Estimating Profitability Decomposition Frameworks via Machine Learning: Implications for Earnings Forecasting and Financial Statement Analysis

Oliver Binz
ESMT Berlin
Schlossplatz 1, 10178 Berlin, Germany
oliver.binz@esmt.org

Katherine Schipper
Duke University
Fuqua School of Business
100 Fuqua Drive, Durham, NC 27708, USA
katherine.schipper@duke.edu

Kevin R. Standridge
Duke University
Fuqua School of Business
100 Fuqua Drive, Durham, NC 27708, USA
kevin.standridge@duke.edu

July 2024

We thank Dan Bens, Qi Chen, Wilbur Chen (discussant), Stephanie Cheng, Scott Dyreng, Mory Elsaify, John Graham, Luzi Hail, 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), 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?"

Estimating Profitability Decomposition Frameworks via Machine Learning: Implications for Earnings Forecasting and Financial Statement Analysis

Abstract

We find that nonlinear estimation of profitability decomposition frameworks yields more accurate out-of-sample profitability forecasts than forecasts from 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 teaching and practice of fundamental analysis and find robust evidence that higher levels of decomposition and focusing on core items improve forecast accuracy, and mixed evidence supporting the use of a long series of historical information. The benefits are pronounced for firms with extreme profitability levels and during the growth, maturity, decline, and shakeout stages of firms' lifecycles. We find that our forecasts predict returns and profitability changes before and after controlling for analyst forecasts and state-of-the-art asset pricing factors.

JEL Classification: C53, G10, M41

Keywords: Financial Statement Analysis, Machine Learning, Earnings Forecasting

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., Penman 2012; Palepu and Healy 2012; Wahlen, Baginski, and Bradshaw 2018; Yohn 2020; Sommers, Easton, and Drake 2021). Our analysis uses Nissim and Penman's (2001; 2003) (hereafter NP) profitability decomposition framework, which is pervasive in textbooks and is cited by Monahan (2018, p. 168) as the authoritative work on the issue. We use machine learning to estimate NP's 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 teaching and practice of financial statement analysis, we analyze several financial statement analysis 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 both because it is an arithmetically sound (tautological) structure of accounting equalities and because it uses an intentionally restricted information set that reveals the hierarchical structure of relations between profitability and its accounting drivers. We apply neural-network estimation, a flexible and popular machine learning algorithm, to find the functional form that uses this information most effectively.[1] A structured restricted-information approach can outperform, in out-of-sample tests,

[1] Using machine learning algorithms requires a number of design choices. Appendix B summarizes our design choices.

data mining approaches applied to very large sets of predictors with few or no restrictions beyond data availability (Bertomeu, Cheynel, Liao, and Milone 2021; Liu 2021). Regardless of whether structured, restricted-information 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 the recent 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). Furthermore, a hierarchical profitability decomposition framework facilitates systematic analysis of the effects of variation in design choices, specifically, the granularity of profitability driver disaggregation, the treatment of transitory/non-core items, and the amount of historical information to use. Put another way, at the possible cost of sacrificing an unknown amount of predictive ability, the structured, restricted-information approach allows us to provide insights into certain choices financial statement users typically make on empirical grounds.

We first confirm many of NP's findings for their sample period (1963-1999) for our longer sample period (1963-2022). We show that current-period ratios are individually and interactively associated with future profitability in nonlinear ways, including contorted S-shaped, U-shaped, and concave patterns. 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 neural networks, a machine learning algorithm that automatically 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 neural-network-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, NP's framework, or 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 8.30%, suggesting that nonlinear estimation is important for profitability forecasting. This improvement appears large at face value, and appears even more significant 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 error that would be predictable if one had access to the true model connecting future profitability to all known 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]

[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).

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 linear estimation of NP's framework does not improve forecasting performance; in fact, performance deteriorates. We also find that nonlinear estimation of NP's framework significantly improves forecast accuracy, and 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 model decreases absolute forecast errors by 2.55%, suggesting meaningful synergies between nonlinear estimation and the use of profitability decomposition frameworks.

Our inferences are robust to several alternative research design choices, including truncating rather than winsorizing the data and employing alternative machine learning algorithms. 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. Cross-sectional analyses show the forecast-accuracy benefits are pervasive, not concentrated in subsamples of firms for which forecasting is particularly difficult, thereby demonstrating the information loss from estimating an inherently nonlinear profitability decomposition framework using linear methods. Further, the benefits are pronounced for firms with extreme profitability levels and during the growth, maturity, decline, and shakeout stages of

---

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).

firms' lifecycles (but not the introduction stage).

Extending the profitability decomposition results, we examine how two essential financial statement analysis design choices that must be made on empirical grounds in applying NP's framework affect forecast accuracy. First, we find that focusing on core items (i.e., items more likely to recur) improves forecast accuracy. Second, we find mixed evidence that using more historical information (up to three years) improves forecast accuracy. While using one additional year of financial statement data beyond the most recent year for which financial statements are available improves performance in deciles 1 through 9 of the absolute forecast error distribution, the improvements for decile 10 and for the full sample are not significant at conventional levels. Results are similar using two-years lagged data; using three-years lagged data, performance does not improve for deciles 5, 6 and 10 and for the full sample. This evidence points to increasing staleness in financial statement information, suggesting that adding more historical information beyond one or two years does not provide incremental benefits for 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, we test whether the forecasts derived from nonlinear estimation of NP's profitability decomposition framework predict stock returns and changes in profitability before and after controlling for state-of-the-art asset pricing factors and analyst forecasts. After controlling for the asset pricing factors in Lee, Shi, Sun, and Zhang (2024) and the consensus analyst forecast, we find that a one-standard-deviation change in the forecast obtained from nonlinear estimation of NP's framework is associated with a 0.071 (0.589) standard-deviation change in year-ahead stock returns (change in profitability), suggesting that investors and analysts could use the methodology outlined in this paper to enhance trading profits and profitability forecast accuracy.

Our research contributes to two literatures. The first analyzes structural accounting-based profitability decomposition frameworks, in particular, NP's. We extend this literature in four distinct and related ways. First, although NP both argue and demonstrate that evaluating their model requires accommodating its essential nonlinear structure, prior research including Fairfield and Yohn (2001), Soliman (2008), and Esplin, Hewitt, Plumlee, and Yohn (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 the nonlinearities in profitability decomposition frameworks. Second, we find strong evidence, as do Esplin et al. (2014), that focusing on core (i.e., likely recurring) items improves forecast accuracy. Third, consistent with recent findings that using more data does not always increase forecasting performance (Gallo, Labro, and Omartian 2023), we find mixed evidence supporting the use of historical information up to three years preceding the latest financial statements available. Fourth, we provide cross-sectional evidence that the benefits of nonlinear estimation of profitability decomposition frameworks are pervasive and particularly pronounced for extreme profitability outcomes and during the growth, maturity, decline, and shakeout stages of firms' lifecycles. This finding complements Vorst and Yohn's (2018) result that the relative performance of autoregressive profitability forecasting models (i.e., models that use limited financial statement information) is most pronounced during the

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

introduction stage, and is consistent with Anderson, Hyun, Muslu, and Yu's (2024) result that the usefulness of Dupont analysis increases for more mature firms.

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, Kwan, Yip, and Yuan (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), Hunt, Myers, and Myers (2019), Anand, Brunner, Ikegwu, and Sougiannis (2020), Cao and You (2020), van Binsbergen, Han, and Lopez-Lira (2022), and Chen, Cho, Dou, and Lev (2022), 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 empirical and atheoretical 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 avoid (or at least mitigate) overfitting.

From the perspective of providing insights for teaching and practical applications, a purely empirical, atheoretical approach has two interrelated weaknesses that our approach, grounded in NP's structured profitability decomposition framework, addresses. First, 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., Armstrong (2001) and 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

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

variables that drive the predictive power. A purely empirical, atheoretical approach to understanding these causal 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. 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 an atheoretical approach with a nonlinear algorithm like a neural network can make it impossible for a researcher to determine the incremental value of any one variable because the actual degree of collinearity is unknown. 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. That is, as shown in Figure 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 and teaching as to how best to use structural profitability decomposition frameworks.

Our paper proceeds as follows. Section 2 discusses profitability decomposition frameworks and neural networks. 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 and Appendix B discusses the hyperparameter choices that underlie our neural network 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 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 (Figure 1):

Level 1. ROCE = ROTCE × MSR: ROTCE denotes return on total common equity (= (CNI + MII)/(CSE + MI)), MSR minority sharing ratio (= $\frac{\text{CNI/(CNI + MII)}}{\text{CSE/(CSE + MI)}}$), 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):

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

Equation (1) illustrates that the predictors in NP's framework are nonlinearly linked to ROCE. Analyses reported by NP and in our Section 3.3 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: neural networks, a machine learning algorithm.

2.2. Neural Networks

A neural network generalizes estimators such as OLS to model complex nonlinear relations among independent and (possibly multiple) dependent variables through a layered system of equations. The basic building block of a neural network is a neuron, a function that takes in variables as inputs, combines them through a linear equation, and transforms the output of that equation through a (typically) nonlinear function known as an activation function. OLS and logistic regression are examples of single neurons that create a linear relation among independent variables and transform the relation by multiplying by one and $(1 + e^x)^{-1}$, respectively.

A neural network organizes relations among inputs into layers. Each layer, including the input layer (independent variables) and output layer (dependent variables), consists of a series of neurons. Layers between the input and output layers are hidden layers. A neural network with one (multiple) hidden layer(s) is referred to as shallow (deep). While the connections between layers can be set arbitrarily, the most common implementations are sequential models (feedforward neural networks) that fully connect each neuron in the preceding layer to each neuron in the next layer. In other words, the neurons of one layer become the independent variables that are the inputs for the neurons in the next layer. We use a fully connected sequential model with multiple variables in the input layer, constant activation functions, and a fixed number of neurons in each hidden layer. By combining the activation functions with a series of layers, the neural network can model complex nonlinear and interactive relations without the need to specify a functional form. While in principle the same outcome is achievable by including higher-order polynomials and interaction terms in linear regression models, the dimensionality of this problem quickly makes estimation infeasible. A simple linear regression with 10 independent variables would estimate 1 + 10 = 11 parameters. Including the squares and cubes of each independent variable increases the number of parameters to be estimated to 31 [= 1 + 10 + 10 + 10]. Including interactions among these 30 independent variables increases the number of parameters to 1 + 10 + 10 + 10 + 29! = 8.84 × e^{30}. Once the number of parameters exceeds the number of observations in the dataset the model is not estimable via OLS. In contrast, neural networks allow researchers to capture higher-order and interactive relations without explicitly specifying them. The combination of hidden layers connected through nonlinear activation functions approximates such relations automatically (Hornik, Stinchcombe, and White 1989; Cybenko 1989; Tsang, Cheng, and Liu 2017), reducing computing and implementation time, limiting subjective research design choices, and making neural networks prime candidates for modeling the complex nonlinear relations between the past and future profitability-determining fundamentals discussed in the previous section. Appendix B describes our machine learning design choices and reports results of robustness tests.

2.3. Forecasting Procedure

We use neural network estimation to predict out-of-sample ROCE using the ratios from each level of the NP framework as independent variables.[5] To generate an annual forecast, we train the neural network on 10 years of lagged data and generate forecasts for the 11th year. Because our training data begin in 1964 (i.e., 1963 is the first year in which Compustat is free from survivorship 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 neural network 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 2022.

3. Data and Descriptive Evidence

3.1. Data

We obtain annual data for 1963 to 2022 from Compustat. We require all firm-years to have non-missing values for all variables within NP's framework in the current year and for ROCE one

[5] In untabulated analyses, we compare this approach to three alternatives. Specifically, we obtain ROCE forecasts by combining the predictions for each Level 4 component obtained from (1) estimating autoregressive models via separate neural networks, (2) estimating multivariate models that include all variables identified in NP's framework as predictors via separate neural networks, and (3) estimating multivariate models that include all variables identified in NP's framework as predictors jointly via a single neural network. All three approaches yield larger absolute forecast errors than our baseline approach on average and across the absolute forecast error distribution.

year in the future. To ensure sufficient liquidity for the returns analysis, we follow Hou, Xue, and Zhang (2020) and require firms to have a market capitalization above the 20th percentile of NYSE firms, an SIC code, and assets exceeding $10 million. Following NP, we require non-negative values for CSE, NOA, OA, and OL at beginning and end of the year of observation and winsorize all ratios in the NP framework by year at the 1st and 99th percentiles.

