly factors (SMB, HML, RMW, CMA, UMD). The Beta and factor coefficients are used as that firm-year's Beta and factors. We require each firm-year to have at least 55 prior months of returns over the past five years. Excess market returns and monthly factors are given by the WRDS Fama French factor database.

Panel C.3. Analyst Forecast Variables

Variables	Definition
Mean (Median) Analyst Forecast	Formula: GAAP IBES EPS × GAAP IBES Shares/CSE. The mean (median) GAAP analyst forecast of EPS is the first forecast after nine months before the forecasted fiscal year-end. For example, if the fiscal year end is 12/31/2019, the analyst forecast is the first analyst forecast in April 2019 (if available). GAAP IBES shares are derived on the same date as the forecast. We require that the IBES GAAP EPS forecast be within $0.01 of epsfi in Compustat. We derive IBES GAAP net income as GAAP IBES EPS × GAAP IBES shares, which we scale by CSE to create an IBES forecast of ROCE.

Panel C.4. Dupont Variables

Variables	Definition
Leverage	NOA/CSE
Profit Margin	ROCE × CSE/sale
Asset Turnover	Sale/NOA

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jacceco.2025.101805.

References

Anand, V., Brunner, R., Ikegwu, K., Sougiannis, T., 2020. Predicting profitability using machine learning. Working paper.
Ball, R., Brown, P., 1967. Some preliminary findings on the association between the earnings of a firm, its industry, and the economy. J. Account. Res. 5, 55–77.
Ball, R., Sadka, G., Sadka, R., 2009. Aggregate earnings and asset prices. J. Account. Res. 47 (5), 1097–1133.
Barnea, A., Ronen, J., Sadan, S., 1976. Classificatory smoothing of income with extraordinary items. Account. Rev. 51 (1), 110–122.
Bertomeu, J., 2020. Machine learning improves accounting: discussion, implementation and research opportunities. Rev. Account. Stud. 25 (3), 1135–1155.
Bertomeu, J., Cheynel, E., Liao, Y., Milone, M., 2025. Using machine learning to measure conservatism. Manag. Sci. 71 (2), 1504–1522.
Binz, O., 2022. Managerial response to macroeconomic uncertainty: Implications for firm profitability. Account. Rev. 97 (5), 89–117.
Binz, O., Joos, P., Kubic, M., 2023. The income statement channel of monetary policy. Working paper.
Binz, O., Mayew, W.J., Nallareddy, S., 2022. Firms' response to macroeconomic estimation errors. J. Account. Econ. 73 (2–3).
Bonsall, S.B., Bozanic, Z., Fischer, P.E., 2013. What do management earnings forecasts convey about the macroeconomy? J. Account. Res. 51 (2), 225–266.
Callen, J.L., Kwan, C.C., Yip, P.C., Yuan, Y., 1996. Neural network forecasting of quarterly accounting earnings. Int. J. Forecast. 12 (4), 475–482.
Cao, K., You, H., 2023. Fundamental analysis via machine learning. Financ. Anal. J. 80 (2), 74–98.
Chen, X., Cho, Y.H., Dou, Y., Lev, B., 2022. Fundamental analysis of xbrl data: a machine learning approach. J. Account. Res. 60 (2), 467–515.
Dye, R.A., 2002. Classifications manipulation and nash accounting standards. J. Account. Res. 40 (4), 1125–1162.
Easton, P.D., Kapons, M., Kelly, P., Neuhierl, A., 2020. Beating a random walk. Working paper.
Esplin, A., Hewitt, M., Plumlee, M., Yohn, T.L., 2014. Disaggregating operating and financial activities: Implications for forecasts of profitability. Rev. Account. Stud. 19 (1), 328–362.
Fairfield, P.M., Yohn, T.L., 2001. Using asset turnover and profit margin to forecast changes in profitability. Rev. Account. Stud. 6 (4), 371–385.
Fama, E.F., French, K.R., 1997. Industry costs of equity. J. Financ. Econ. 43 (2), 153–193.
Gerakos, J., Gramacy, R., 2013. Regression-based earnings forecasts. Working paper.
Givoly, D., Hayn, C., D'Souza, J., 2000. Measurement errors and information content of segment reporting. Rev. Account. Stud. 4 (1), 15–43.
Hou, K., Van Dijk, M.A., Zhang, Y., 2012. The implied cost of capital: a New approach. J. Account. Econ. 53 (3), 504–526.
Hou, K., Xue, C., Zhang, L., 2020. Replicating anomalies. Rev. Financ. Stud. 33 (5), 2019–2133.
Hunt, J.O., Myers, J.N., Myers, L.A., 2022. Improving earnings predictions and abnormal returns with machine learning. Account. Horiz. 36 (1), 131–149.
Johnson, T.H., Kaplan, R.S., 1987. Relevance Lost: the Rise and Fall of Management Accounting. Havard Business School Press, Boston, Massachusetts.
Kinney, M., Trezevant, R., 1997. The use of special items to manage earnings and perceptions. J. Financ. Statement Anal. 3, 45–53.
Lee, C.M., Shi, T.T., Sun, S.T., Zhang, R., 2024. Production complementarity and information transmission across industries. J. Financ. Econ. 155, 103812.
Li, K.K., Mohanram, P., 2014. Evaluating cross-sectional forecasting models for implied cost of capital. Rev. Account. Stud. 19 (3), 1152–1185.
Liu, M., 2022. Assessing human information processing in lending decisions: a machine learning approach. J. Account. Res. 60 (2), 607–651.
McVay, S.E., 2006. Earnings management using classification shifting: an examination of core earnings and special items. Account. Rev. 81 (3), 501–531.
Michaud, E.J., Liu, Z., Tegmark, M., 2023. Precision machine learning. Entropy 25 (1), 175.
Monahan, S.J., 2018. Financial statement analysis and earnings forecasting. Foundations and Trends in Accounting 12 (2), 105–215.
Nissim, D., 2024. Earnings quality, fundamental analysis and valuation. Working paper.
Nissim, D., Penman, S.H., 2001. Ratio analysis and equity valuation: from research to practice. Rev. Account. Stud. 6 (1), 109–154.
Nissim, D., Penman, S.H., 2003. Financial statement analysis of leverage and how it informs about profitability and price-to-book ratios. Rev. Account. Stud. 8 (4), 531–560.