3.2. Summary Statistics and Correlations

Table 1 presents descriptive statistics. Although our sample is considerably larger than NP's, our descriptive statistics are generally similar to those they report. Average *ROCE* exceeds average *RNOA*, which indicates the average firm uses financial leverage successfully to increase profitability for its shareholders. Firms tend to use more operating than financial leverage, but operating spreads also tend to be smaller than financial spreads. *Other Items/NOA* and *Other Items/OA* are small, with some large positive outliers at the 99th percentiles. Table 2 presents correlations for selected ratios, with Pearson (Spearman) correlations below (above) the diagonal. Several correlations of profitability drivers with *ROCE* and *RNOA* are economically and statistically significant, suggesting the predictive usefulness of NP's framework. *ROCE* and *ROTCE* are nearly perfectly correlated, while *RNOA* exhibits idiosyncratic variation, with Pearson (Spearman) correlation with *ROCE* and *ROTCE* equal to 0.38 (0.78). Margins are more strongly correlated with contemporaneous profitability than turnovers. Except for current *ROCE* and *ROTCE*, *OLSPREAD* is most highly correlated with future *ROCE*. In contrast, *OLLEV*'s correlations with other profitability components are generally low.

3.3. Visual Evidence of Nonlinearities and Interactive Relations

Panels A to H of Figure 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 Figure 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 Figures 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 neural networks (hereafter, NN) the appropriate estimation tool.

4. Main Results

We first analyze whether linear OLS and nonlinear NN estimation of a simple autoregressive process yield more accurate out-of-sample year-ahead profitability forecasts than a random walk.[6] 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

[6] In untabulated analyses, we find that our inferences are similar for two- and three-year-ahead profitability forecasts.

estimation of ROCE's autoregressive process yield smaller average out-of-sample absolute forecast errors than a random walk for the full sample. While linear estimation yields higher absolute forecast errors than a random walk, nonlinear estimation yields significantly (at the 0.01 level) lower absolute forecast errors, indicating that NN effectively captures the nonlinear autoregressive relations visible in Figure 2 Panel A. In terms of economic magnitude, relative to a random walk, NN estimation decreases average ROCE absolute forecast errors by a meaningful 1.10% (= 0.1215 – 0.1325) of ROCE in absolute terms and by 8.30% (= 0.1215/0.1325 – 1) in relative terms. Figure 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 average 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 in deciles 9 and 10 of the absolute forecast error distribution, absolute forecast errors are larger in deciles 1 through 8. 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 very 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. Figure 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 ratio decomposition further increases forecast accuracy. Table 4 presents the results for OLS estimation (Panel A) and nonlinear NN 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 4, 5 and 7 of the Level 1 disaggregation (splitting ROCE into the parts 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, average 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 RNOA and Net Borrowing Cost), and Level 4 (splitting out operating leverage). 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 explanation is the nonlinear relation between current and future ratios, making accommodating nonlinearities imperative in 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 0.31% [= 0.1184 – 0.1215] of ROCE in absolute terms and by 2.55% [= 0.1184/0.1215 – 1] in relative terms. Figure 6 illustrates this result. The improvements largely derive from smaller absolute forecast errors in deciles 7 through 10 of the absolute forecast error distribution, providing evidence that NP's 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 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. Truncation

In our main tests, we follow NP and winsorize the data at the 1st and 99th percentiles.[7] In Table 5, we test whether our inferences are robust to using truncation. We find that this is the case. Panel A shows that nonlinear estimation improves performance over a random walk for the full sample and for deciles 3 through 10 while linear estimation does not, and Panel B shows that average absolute forecast errors decrease when moving from NN Level 0 to NN Level 3 or Level 4, both for the full sample and for deciles 7 through 10.

---

[7] NP winsorize rather than truncate their sample because "because trimming results in an excessive reduction in the sample" (footnote 14). Consistent with this statement, truncating rather than winsorizing the sample for the Table 5 analysis eliminates 13,995 (= 103,416 – 117,411) observations or about 12% of our sample.

5.2. Alternative Machine Learning Approaches

Panels A and B of Table 6 test the robustness of our inferences to three alternative machine learning approaches. First, we estimate our models using Random Forests and Gradient Boosted Trees, two commonly used machine learning algorithms. Second, to address a possible overfitting concern that arises because firms used in the training sample are also used in the cross-validation sample, we use holdout estimation that repeats the cross-validation after excluding 20% of all firms from the training sample; we evaluate model performance on these holdout firms. We find that Gradient-Boosted Tree nonlinear estimation and the holdout-sample approach yield lower full-sample absolute ROCE forecast errors than a random walk and that nonlinear estimation of NP's framework further improves performance.

5.3. Additional Predictors

Table 7 reports results of tests that analyze whether including five sets of additional predictors not identified in NP's framework improves the accuracy of our models. First, we include profit margins, asset turnovers, and leverage as identified by a simple Dupont analysis. Second, we include the variables in Hou, Van Dijk, and Zhang's (2012) profitability forecasting model (Hou). Third, building on the notion of fundamental industry-specific differences in firms' profitability and profitability components, we include Fama-French 48-industry fixed effects (Industry).[8] Fourth, we include real GDP growth, inflation, and the unemployment rate, the macroeconomic aggregates that receive most attention by macroeconomic forecasters as evidenced by their extensive coverage in US Federal Reserve communications and databanks (Macro). Fifth, we include the Dupont, Hou, Industry, and Macro variables jointly (Joint). Panel A [Panel B]

[8] In untabulated analyses, we find that our inferences are similar when we use the Fama-French 10-industries classification or estimate our models by (Fama-French 10) industry rather than including industry fixed effects.

shows the results when we use the variables in these models only [in combination with the variables identified in Level 4 of NP's framework] as predictors. We generally do not find evidence that including these alternative sets of predictors increases forecast performance. None of them yields statistically lower (at conventional levels) absolute forecast errors for the full sample.

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:

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

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, Ronen, and Sadan 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 8 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 average absolute forecast errors are significantly (at the 0.05 level) smaller than those of the baseline model for the full sample. Figure 7 Panel A illustrates this result. These improvements derive from medium to large absolute forecast errors in deciles 4 through 9 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 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.[9] 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 9, do not show that including additional lags of historical data significantly (at conventional levels) decreases average absolute forecast errors for the full sample. We find some evidence that including *one* lag of historical data decreases forecast errors (significantly, at the 0.01 level) in deciles 1 through 9 of the absolute forecast error distribution. Including two lags of data appears to have little effect on performance, relative to results based on one lag, while including three lags appears to deteriorate performance. Figure 7 Panel B illustrates this result for the full sample. One explanation is that financial statement information becomes stale after three years and does not provide incremental information beyond the information contained in the two most recent annual financial statements. Viewed as a whole,

---

[9] See, for example, Brown and Ball (1967), Ball, Sadka, and Sadka (2009), Bonsall, Bozanic, and Fischer (2013), Binz, Mayew, and Nallareddy (2022), Binz (2022), and Binz, Joos, and Kubic (2023).

these results provide mixed evidence that using more past data improves forecast performance.

6.3. Cross-Sectional Analysis

To shed light on the contexts in which nonlinear estimation of profitability decomposition frameworks is especially useful, we examine how our results vary with two firm-level variables that likely make forecasting more difficult: extreme profitability outcomes and variation in firms' lifecycles.

6.3.1. Extreme Profitability Outcomes

As illustrated in Figure 2, mean reversion in extreme ROCE deciles is greater than in middle deciles. Building on this visual evidence, Table 10 Panel A shows average absolute forecast errors from a random walk and from nonlinear NN estimation of the Level 4 decomposition of NP's framework separately for observations within the 10^{th} and 90^{th} ROCE percentiles (Interior) and for observations below the 10^{th} or above the 90^{th} ROCE percentile (Extreme). We find that nonlinear estimation of NP's framework yields lower average absolute forecast errors than a random walk for extreme as well as interior observations, and the improvement is greater for extreme observations. Figure 8 Panel A illustrates this result. Further, while we find significant (at the 0.01 level) improvements across deciles 2 through 10 of the absolute forecast error distribution for extreme observations, we find improvements only for deciles 6 through 9 of interior observations. This evidence suggests that one channel through which nonlinear estimation of NP's framework improves performance over a random walk is through its more efficient handling of relatively extreme profitability outcomes.

6.3.2. Corporate Lifecycle Stage

Mirroring the structure of Panel A, Table 10 Panel B analyzes how average absolute forecast errors vary across Dickinson's (2011) Introduction, Growth, Maturity, Decline, and

Shakeout lifecycle stages.[10] Nonlinear estimation of NP's framework yields significantly (at the 0.01 level) lower absolute forecast errors than a random walk for the full sample during the Growth, Maturity, Decline, and Shakeout stages but not during the Introduction stage, possibly because Introduction stage firms have not accumulated a sufficiently stable operating history for analysis of their historical data to yield precise forecasts. However, for all lifecycle stages including the Introduction stage, we find that nonlinear estimation of NP's framework improves performances for larger forecast error-deciles of the absolute forecast error distribution. Figure 8 Panel B illustrates these results for the full sample. In sum, these findings suggest the benefits of using nonlinear NN estimation and NP's framework are pervasive across most lifecycle stages, particularly for larger absolute forecast errors.

6.4. Can Forecasts from Nonlinear Estimation of NP's Profitability Framework Inform Investors and Analysts?

We analyze 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. Our analysis has three distinct and related components.

First, following Chen et al. (2022), Table 11 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 NN Level 4 earnings forecast and the current earnings realization

[10] Dickinson (2011) follows Gort and Klepper (1982) and defines a shakeout as the life cycle stage "where the number of producers begins to decline" (p. 1971). This definition aligns with those in Klepper and Miller (1995) and Klepper and Simons (2005). Based on her assessment that life cycle theory does not provide predictions for how CFO, CFI, and CFF should behave during the shakeout stage (see her Table 1), in Footnote 7, she states that she measures the shakeout stage as an indicator for cash flow patterns that do not neatly map into one of the other life stages. However, in her empirical analysis, she finds some evidence that her measurement of the shakeout stage aligns with life cycle theory's predictions that firms in the shakeout stage are liquidating net operating assets (Table 2 Panel B), that firms transition the shakeout stage relatively quickly (Table 3 Panel B), and that firms in the shakeout stage experience lower future profitability (Table 5 "CF Pattern" columns).

(*NN Level 4*), firm fixed effects, and month-year fixed effects before and after including Lee et al.'s (2024) six asset pricing factors. We cluster standard errors by firm and year and standardize all variables to facilitate interpretation. We find evidence that the forecasts generated through nonlinear estimation of NP's framework predict returns. *NN Level 4*'s slope coefficient is reliably (at the 0.01 level) positive before and after controlling for Lee et al.'s (2004) asset pricing factors. In terms of economic magnitude, the Column (2) estimates suggest that a forecasted one standard-deviation profitability change is associated with a 0.068 standard-deviation change in returns, which is large relative to the magnitude of the slope coefficients of the Lee et al. (2024) factors.

Second, for a small sample of observations for which IBES analyst consensus GAAP earnings forecasts are available, we test whether our forecasts are informative beyond analyst forecasts.[11] Specifically, Table 11, Panel A Columns (3) to (6) repeat the analysis in Columns (1) and (2) including the first mean (*Mean Analyst Forecast*) or median (*Median Analyst Forecast*) consensus analyst ROCE forecasts that become available three months after the fiscal year end. 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.

Third, Table 11 Panel B examines a potential mechanism or channel through which our forecasts predict returns incrementally to analyst forecasts, namely, incremental profitability forecast information. Specifically, we test whether our forecasts continue to predict year-ahead changes in ROCE (Δ*ROCE*) after controlling for the mean or median consensus analyst forecast. We find strong evidence that this is the case. For example, the estimates in Columns (3) to (6)

[11] 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).

show that after controlling for the mean or median consensus analyst forecast a one-standard-deviation change in our model's forecasts predicts a 0.589 standard-deviation change 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 (NP; 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 neural networks, 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 empirically for the best profitability predictors, as might be done, for example, by applying machine learning to a large set of predictors atheoretically, without a framework that specifies 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 teaching and 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.

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 neural networks 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 truncating rather than winsorizing the data and 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 and a focus on core items improve forecast accuracy, and we find mixed evidence that using more historical information (beyond two lags) improves forecast accuracy. Cross-sectional analyses suggest the benefits of nonlinear estimation are pervasive, not concentrated in small pockets of firms. Nonlinear estimation of profitability decomposition frameworks provides forecasting advantages for firms with extreme profitability outcomes and during growth, maturity, decline, and shakeout lifecycle stages. Finally, we find that that our forecasts predict returns and changes in profitability before and after controlling for analyst forecasts and state-of-the-art asset pricing factors, suggesting that the kinds of analyses in this paper could be useful to analysts and investors.

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		
3	Net Borrowing Costs	*NBC*	NFE/NFO	*Core NBC*	Core NFE/NFO
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*	NBC/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 Shareholder's 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 (2018-2022).
CORE NFE	Core Net Financial Expense	xint × (1 – MTR) + dvp – idit × (1 – MTR)
UFE	Unusual Financial Expense	Δmsa
NFE	Net Financial Expense	Core NFE + UFE
CSA	Clean Surplus Adjustment	Δmsa + Δrecta
CNI	Comprehensive Net Income	ni – dvp + CSA
OI	Operating Income	NFE + CNI + mii
UOI	Unusual Operating Income	(1 – MTR) × (nopio + spi – esub) + xido + Δrecta
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
IO	Implicit Interest Charge on Current Liabilities	RF × (OL – txditc)

Panel C. Other Variables

Panel C.1. Cross-Sectional Variables

Variables	Definition
Extreme	Indicator for whether an observation has a current ROCE value above the 90th or below the 10th percentile in a given year.
Interior	Indicator for whether an observation has a current ROCE value that is within the 10th and 90th percentile in a given year.
Introduction	Indicator for whether a given observation falls within the Introduction category of the Dickinson (2011) life cycle framework. The Introduction category is defined as: Operating Cash Flow < 0, Investing Cash Flow < 0, Financing Cash Flow > 0.
Growth	Indicator for whether a given observation falls within the Growth category of the Dickinson (2011) life cycle framework. The Growth category is defined as: Operating Cash Flow > 0, Investing Cash Flow < 0, Financing Cash Flow > 0.
Maturity	Indicator for whether a given observation falls within the Maturity category of the Dickinson (2011) life cycle framework. The Maturity category is defined as: Operating Cash Flow > 0, Investing Cash Flow < 0, Financing Cash Flow < 0.
Decline	Indicator for whether a given observation falls within the Decline category of the Dickinson (2011) life cycle framework. The Decline category is defined as: Operating Cash Flow < 0, Investing Cash Flow > 0.
Shakeout	Indicator for whether a given observation falls within the Shakeout category of the Dickinson (2011) life cycle framework. The Shakeout category is a residual category, comprising observations not contained in one of the other four categories.

Panel C.2. 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 to 2022: ib – oancf Before 1988: Δ (act – ch) – Δ(lct – dlc – txp + dp)

__Panel C.3. 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 3/31/2019 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 would regress the monthly return of the firm from 1/1/2015 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.4. 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. 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.5. Dupont Variables__

Variables	Definition
Leverage	NOA/CSE
Profit Margin	CNI/sale
Asset Turnover	sale/NOA

References

Anand, V., R. Brunner, K. Ikegwu, and T. Sougiannis. 2020. Predicting profitability using machine learning. Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3466478.

Anderson, M., S. Hyun, V. Muslu, and D. Yu. 2024. Earnings prediction with DuPont components and calibration by life cycle. *Review of Accounting Studies* 29 (2):1456-1490.

Armstrong, J. S. 2001. *Principles of forecasting: A handbook for researchers and practitioners*. Vol. 30: Springer Science & Business Media.

Ball, R., G. Sadka, and R. Sadka. 2009. Aggregate earnings and asset prices. *Journal of Accounting Research* 47 (5):1097-1133.

Barnea, A., J. Ronen, and S. Sadan. 1976. Classificatory smoothing of income with extraordinary items. *The Accounting Review* 51 (1):110-122.

Bertomeu, J. 2020. Machine learning improves accounting: discussion, implementation and research opportunities. *Review of Accounting Studies* 25 (3):1135-1155.

Bertomeu, J., E. Cheynel, Y. Liao, and M. Milone. 2021. Using Machine Learning to Measure Conservatism. Working paper.

Binz, O. 2022. Managerial Response to Macroeconomic Uncertainty: Implications for Firm Profitability. *The Accounting Review* 97 (5):89-117.

Binz, O., P. Joos, and M. Kubic. 2023. How Does Monetary Policy Affect Corporate Earnings? Working Paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3835653.

Binz, O., W. J. Mayew, and S. Nallareddy. 2022. Firms' Response to Macroeconomic Estimation Errors. *Journal of Accounting and Economics* 73 (2-3).

Bonsall, S. B., Z. Bozanic, and P. E. Fischer. 2013. What do management earnings forecasts convey about the macroeconomy? *Journal of Accounting Research* 51 (2):225-266.

Brown, P., and R. Ball. 1967. Some preliminary findings on the association between the earnings of a firm, its industry, and the economy. *Journal of Accounting Research* 5:55-77.

Callen, J. L., C. C. Kwan, P. C. Yip, and Y. Yuan. 1996. Neural network forecasting of quarterly accounting earnings. *International Journal of Forecasting* 12 (4):475-482.

Cao, K., and H. You. 2020. Fundamental Analysis Via Machine Learning. Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3706532.

Chen, X., Y. H. Cho, Y. Dou, and B. Lev. 2022. Fundamental Analysis of XBRL Data: A Machine Learning Approach. *Journal of Accounting Research* 60 (2):467-515.

Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* 2 (4):303-314.

Dickinson, V. 2011. Cash flow patterns as a proxy for firm life cycle. *The Accounting Review* 86 (6):1969-1994.

Dye, R. A. 2002. Classifications manipulation and Nash accounting standards. *Journal of Accounting Research* 40 (4):1125-1162.

Esplin, A., M. Hewitt, M. Plumlee, and T. L. Yohn. 2014. Disaggregating operating and financial activities: Implications for forecasts of profitability. *Review of Accounting Studies* 19 (1):328-362.

Fairfield, P. M., and T. L. Yohn. 2001. Using asset turnover and profit margin to forecast changes in profitability. *Review of Accounting Studies* 6 (4):371-385.

Gallo, L., E. Labro, and J. Omartian. 2023. Overreliance on data in forecasting. *Working Paper*.

Gerakos, J., and R. Gramacy. 2013. Regression-based earnings forecasts. Working paper. Available at *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2112137*.

Givoly, D., C. Hayn, and J. D'Souza. 2000. Measurement errors and information content of segment reporting. *Review of Accounting Studies* 4 (1):15-43.

Gort, M., and S. Klepper. 1982. Time paths in the diffusion of product innovations. *The economic journal* 92 (367):630-653.

Hornik, K., M. Stinchcombe, and H. White. 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2 (5):359-366.

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.

Hou, K., C. Xue, and L. Zhang. 2020. Replicating anomalies. *The Review of Financial Studies* 33 (5):2019-2133.

Hunt, J. O., J. N. Myers, and L. A. Myers. 2019. Improving Earnings Predictions with Machine Learning. Working paper. Available at *https://zicklin.baruch.cuny.edu/wp-content/uploads/sites/10/2019/12/Improving-Earnings-Predictions-with-Machine-Learning-Hunt-Myers-Myers.pdf*.

Johnson, T. H., and R. S. Kaplan. 1987. *Relevance lost: the rise and fall of management accounting*. Boston, Massachusetts: Havard Business School Press.

Kingma, D. P., and J. Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kinney, M., and R. Trezevant. 1997. The use of special items to manage earnings and perceptions. *Journal of Financial Statement Analysis* 3:45-53.

Klepper, S., and J. H. Miller. 1995. Entry, exit, and shakeouts in the United States in new manufactured products. *International journal of industrial organization* 13 (4):567-591.

Klepper, S., and K. L. Simons. 2005. Industry shakeouts and technological change. *International journal of industrial organization* 23 (1-2):23-43.

Lee, C. M., T. T. Shi, S. T. Sun, and R. Zhang. 2024. Production complementarity and information transmission across industries. *Journal of Financial Economics* 155:103812.

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.

Liu, M. 2021. Assessing Human Information Processing in Lending Decisions: A Machine Learning Approach. Working paper. Available at *https://www.proquest.com/docview/2427239457?pq-origsite=gscholar&fromopenview=true*.

McVay, S. E. 2006. Earnings management using classification shifting: An examination of core earnings and special items. *The Accounting Review* 81 (3):501-531.

Monahan, S. J. 2018. Financial Statement Analysis and Earnings Forecasting. *Foundations and Trends in Accounting* 12 (2):105-215.

Nissim, D. 2024. Earnings Quality. Working paper. Available at *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3794378*.

Nissim, D., and S. H. Penman. 2001. Ratio analysis and equity valuation: From research to practice. *Review of Accounting Studies* 6 (1):109-154.

Nissim, D., and S. H. Penman. 2003. Financial statement analysis of leverage and how it informs about profitability and price-to-book ratios. *Review of Accounting Studies* 8 (4):531-560.

Palepu, K. G., and P. M. Healy. 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., and X.-J. Zhang. 2002. Accounting conservatism, the quality of earnings, and stock returns. *The Accounting Review* 77 (2):237-264.
Sloan, R. G. 2019. Fundamental analysis redux. *The Accounting Review* 94 (2):363-377.
Soliman, M. T. 2008. The use of DuPont analysis by market participants. *The Accounting Review* 83 (3):823-853.
Sommers, G., P. Easton, and P. Drake. 2021. *Valuation Using Financial Statements*. 2 ed: Cambridge Business Publishers, LLC.
Tsang, M., D. Cheng, and Y. Liu. 2017. Detecting statistical interactions from neural network weights. *arXiv preprint arXiv:1705.04977*.
van Binsbergen, J. H., X. Han, and A. Lopez-Lira. 2022. Man versus machine learning: The term structure of earnings expectations and conditional biases. *The Review of Financial Studies*.
Vorst, P., and T. L. Yohn. 2018. Life cycle models and forecasting growth and profitability. *The Accounting Review* 93 (6):357-381.
Wahlen, J., S. Baginski, and M. Bradshaw. 2018. *Financial Reporting, Financial Statement Analysis and Valuation: A Strategic Approach*. edited by C. Learning. 9 ed.
Yohn, T. L. 2020. Research on the use of financial statement information for forecasting profitabiilty. *Accounting and Finance* 60:3163-3181.

Appendix B. Hyperparameter Optimization

Unlike OLS, machine learning algorithms require researchers to specify a variety of parameters (hyperparameters) that can affect a model's performance. For example, in a neural network consisting of layers of functions called "neurons," the researcher might specify the number of layers and the number of neurons per layer. Other common hyperparameters in a neural network include an activation function, an optimizer, a learning rate, and a method of regularization.

To determine the best hyperparameters for a given model, machine learning experts typically use hyperparameter optimization, which involves trying out several values for hyperparameters and evaluating which values perform best in the training sample. A researcher does not use the test sample to choose hyperparameters; doing so risks corrupting the test sample with lookahead bias. In our research setting, hyperparameter optimization is important to ensure that models which perform well do so because the financial statement analysis design choices are truly superior and not because the set of financial statement analysis variables is better suited to a given set of hyperparameters. Because this process requires re-estimating the neural network for each (trial) set of hyperparameters, the main disadvantage of hyperparameter optimization is computational intensity.

Our research setting estimates over 1,000 neural networks (over 24 models for nearly 50 years); it is therefore crucial to reduce the set of hyperparameters over which we optimize to reduce computation time. First, we optimize hyperparameters only in 1991, as opposed to annually.[12] Second, we optimize only the number of layers in the neural network and the number of neurons

[12] To prevent researcher bias, an anonymous reviewer selected 1991 as the year for which we conduct the hyperparameter optimization.

per layer, not all hyperparameters.

We optimize the number of layers and the number of neurons as follows. For each model, we randomly select 80% of the observations as a training set and use the remaining 20% as a test set. We then perform a grid search over the number of layers in the network and the number of neurons per layer. We display the grid in Table B.1 Panel A. In Panel B, we display the results of this grid search for the five neural networks analyzed in Tables 3 and 4. We show that the optimal number of layers ranges between 3 and 5 and that the optimal number of neurons ranges between 5 and 50. Both results are within the interior of our grid, suggesting the grid is wide enough to capture the optimal hyperparameters.

There are several hyperparameters that we specify but do not optimize. First, we use the Adam optimizer proposed by Kingma and Ba (2014). Second, we use Adam's default learning rate of 0.001. Third, we use the Leaky ReLu as the activation function because it is the most commonly used activation function for non-classification problems. Leaky ReLu modifies the traditional ReLu activation function to prevent the neural network from developing dead neurons (i.e., neurons caught in a local minimum at 0). Fourth and fifth, we use L-2 regularization and L-2 regularization's default regularization rate of 0.0001. Sixth, we use a batch size of 128. Seventh, we train the data for a minimum of 10 epochs and a maximum of 50 epochs. In this range, we implement early stopping if the network does not improve by at least 0.0001 for 5 epochs.