Palepu, K.G., Healy, P.M., 2012. Business Analysis and Valuation, 5 ed. Cengage Learning.
Penman, S., 2010. Accounting for Value. Columbia University Press.
Penman, S.H., 2012. Financial Statement Analysis and Security Valuation. McGraw-Hill, New York.
Penman, S.H., Zhang, X.-J., 2002. Accounting conservatism, the quality of earnings, and stock returns. Account. Rev. 77 (2), 237–264.
Poulinakis, K., Drikakis, D., Kokkinakis, I.W., Spottswood, S.M., 2023. Machine-learning methods on noisy and sparse data. Mathematics 11 (1), 236.
Shwartz-Ziv, R., Armon, A., 2022. Tabular data: deep learning is not all You need. Inf. Fusion 81, 84–90.
Sloan, R.G., 2019. Fundamental analysis redux. Account. Rev. 94 (2), 363–377.
Soliman, M.T., 2008. The use of Dupont analysis by market participants. Account. Rev. 83 (3), 823–853.
Sommers, G., Easton, P., Drake, P., 2021. Valuation Using Financial Statements, 2 ed. LLC, Cambridge Business Publishers.
van Binsbergen, J.H., Han, X., Lopez-Lira, A., 2023. Man versus machine learning: the term structure of earnings expectations and conditional biases. Rev. Financ. Stud. 36 (6), 2361–2396.
Wahlen, J., Baginski, S., Bradshaw, M., 2018. Financial reporting. In: Learning, C. (Ed.), Financial Statement Analysis and Valuation: A Strategic Approach, p. 9.
Yohn, T.L., 2020. Research on the use of financial statement information for forecasting profitabiilty. Account. Finance 60, 3163–3181.