Our decision not to optimize all hyperparameters may reduce the efficiency of the neural networks. With the substantially greater resources available to many hedge funds and other large investors, it may be possible to derive more precise estimates. Therefore, our results likely understate the potential for improvement from applying machine-learning algorithms to the NP framework. To test this possibility, we proceed in two steps. First, fixing the number of layers at

5 and neurons per layer at 25, we optimize all other hyperparameters over the grid search values listed in Table B.2 Panel A one at a time.[13] Table B.2 Panel B shows the optimized values for each hyperparameter. Second, we re-estimate the results in Table 3 and Table 4 Panel B changing each optimized hyperparameter one at a time while holding all others constant. The results, shown in Table B.3 Panel A and B, are nearly identical to those in Table 3 and Table 4 Panel B, suggesting that changing hyperparameters other than the number of hidden layers and neurons per layer does not alter our inferences.

[13] A preferable approach would be to optimize all hyperparameters jointly as we do for the number of hidden layers and neurons per layer in Table B.1. However, this exercise would involve a grid search over 5 variations in the number of hidden layers × 8 variations in the number of neurons per hidden layer × 5 optimizing algorithms × 5 learning rates × 4 regularization types × 6 regularization rates × 5 activation functions × 7 batch sizes × 4 variations in the number of epochs × (5 levels of ratio disaggregation for Tables 3 and 4 + 5 levels of ratio disaggregation for Table 5 + 4 alterations in additional predictors in Table 7 + separate estimation for core items in Table 8 + 3 alternative lags of data in Table 9) = 60,480,000 cells. Each grid search iteration takes approximately 10 minutes, so optimizing all hyperparameters jointly for our analyses would take approximately 1,150 years.

Table B.1. Optimizing the Number of Hidden Layers and Neurons

Panel A shows the set of hidden layers and number of neurons per layer over which we perform a grid search. Panel B displays the optimal values.

Panel A. Grid Search

Hyperparameter	Grid Search Values
# of Layers	1, 3, 5, 10, 20
# of Neurons	1, 5, 10, 25, 50, 100, 250, 500

Panel B. Chosen Hyperparameters

Hyperparameter	Level 0	Level 1	Level 2	Level 3	Level 4
# of Layers	5	5	5	5	3
# of Neurons	5	25	25	5	50

Table B.2. Optimizing Hyperparameters Other than the Number of Hidden Layers and Neurons

Panel A shows the set of learning rates, regularization rates, activation functions, optimizing algorithms, batch sizes, regularization types, and epochs over which we perform a grid search. Panel B displays the optimal values.

Panel A. Grid Search

Hyperparameter	Grid Search Values
Optimizing Algorithm	SGD, RMSProp, Adam, Adagrad, Adadelta
Learning Rate	0.1, 0.01, 0.001, 0.0001, 0.00001
Regularization Type	L1, L2, 50% Dropout, 20% Dropout
Regularization Rate	0.1, 0.01, 0.001, 0.0001, 0.00001, 0
Activation Function	Leaky ReLu, ReLu, ELU, Sigmoid, Tanh
Batch Size	16, 32, 64, 128, 256, 512, 1024
Number of Epochs	10, 50, 100, 200

Panel B. Chosen Hyperparameters

Hyperparameter	Level 0	Level 4
Optimizing Algorithm	Adam	Adam
Learning Rate	0.0001	0.001
Regularization Type	20% Dropout	20% Dropout
Regularization Rate	0.001	0
Activation Function	ReLu	Leaky ReLu
Batch Size	512	512
Number of Epochs	100	100

Table B.3. Table 3 after Optimizing Hyperparameters Other than the Number of Hidden Layers and Neurons
This table shows the Table 3 results after estimating the underlying neural networks using the optimized hyperparameter values shown in Table B.2 Panel B one at a time.

Error Decile	RW	Optimizing Algorithm	Learning Rate	Activation Function	Regularization Type	Regularization Rate	Batch Size	Number of Epochs
All	**0.1325**	**0.1215*** **	**0.1215*** **	**0.1216*** **	**0.1236*** **	**0.1220*** **	**0.1217*** **	**0.1234*** **
1	0.0027	0.0031	0.003	0.003	0.0034	0.0032	0.003	0.0034
2	0.0089	0.0095	0.0092	0.0095	0.0104	0.0098	0.0093	0.0104
3	0.0165	0.0168	0.0166	0.0168	0.0183	0.0172	0.0166	0.0182
4	0.0257	0.0257	0.0254***	0.0257	0.0276	0.0261	0.0255***	0.0274
5	0.0378	0.0369***	0.0366***	0.0369***	0.039	0.0374***	0.0368***	0.0388
6	0.0541	0.0519***	0.0516***	0.0519***	0.0540**	0.0524***	0.0518***	0.0538***
7	0.0789	0.0736***	0.0736***	0.0736***	0.0753***	0.0740***	0.0738***	0.0754***
8	0.1186	0.1093***	0.1092***	0.1094***	0.1103***	0.1097***	0.1095***	0.1105***
9	0.2014	0.1825***	0.1826***	0.1823***	0.1839***	0.1836***	0.1830***	0.1840***
10	0.7804	0.7059***	0.7069***	0.7064***	0.7139***	0.7068***	0.7083***	0.7122***

Table B.4. Table 4 Panel B after Optimizing Hyperparameters Other than the Number of Hidden Layers and Neurons

Panels A to G show the Table 4 Panel B results after estimating the underlying neural networks using the optimized hyperparameter values shown in Table B.2 Panel B one at a time.

Panel A. Optimizing Algorithm

Error Decile	NN Level 0	NN Level 4
All	**0.1215**	**0.1184*****
1	0.0343	0.0032***
2	0.0166	0.0099***
3	0.0220	0.0173***
4	0.0296	0.0261***
5	0.0392	0.0373***
6	0.0534	0.0518***
7	0.0753	0.0727***
8	0.1103	0.1064***
9	0.1800	0.1750***
10	0.6545	0.6846

Panel B. Learning Rate

Error Decile	NN Level 0	NN Level 4
All	**0.1215**	**0.1183*****
1	0.0326	0.0032***
2	0.0156	0.0099***
3	0.0216	0.0174***
4	0.0293	0.0263***
5	0.0391	0.0373***
6	0.0536	0.0519***
7	0.0756	0.0724***
8	0.1109	0.1060***
9	0.1803	0.1748***
10	0.6561	0.6840

Panel C. Activation Function

Error Decile	NN Level 0	NN Level 4
All	**0.1216**	**0.1183*****
1	0.0344	0.0032***
2	0.0163	0.0098***
3	0.0219	0.0173***
4	0.0296	0.0261***
5	0.0392	0.0371***
6	0.0535	0.0519***
7	0.0754	0.0725***
8	0.1105	0.1064***
9	0.1801	0.1753***
10	0.6545	0.6833

Panel D. Regularization Rate

Error Decile	NN Level 0	NN Level 4
All	**0.122**	**0.1181*****
1	0.037	0.0032***
2	0.0177	0.0098***
3	0.0226	0.0173***
4	0.0296	0.0261***
5	0.0392	0.0369***
6	0.0533	0.0517***
7	0.0754	0.0723***
8	0.1107	0.1059***
9	0.1808	0.1741***
10	0.6539	0.6836

Panel E. Batch Size

Error Decile	NN Level 0	NN Level 4
All	**0.1217**	**0.1186*****
1	0.0329	0.0032***
2	0.0157	0.0098***
3	0.0217	0.0173***
4	0.0295	0.0263***
5	0.0395	0.0373***
6	0.0539	0.0520***
7	0.0759	0.0729***
8	0.1113	0.1067***
9	0.1806	0.1752***
10	0.6566	0.6851

Panel F. Regularization Type

Error Decile	NN Level 0	NN Level 4
All	**0.1236**	**0.1211****
1	0.0461	0.0034***
2	0.024	0.0105***
3	0.0291	0.0183***
4	0.0355	0.0275***
5	0.0436	0.0387***
6	0.0563	0.0534***
7	0.0762	0.0742***
8	0.109	0.1078*
9	0.1737	0.1782
10	0.6427	0.6988

Panel G. Number of Epochs

Error Decile	NN Level 0	NN Level 4
All	**0.1234**	**0.1205****
1	0.0465	0.0034***
2	0.0232	0.0105***
3	0.0284	0.0183***
4	0.035	0.0274***
5	0.0434	0.0386***
6	0.0563	0.0534***
7	0.0764	0.0740***
8	0.1092	0.1074**
9	0.1743	0.1775
10	0.6413	0.6948

Figure 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.

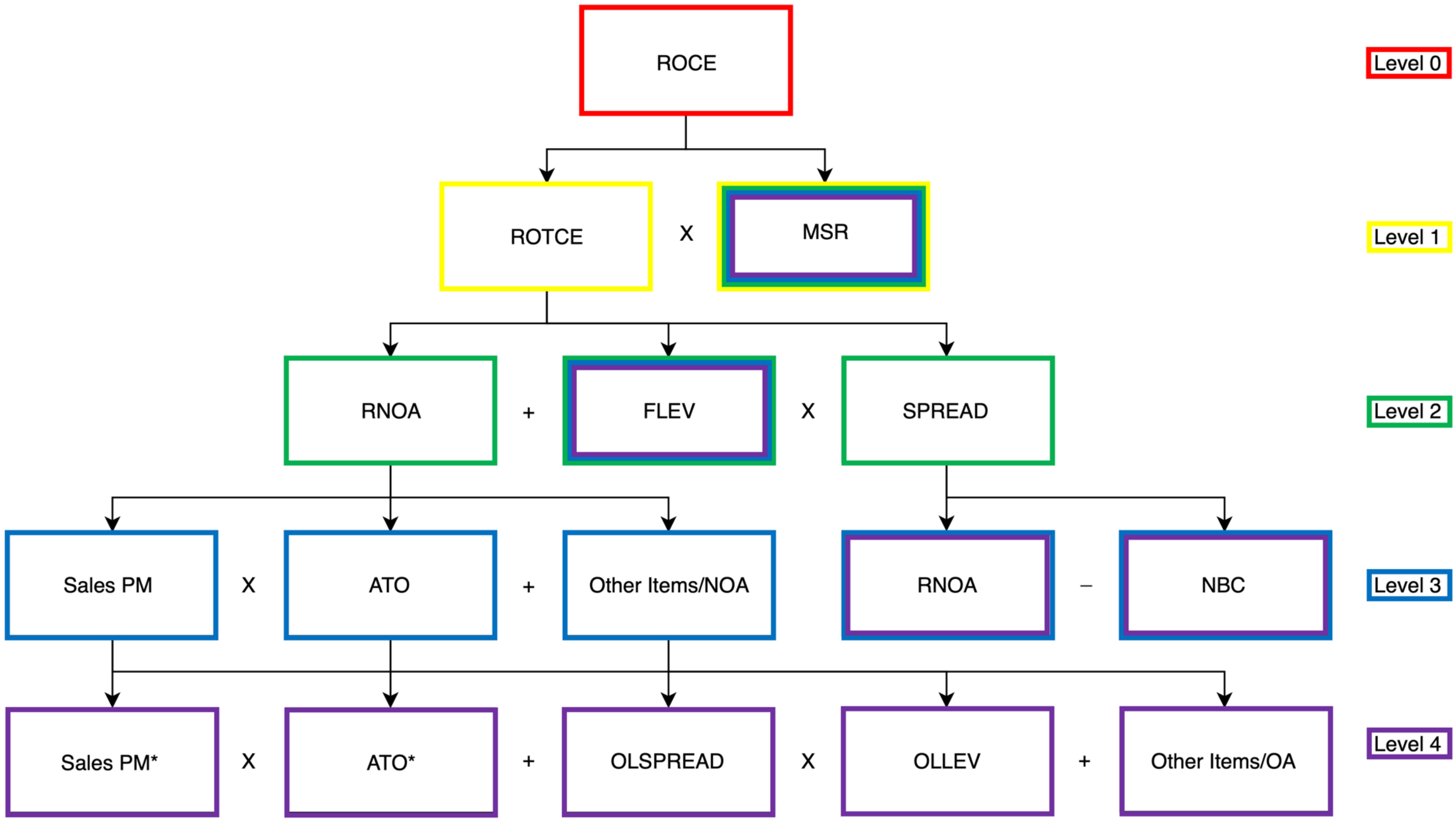


Figure 2. Univariate Time-Series Plots

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.

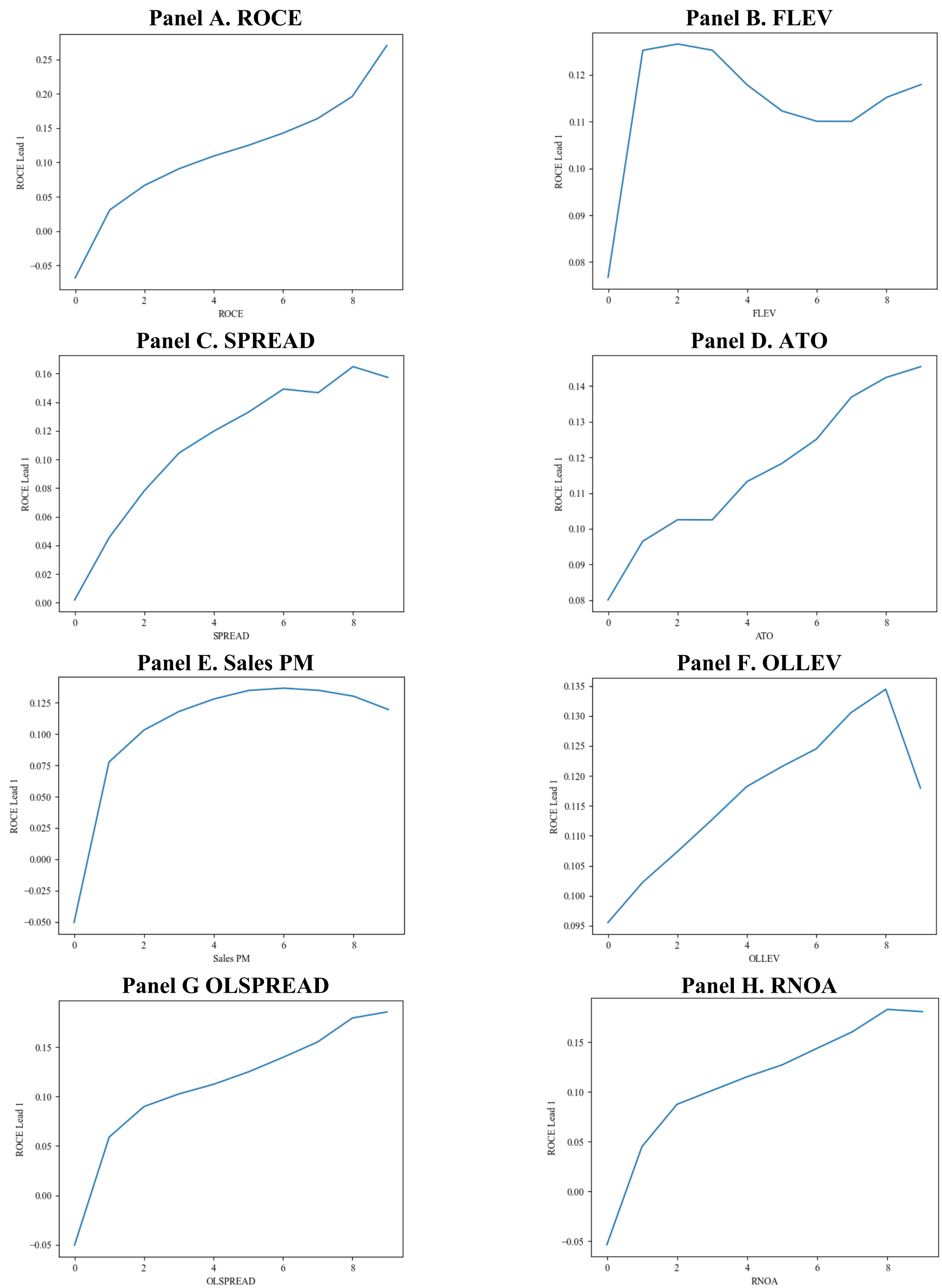
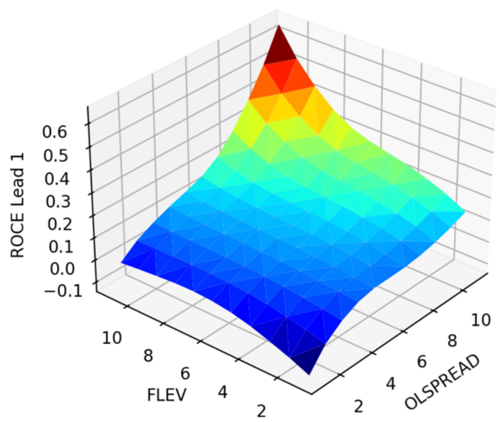


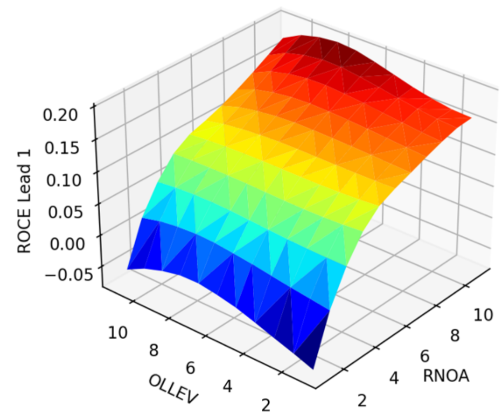
Figure 3. Interactive Relations across Variables in ROCE Prediction

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.

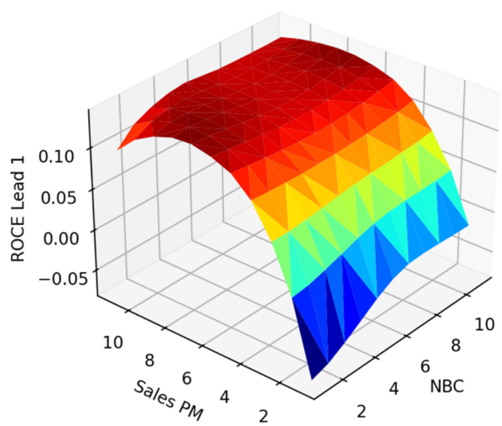
Panel A. FLEV & OLSPREAD



Panel B. OLLEV & RNOA



Panel C. Sales PM & NBC



Panel D. SPREAD & ATO

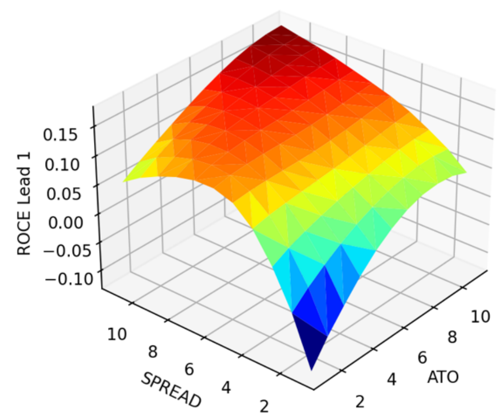


Figure 4. Average Absolute Forecast Error Comparison: Random Walk, OLS, and NN
This figure plots average absolute forecast errors for the full sample for the random walk (RW), OLS, and NN (neural network) models shown in Table 3.

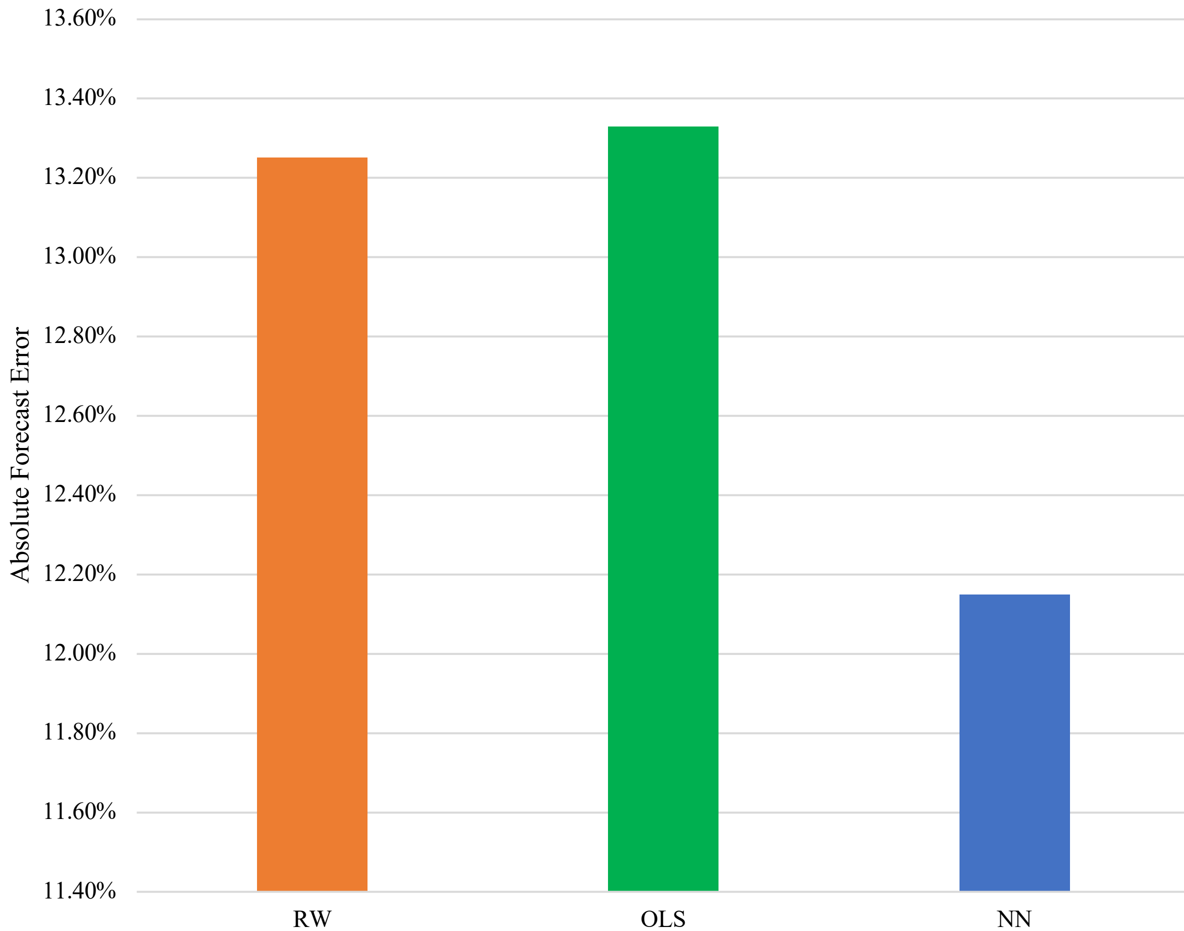


Figure 5. OLS and NN relative to Random Walk Absolute Forecast Errors across the Absolute Forecast Error Distribution

This figure plots average absolute forecast errors across the absolute forecast error deciles for OLS and Neural Network (NN) models scaled by random walk forecast errors shown in Table 3.

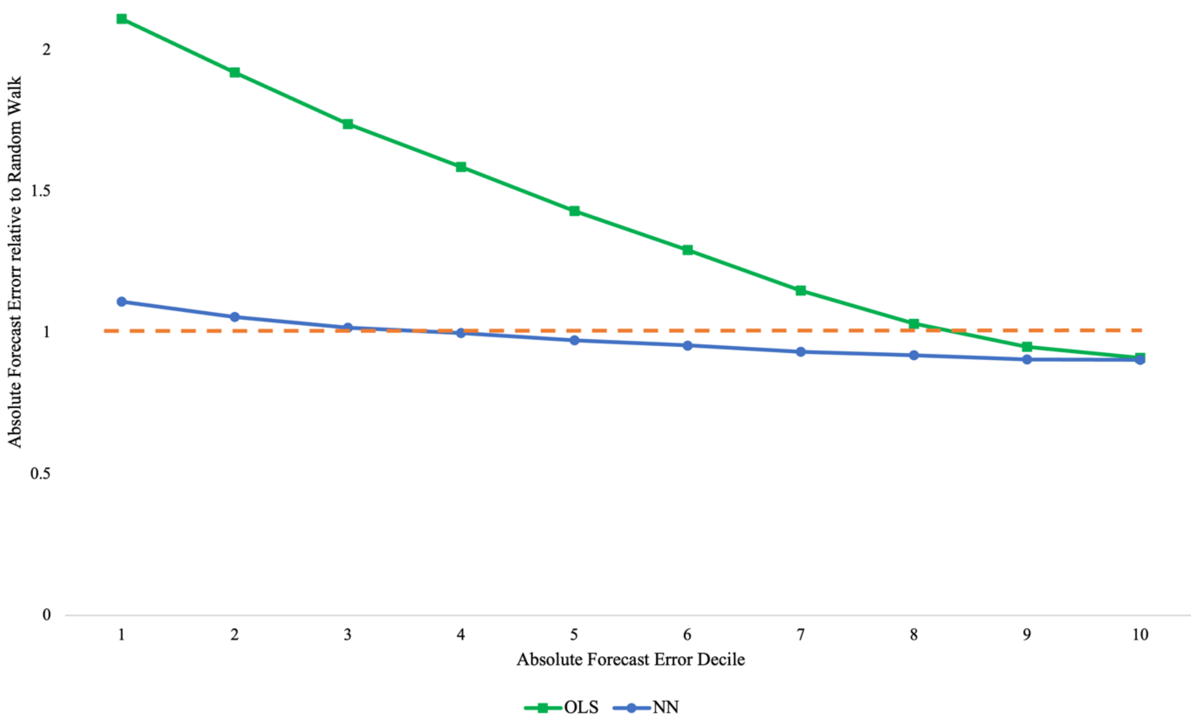


Figure 6. Average Absolute Forecast Errors across Decomposition Levels
This figure plots average absolute forecast errors for the full sample across Level 0 to Level 4 of NP's profitability decomposition framework shown in Figure 1 and NN models shown in Table 4 Panel B.

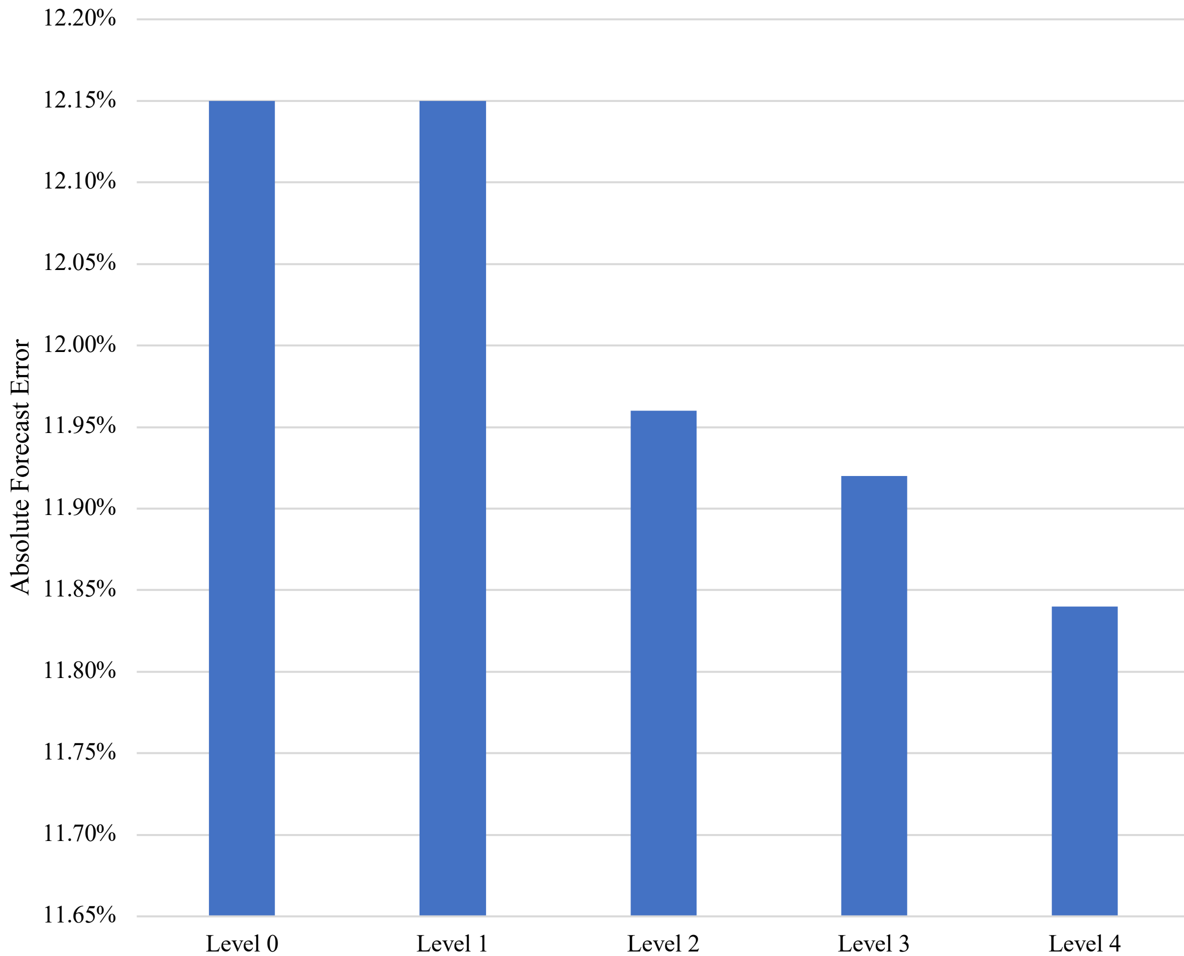
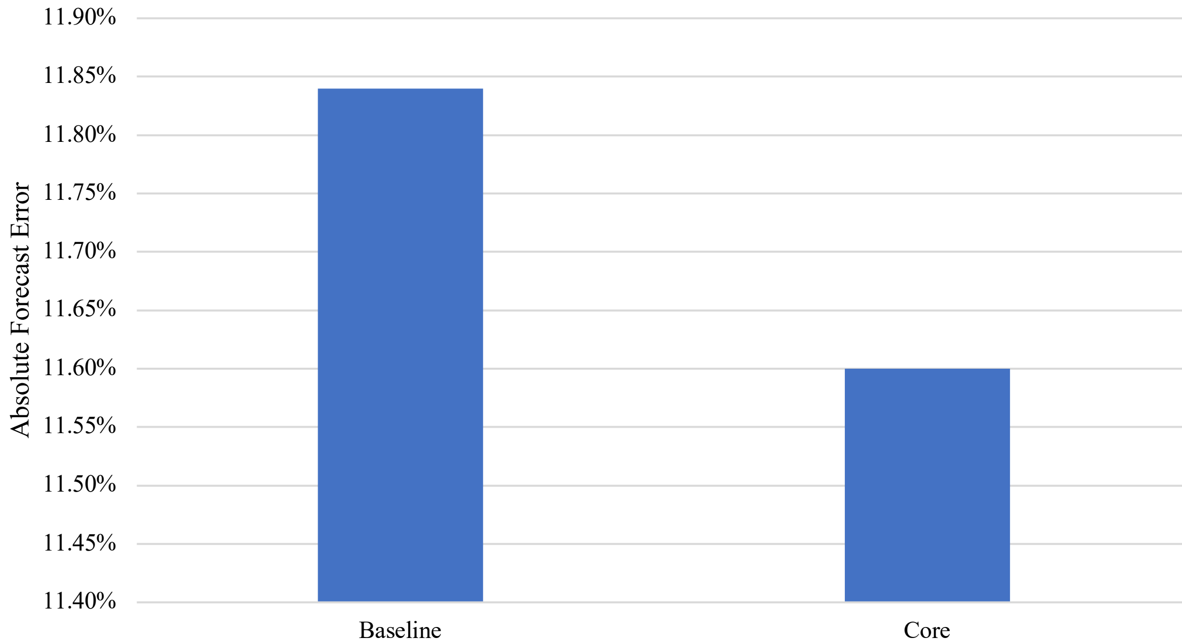


Figure 7. Average Absolute Forecast Error Comparison: Core Items (Panel A) and Historical Information (Panel B)

Panel A plots average absolute forecast errors for models including core and transitory items (Baseline) and models focusing on core items (Core Only) shown in Table 8. Panel B plots average 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 9.

Panel A. Core Items



Panel B. Historical Information

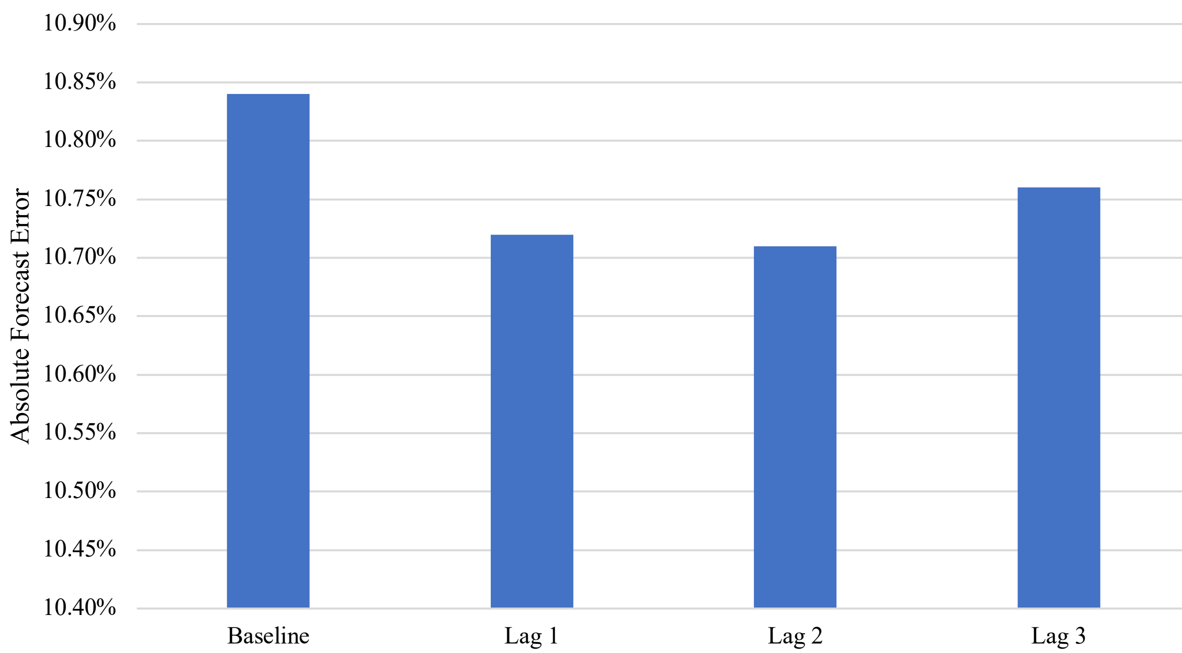
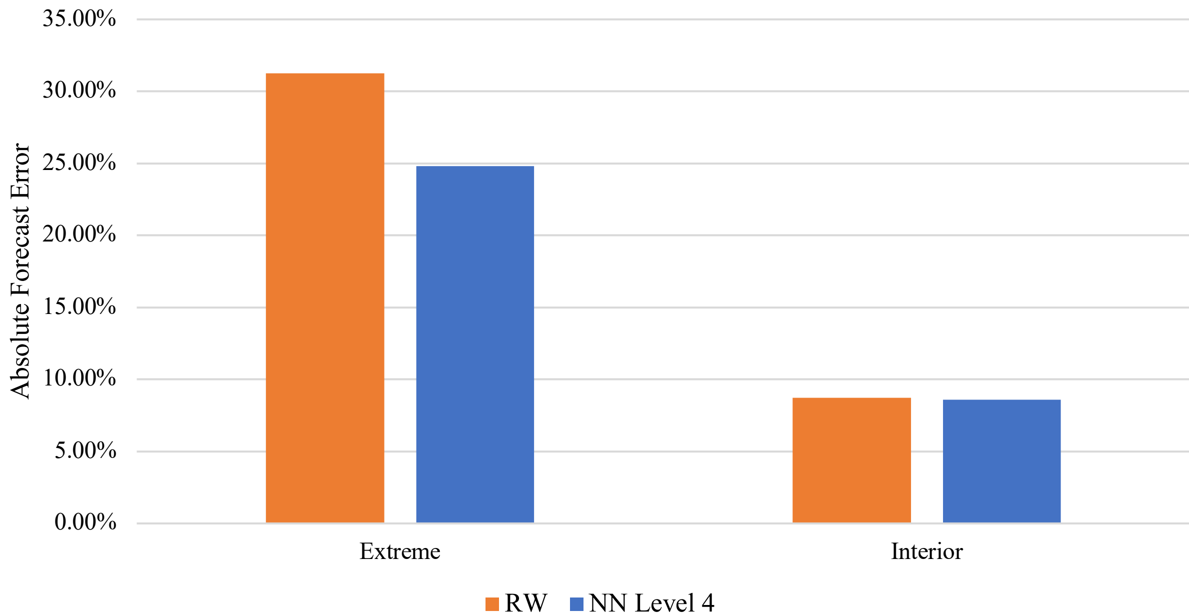


Figure 8. Average Absolute Forecast Error Comparison: Extreme vs Interior Values (Panel A) and Life Cycle (Panel B)

Panel A plots average absolute forecast errors for the random walk (RW) and Neural Network Level 4 (NN Level 4) models separately for observations with current profitability above the 90th or below the 10th percentile (Extreme) and for observations with current profitability within the 90th and 10th percentiles (Interior) shown in Table 10 Panel A. Panel B plots average absolute forecast errors for the random walk (RW) and Neural Network Level 4 (NN Level 4) models separately for observations within different lifecycle stages (Introduction, Growth, Maturity, Decline, Shakeout) shown in Table 10 Panel B.

Panel A. Extreme vs Interior Values



Panel B. Life Cycle

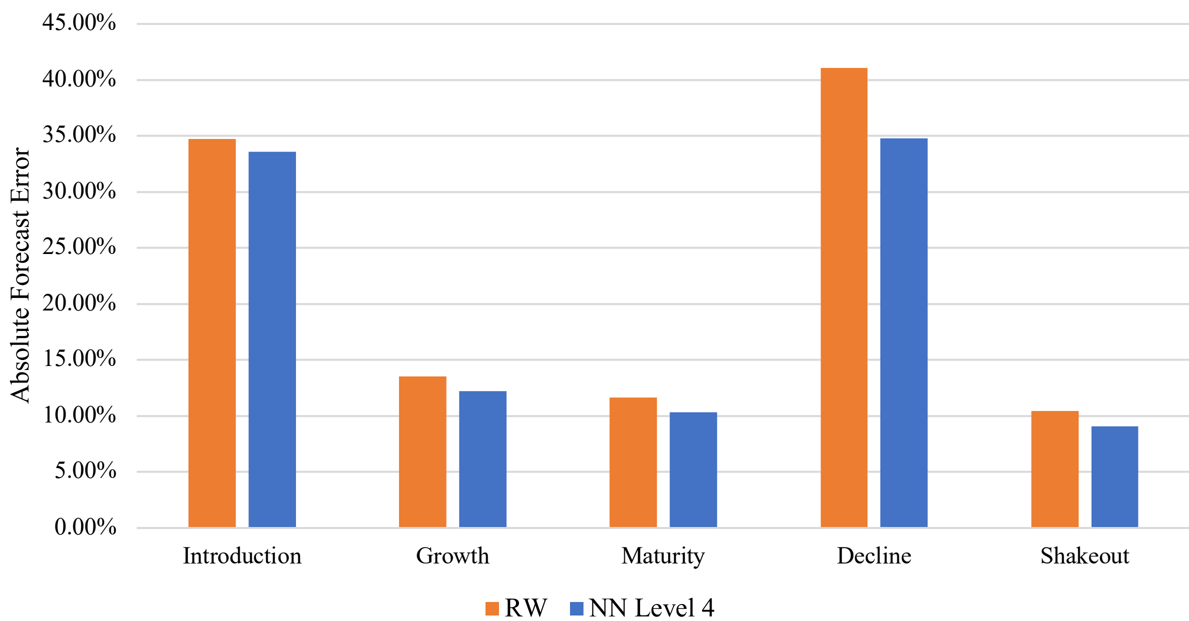


Table 1. Descriptive Statistics

This table presents descriptive statistics for the full sample. All variables are defined in Appendix A.

	N	Mean	SD	P1	P25	Median	P50	P99
ROCE	117,411	0.10	0.27	-0.83	0.06	0.12	0.18	0.75
ROTCE	117,411	0.10	0.27	-0.83	0.06	0.12	0.18	0.76
MSR	117,411	0.99	0.04	0.73	1.00	1.00	1.00	1.08
RNOA	117,411	0.13	0.51	-1.44	0.06	0.10	0.18	2.21
FLEV	117,411	0.65	1.71	-0.97	-0.12	0.31	0.89	8.51
SPREAD	117,411	0.09	0.67	-2.14	0.00	0.05	0.15	2.81
Sales PM	117,411	0.04	0.60	-1.40	0.04	0.07	0.13	0.63
ATO	117,411	2.23	2.75	0.11	0.79	1.57	2.56	15.87
Other Items/NOA	117,411	0.00	0.01	-0.02	0.00	0.00	0.00	0.07
NBC	117,411	0.04	0.14	-0.58	0.01	0.04	0.06	0.65
*Sales PM**	117,411	0.06	0.59	-1.38	0.04	0.08	0.15	0.70
*ATO**	117,411	1.23	0.93	0.06	0.55	1.06	1.62	5.00
OLSPREAD	117,411	0.05	0.20	-0.70	0.01	0.04	0.09	0.69
OLLEV	117,411	0.91	2.43	0.05	0.26	0.40	0.66	14.18
Other Items/OA	117,411	0.00	0.01	-0.01	0.00	0.00	0.00	0.04

Table 2. Correlations

This table presents Pearson (Spearman) correlations below (above) the diagonal for the full sample. * indicates significance at the 5% level. All variables are defined in Appendix A.

Variable		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
ROCE Lead 1	1	1	0.64*	0.63*	0.03*	0.50*	0.03*	0.41*	0.30*	0.22*	0.05*	0.01	0.28*	0.22*	0.48*	0.13*	0.05*
ROCE	2	0.42*	1	1.00*	0.05*	0.78*	0.00	0.64*	0.54*	0.25*	0.06*	-0.01*	0.51*	0.26*	0.74*	0.12*	0.06*
ROTCE	3	0.42*	1.00*	1	0.02*	0.78*	0.00	0.64*	0.54*	0.25*	0.06*	-0.01*	0.52*	0.26*	0.74*	0.12*	0.06*
MSR	4	0.00	0.01*	-0.02*	1	0.01	0.01*	-0.01*	-0.03*	0.03*	-0.06*	0.03*	-0.01*	0.05*	-0.04*	-0.04*	-0.05*
RNOA	5	0.23*	0.38*	0.38*	-0.01*	1	-0.43*	0.83*	0.57*	0.45*	0.06*	-0.18*	0.55*	0.40*	0.90*	0.29*	0.05*
FLEV	6	0.03*	-0.06*	-0.06*	-0.02*	-0.07*	1	-0.42*	-0.03*	-0.44*	0.07*	0.44*	-0.03*	-0.35*	-0.35*	-0.33*	0.07*
SPREAD	7	0.18*	0.29*	0.29*	-0.01*	0.85*	-0.07*	1	0.52*	0.36*	0.03*	-0.56*	0.50*	0.31*	0.78*	0.23*	0.03*
Sales PM	8	0.26*	0.34*	0.34*	-0.02*	0.54*	0.06*	0.44*	1	-0.31*	0.01*	-0.13*	0.98*	-0.33*	0.64*	-0.12*	0.01*
ATO	9	0.04*	0.02*	0.02*	0.02*	0.28*	-0.22*	0.26*	0.02*	1	-0.01*	-0.10*	-0.32*	0.92*	0.29*	0.56*	-0.02*
Other Items/NOA	10	0.04*	0.05*	0.05*	-0.05*	0.13*	-0.04*	0.11*	0.05*	0.03*	1	0.05*	0.02*	-0.02*	0.06*	0.03*	1.00*
NBC	11	0.02*	0.01*	0.01*	0.00	-0.04*	0.07*	-0.42*	0.01*	-0.06*	-0.03*	1	-0.12*	-0.04*	-0.20*	-0.13*	0.05*
*Sales PM**	12	0.26*	0.34*	0.34*	-0.02*	0.54*	0.06*	0.44*	1.00*	0.02*	0.05*	0.01*	1	-0.37*	0.57*	-0.06*	0.02*
*ATO**	13	0.10*	0.11*	0.11*	0.06*	0.21*	-0.22*	0.18*	0.05*	0.73*	-0.01*	-0.02*	0.03*	1	0.30*	0.30*	-0.02*
OLSPREAD	14	0.30*	0.48*	0.48*	-0.02*	0.88*	-0.07*	0.73*	0.65*	0.18*	0.14*	-0.04*	0.64*	0.21*	1	0.14*	0.06*
OLLEV	15	-0.01*	-0.05*	-0.05*	-0.01*	0.13*	-0.13*	0.14*	-0.01*	0.46*	0.02*	-0.07*	0.03*	0.00	0.00	1	0.02*
Other Items/OA	16	0.04*	0.05*	0.05*	-0.04*	0.10*	-0.04*	0.09*	0.05*	-0.01*	0.96*	-0.02*	0.05*	-0.03*	0.12*	-0.01*	1

Table 3. Absolute Forecast Errors from Random Walk, Linear (OLS) Estimation, and Nonlinear (NN) Estimation of Year-Ahead Profitability in an Autoregressive Model

This table shows mean absolute year-ahead ROCE forecast errors of models based on a random walk (RW), linear (OLS) estimation and nonlinear (NN) 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.

Error Decile	RW	OLS	NN
All	**0.1325**	**0.1333**	**0.1215*** **
1	0.0027	0.0057	0.0030
2	0.0089	0.0171	0.0094
3	0.0165	0.0287	0.0168
4	0.0257	0.0408	0.0257**
5	0.0378	0.0541	0.0368***
6	0.0541	0.0700	0.0517***
7	0.0789	0.0907	0.0736***
8	0.1186	0.1225	0.1092***
9	0.2014	0.1914***	0.1826***
10	0.7804	0.7114***	0.7066***

Table 4. Absolute One-Year Ahead Profitability Forecast Errors from Linear (OSL) and Nonlinear (NN) Estimation of NP's Framework

This table shows mean absolute year-ahead ROCE forecast errors of models based on linear (OLS) and nonlinear (NN) 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	**0.1333**	**0.1332**	**0.1518**	**0.1497**	**0.1456**
1	0.0057	0.0057	0.0071	0.0065	0.0066
2	0.0171	0.0171	0.0213	0.0198	0.0197
3	0.0287	0.0286*	0.0356	0.0336	0.0329
4	0.0408	0.0407***	0.0505	0.0481	0.0467
5	0.0541	0.0540**	0.0670	0.0643	0.0620
6	0.0700	0.0700	0.0871	0.0838	0.0807
7	0.0907	0.0906*	0.1136	0.1096	0.1057
8	0.1225	0.1225	0.1524	0.1481	0.1433
9	0.1914	0.1913	0.2277	0.2231	0.2179
10	0.7114	0.7113	0.7551	0.7603	0.7407

Panel B. Nonlinear Estimation (NN)

Error Decile	Level 0	Level 1	Level 2	Level 3	Level 4
All	**0.1215**	**0.1215**	**0.1196***	**0.1192****	**0.1184*****
1	0.0030	0.0030	0.0030	0.0031	0.0032
2	0.0094	0.0095	0.0095	0.0096	0.0098
3	0.0168	0.0168	0.0168	0.0170	0.0173
4	0.0257	0.0257	0.0257	0.0259	0.0262
5	0.0368	0.0369	0.0368*	0.0369	0.0373
6	0.0517	0.0519	0.0516**	0.0517	0.0520
7	0.0736	0.0737	0.0729***	0.0729***	0.0727***
8	0.1092	0.1094	0.1075***	0.1069***	0.1065***
9	0.1826	0.1821	0.1780***	0.1770***	0.1753***
10	0.7066	0.7055	0.6941	0.6906*	0.6842**

Table 5. Robustness Test: Truncation

This table shows mean absolute year-ahead ROCE forecast errors of models based on a random walk (RW), on linear (OLS) estimation, and on nonlinear (NN) estimation of Level 0 to Level 4 of NP's profitability decomposition framework after truncation at the 1st and 99th percentiles. Panel A compares linear (OLS) and nonlinear (NN) estimation to a random walk. Panel B compares nonlinear estimation of Level 0 with nonlinear estimation of Level 1 through Level 4 of NP's profitability decomposition framework. 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 (Panel A) and than that of the Level 0 model in the second column of the same row (Panel B).

Panel A. Nonlinear Estimation

Error Decile	RW	OLS	NN
All	**0.1039**	**0.1029**	**0.0960******
1	0.0028	0.0048	0.0028
2	0.0087	0.0143	0.0088
3	0.0159	0.0240	0.0157***
4	0.0245	0.0339	0.0238***
5	0.0357	0.0450	0.0340***
6	0.0505	0.0585	0.0474***
7	0.0725	0.0768	0.0666***
8	0.1071	0.1045***	0.0968***
9	0.1749	0.1610***	0.1556***
10	0.5461	0.5065***	0.5090***

Panel B. NP Framework

Error Decile	Level 0	Level 1	Level 2	Level 3	Level 4
All	**0.0960**	**0.0962**	**0.0954**	**0.0944****	**0.0944****
1	0.0028	0.0029	0.0030	0.0030	0.0031
2	0.0088	0.0092	0.0094	0.0093	0.0095
3	0.0157	0.0162	0.0163	0.0162	0.0163
4	0.0238	0.0244	0.0245	0.0243	0.0245
5	0.0340	0.0345	0.0346	0.0343	0.0344
6	0.0474	0.0478	0.0479	0.0474	0.0475
7	0.0666	0.0669	0.0667	0.0663***	0.0661***
8	0.0968	0.0971	0.0964***	0.0950***	0.0947***
9	0.1556	0.1557	0.1538***	0.1515***	0.1512***
10	0.5090	0.5076	0.5015	0.4966**	0.4971**

Table 6. 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 gradient-boosted tree (Gradient-Boosted Tree) 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 than that of the corresponding Level 0 model (Panel B) in the adjacent column of the same row.

Panel A. Nonlinear Estimation

Error Decile	RW	Random Forest	Gradient-Boosted Tree	Holdout
All	**0.1325**	**0.1313**	**0.1307***	**0.1225*****
1	0.0027	0.0049	0.0051	0.0031
2	0.0089	0.0145	0.0150	0.0097
3	0.0165	0.0242	0.0251	0.0173
4	0.0257	0.0350	0.0359	0.0263
5	0.0378	0.0476	0.0482	0.0377
6	0.0541	0.0639	0.0640	0.0530***
7	0.0789	0.0867	0.0862	0.0751***
8	0.1186	0.1239	0.1226	0.1108***
9	0.2014	0.2025	0.1997***	0.1842***
10	0.7804	0.7096***	0.7056***	0.7079***

Panel B. NP Framework

	Random Forest		Gradient-Boosted Tree		Holdout	
Error Decile	NN Level 0	NN Level 4	NN Level 0	NN Level 4	NN Level 0	NN Level 4
All	**0.1313**	**0.1280*****	**0.1307**	**0.1288***	**0.1225**	**0.1193*****
1	0.0049	0.0045***	0.0051	0.0045***	0.0031	0.0032
2	0.0145	0.0137***	0.0150	0.0138***	0.0097	0.0100
3	0.0242	0.0232***	0.0251	0.0235***	0.0173	0.0177
4	0.0350	0.0338***	0.0359	0.0340***	0.0263	0.0265
5	0.0476	0.0463***	0.0482	0.0465***	0.0377	0.0374***
6	0.0639	0.0624***	0.0640	0.0627***	0.0530	0.0521***
7	0.0867	0.0856***	0.0862	0.0854***	0.0751	0.0735***
8	0.1239	0.1225***	0.1226	0.1222**	0.1108	0.1074***
9	0.2025	0.1983***	0.1997	0.1982***	0.1842	0.1764***
10	0.7096	0.6892**	0.7056	0.6976	0.7079	0.6889**

Table 7. Robustness Test: Additional Predictors

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); indicators for each Fama-French 48-industry (Industry); annual real GDP growth, inflation, and unemployment (Macro); and adding the Dupont, Hou, Industry, and Macro variables jointly (Joint). Panel A displays results using the Hou, Macro, Industry, and Joint variables only as predictors. Panel B displays results using the Hou, Macro, Industry, and Joint variables in conjunction with the Baseline model variables as predictors. 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.

Panel A. Without NP Variables

Error Decile	Baseline	Dupont	Hou	Industry	Macro	Joint
All	**0.1298**	**0.1364**	**0.1406**	**0.1603**	**0.1623**	**0.1393**
1	0.0036	0.0041	0.0046	0.0054	0.0058	0.0045
2	0.0112	0.0125	0.0141	0.0168	0.0176	0.0137
3	0.0196	0.0217	0.0241	0.0293	0.0305	0.0238
4	0.0293	0.0325	0.0355	0.0433	0.0449	0.0349
5	0.0417	0.0459	0.0492	0.0597	0.0616	0.0486
6	0.0581	0.0633	0.0664	0.0800	0.0823	0.0655
7	0.0811	0.0876	0.0901	0.1079	0.1100	0.0893
8	0.1186	0.1264	0.1284	0.1523	0.1543	0.1271
9	0.1937	0.2054	0.2094	0.2457	0.2486	0.2073
10	0.7415	0.7645	0.7842	0.8626	0.8672	0.7781

Panel B. With NP Variables

Error Decile	Baseline	Dupont	Hou	Industry	Macro	Joint
All	**0.1298**	**0.1301**	**0.1414**	**0.1302**	**0.1803**	**0.1517**
1	0.0036	0.0037	0.0046	0.0038	0.0044	0.0047
2	0.0112	0.0112	0.0142	0.0114	0.0135	0.0146
3	0.0196	0.0196	0.0244	0.0198	0.0234	0.0251
4	0.0293	0.0295	0.0358	0.0298	0.0347	0.0373
5	0.0417	0.0415**	0.0495	0.0422	0.0486	0.0521
6	0.0581	0.0581	0.0666	0.0584	0.0673	0.0707
7	0.0811	0.0814	0.0906	0.0813	0.0936	0.0967
8	0.1186	0.1190	0.1293	0.1188	0.1372	0.1394
9	0.1937	0.1940	0.2111	0.1949	0.2345	0.2296
10	0.7415	0.7429	0.7877	0.7414	1.1456	0.8464

Table 8. Additional Test: Core Items

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.

Error Decile	Baseline	Core
All	**0.1184**	**0.1160****
1	0.0032	0.0033
2	0.0098	0.0100
3	0.0173	0.0174
4	0.0262	0.0260***
5	0.0373	0.0365***
6	0.0520	0.0503***
7	0.0727	0.0700***
8	0.1065	0.1020***
9	0.1753	0.1691***
10	0.6842	0.6752

Table 9. Additional Test: Historical Information

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.

Error Decile	Baseline	Lag 1	Lag 2	Lag 3
All	**0.1084**	**0.1072**	**0.1071**	**0.1076**
1	0.0033	0.0032***	0.0032***	0.0032***
2	0.0100	0.0097***	0.0099***	0.0100
3	0.0174	0.0170***	0.0173***	0.0173***
4	0.0261	0.0256***	0.0258***	0.0259***
5	0.0367	0.0362***	0.0364***	0.0366
6	0.0506	0.0500***	0.0504**	0.0507
7	0.0704	0.0693***	0.0696***	0.0700***
8	0.1012	0.0995***	0.0999***	0.1003***
9	0.1630	0.1602***	0.1598***	0.1605***
10	0.6057	0.6010	0.5985	0.6017

Table 10. Additional Test: Cross-Sectional Variation in Profitability (Panel A) and Lifecycle Stage (Panel B)

Panel A compares mean absolute year-ahead ROCE forecast errors of models based on nonlinear (NN) estimation of Level 4 of NP's profitability decomposition framework to those of a random walk separately for observations within the 10th and 90th ROCE percentile (Interior) and for observations below the 10th or above the 90th ROCE percentile (Extreme). Panel B displays similar forecast error comparisons for observations in different Dickinson (2011) life cycle stages (Introduction, Growth, Maturity, Decline, Shakeout). 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 RW (random walk) model in the adjacent column of the same row.

Panel A. Extreme vs Interior Profitability

	Extreme		Interior	
Error Decile	RW	NN Level 4	RW	NN Level 4
All	**0.3127**	**0.2480*** **	**0.0874**	**0.0860***
1	0.0058	0.0084	0.0025	0.0027
2	0.0254	0.0247***	0.0078	0.0084
3	0.0494	0.0435***	0.0140	0.0148
4	0.0809	0.0652***	0.0214	0.0221
5	0.1181	0.0924***	0.0307	0.0310
6	0.1667	0.1268***	0.0428	0.0424***
7	0.2325	0.1727***	0.0598	0.0581***
8	0.3312	0.2462***	0.0863	0.0817***
9	0.5305	0.3862***	0.1349	0.1278***
10	1.5861	1.3142***	0.4735	0.4707

Panel B. Life Cycle

	Introduction		Growth		Maturity		Decline		Shakeout	
Error Decile	RW	NN Level 4	RW	NN Level 4	RW	NN Level 4	RW	NN Level 4	RW	NN Level 4
All	**0.3472**	**0.3361**	**0.1356**	**0.1221*****	**0.1164**	**0.1032*****	**0.4107**	**0.3476*****	**0.1043**	**0.0906*****
1	0.0070	0.0074	0.0032	0.0036	0.0026	0.0035	0.0089	0.0088	0.0022	0.0024
2	0.0224	0.0220	0.0101	0.0110	0.0093	0.0109	0.0286	0.0290	0.0070	0.0074
3	0.0400	0.0381***	0.0183	0.0195	0.0174	0.0189	0.0547	0.0520***	0.0128	0.0131
4	0.0631	0.0597***	0.0282	0.0290	0.0270	0.0282	0.0896	0.0786***	0.0199	0.0197***
5	0.0963	0.0896***	0.0409	0.0406***	0.0391	0.0398	0.1326	0.1157***	0.0292	0.0281***
6	0.1433	0.1300***	0.0587	0.0561***	0.0551	0.0546***	0.1901	0.1587***	0.0420	0.0394***
7	0.2179	0.1970***	0.0850	0.0772***	0.0788	0.0749***	0.2804	0.2318***	0.0605	0.0552***
8	0.3333	0.3119***	0.1260	0.1133***	0.1158	0.1059***	0.4415	0.3391***	0.0912	0.0804***
9	0.6038	0.5630***	0.2075	0.1820***	0.1853	0.1645***	0.8073	0.5872***	0.1570	0.1338***
10	1.9435	1.9408	0.7782	0.6890***	0.6340	0.5312***	2.0736	1.8754**	0.6217	0.5260***

Table 11. Informativeness of Profitability Forecasts from Nonlinear Estimation of NP's Profitability Framework for Investors (Panel A) and Analysts (Panel B)

Panel A [Panel B] regresses stock returns (*Returns*) [changes in profitability (Δ*ROCE*)] over the subsequent year on the change in profitability predicted by NP's Level 4 model estimated via neural networks (*NN Level 4*), the change in profitability predicted by the mean (*Mean Analyst Forecast*) or median (*Median Analyst Forecast*) analyst consensus, 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.

Panel A. Stock Returns

	(1)	(2)	(3)	(4)	(5)	(6)
Variable			*Returns*			
NN Level 4	0.071***	0.068***	0.074***	0.070***	0.074***	0.071***
	(3.398)	(3.018)	(3.425)	(3.044)	(3.445)	(3.062)
Median Analyst Forecast			-0.013	-0.013		
			(-0.997)	(-0.963)		
Mean Analyst Forecast					-0.014	-0.014
					(-1.266)	(-1.222)
Beta		0.006		0.006		0.006
		(0.239)		(0.244)		(0.245)
SMB		0.028		0.028		0.028
		(1.083)		(1.091)		(1.093)
HML		0.034*		0.033*		0.033*
		(2.022)		(2.018)		(2.014)
RMW		-0.046		-0.047		-0.047
		(-1.452)		(-1.452)		(-1.453)
CMA		0.024		0.024		0.023
		(0.781)		(0.781)		(0.780)
UMD		-0.016		-0.016		-0.016
		(-0.528)		(-0.518)		(-0.517)
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,654	19,654	19,654	19,654	19,654	19,654
R-squared	0.475	0.477	0.475	0.477	0.475	0.477

Panel B. Changes in Profitability

Variable	(1)	(2)	(3)	(4)	(5)	(6)
			$\Delta ROCE$			
NN Level 4	0.596***	0.596***	0.589***	0.589***	0.589***	0.589***
	(15.478)	(15.510)	(15.447)	(15.479)	(15.346)	(15.378)
Median Analyst Forecast			0.037**	0.036**		
			(2.190)	(2.158)		
Mean Analyst Forecast					0.034**	0.034*
					(2.102)	(2.066)
Beta		0.018		0.018		0.018
		(0.886)		(0.869)		(0.869)
SMB		0.016		0.016		0.016
		(1.026)		(0.986)		(0.983)
HML		-0.020		-0.019		-0.019
		(-1.634)		(-1.587)		(-1.579)
RMW		0.003		0.003		0.003
		(0.203)		(0.222)		(0.224)
CMA		-0.008		-0.008		-0.008
		(-0.609)		(-0.605)		(-0.604)
UMD		0.003		0.002		0.002
		(0.229)		(0.159)		(0.163)
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,654	19,654	19,654	19,654	19,654	19,654
R-squared	0.429	0.429	0.430	0.430	0.430	0.430