

What Can Analysts Learn from Artificial Intelligence about Fundamental Analysis?*

Oliver Binz
INSEAD
oliver.binz@insead.edu

Katherine Schipper
Duke University
katherine.schipper@duke.edu

Kevin Standridge
Duke University
kevin.standridge@duke.edu

November 2020

* We thank John Graham, Matt Kubic, Charles Lee, Robert Hills, Mory Elsaify, Bill Mayew, Suresh Nallareddy, Stephen Penman, James Pinnington, Xu Jiang, Qi Chen, Scott Dyreng, Wayne Landsman, Ayung Tseng, Dan Bens, Peter Joos, Thomas Keusch, Gans Narayanamoorthy, Stephanie Cheng, and workshop participants at Duke University, INSEAD, Tulane University, the 2018 AAA/Deloitte/J. Michael Cook Doctoral Consortium, and the 2019 EAA Annual Meeting for helpful comments and suggestions.

What Can Analysts Learn from Artificial Intelligence about Fundamental Analysis?

Abstract

Taking the perspective of an equity investor seeking to maximize risk-adjusted returns through financial statement analysis, we apply a machine learning algorithm to estimate Nissim and Penman's (2001) structural decomposition framework of profitability. Our approach explicitly takes account of the nonlinearities that precluded Nissim and Penman from estimating their framework. We first forecast profitability and then estimate intrinsic values using different subsets of Nissim and Penman's framework and different fundamental analysis design choices; we find that trading on these estimates generates substantial risk-adjusted returns. Choices that improve performance include increasingly granular ratio disaggregation and long-horizon forecasts of operating performance. Perhaps surprisingly, we find only weak evidence of benefits from a fundamental analysis that incorporates historical financial statement information beyond the current-period information or focuses only on core items. While taking account of non-linearities improves model performance for all firms, the effect is strongest for small, loss-making, technology, and financially distressed firms.

JEL Classification: C53, G10, M41

Keywords: Financial Statement Analysis, Machine Learning, Earnings Forecasting

1. Introduction

We combine the capabilities of machine learning with Nissim and Penman's (hereafter NP, 2001) hierarchical approach to financial statement analysis to estimate their structural model of accounting profitability. Applying several financial statement analysis design choices identified by their framework, and consistent with the approach described by, among others, Yohn (2020), we first forecast profitability (accounting earnings) and then calculate intrinsic values using those forecasts. We examine returns to trading strategies based on these value estimates to provide evidence on the decision usefulness of both NP's framework generally and the effects of varying several financial statement analysis design choices within the framework. The strength of NP's framework, an extension of the traditional DuPont decomposition, derives from their use of fundamental financial statement analysis to identify value-relevant information within financial statements while filtering out value-irrelevant noise; this filtering is essential if we accept the premise that trading on information is profitable, while trading on noise is not (Black 1986).

As NP demonstrate, the theoretical and empirical relations among the accounting ratios in their structural profitability model are non-linear, necessitating the use of non-traditional estimation approaches.¹ We apply Deep Learning (hereafter DL), a machine learning algorithm that readily accommodates nonlinearities, to evaluate how the financial statement analysis design choices identified in NP are empirically linked to both accounting profitability and trading profitability. Our findings are informative in three ways. First, they assess the NP framework's usefulness in financial statement analysis. Second, they guide analysts in designing their financial statement analyses. Third,

¹ Describing their results, NP (p. 128) report that their analysis "produced large t-statistics and reasonable R-square values in estimation, but the models performed poorly in prediction." They confirm both the instability of estimates produced from linear estimators applied to pooled data in their setting, and a nonlinear relation between current-period and future-period ratios. They conclude that "pooled, linear models are not likely to work well" and (on p. 148) suggest that their largely descriptive results "may lead to more formal, parsimonious forecast modeling which brings more sophisticated econometrics to the task."

they provide a practical demonstration of the value of artificial intelligence in financial statement analysis, specifically by resolving the difficulties NP encountered in estimating their model.

As discussed in more detail in section 2 and illustrated in Figure 1, NP's model and related discussion substantially extend and enrich the traditional DuPont decomposition. One practical application of their framework is to use past ratio realizations to forecast future ratios that can then be used to estimate intrinsic equity value. The relations among current-period and future-period ratios identified by NP are nonlinear² and cannot be modeled using traditional estimators such as Ordinary Least Squares (OLS) or Least Average Deviation (LAD), leading NP to conclude that these estimators perform poorly out of sample.³ Thus, NP confine their analysis to (mostly) descriptive documentation of empirical regularities, leaving to future research the task of applying their model to forecast profitability. Our analysis undertakes that forecasting task and the task of using the profit forecasts in valuation; in doing so, it shows the practical value of artificial intelligence in the analysis of intrinsic values.

We investigate the effects on model performance of five financial statement analysis design choices within NP's framework: level of disaggregation; including only core items (i.e., discarding unusual (impersistent) items); separately analyzing operating and financing activities; separately analyzing short and long forecast horizons; and using one vs. multiple lags of historical information.⁴

² Examples of non-linearities in valuation include the following: The interaction of turnovers and margins determines profitability; NP's Figure 2 documents a convex, downward sloping relation between margins and turnovers. Further, most value-relevant ratios and growth rates experience convex mean reversion over time (see NP Figures 4 to 6). Leverage has a positive relation with profitability until the costs exceed the benefits of leverage and the relation turns negative (e.g., Myers 1984; Graham 2000). Finally, analysts often assume changing growth rates to avoid logical errors such as the paradox that a firm assumed to grow at a faster rate than the economy will at some point effectively constitute the overall economy. More sophisticated models assume that growth rates decay geometrically over time (Penman 2012).

³ It is conceptually possible to model non-linearities with linear estimators by including higher-order polynomial and interactive terms. As a practical matter, however, the number of predictors grows exponentially rather than linearly under this approach, rendering the estimation infeasible. Section 2.2 elaborates on this point.

⁴ NP define operating activities as all activities except financing activities, defined as activities undertaken to obtain capital from investors. Examples of financial reporting items related to financing activities include interest expense; gains, losses and interest or dividend income on trading investments; and financing fees. NP define core activities as activities judged as persistent rather than unusual or transitory. Examples of transitory items include currency translation adjustments,

A decision maker applying fundamental financial statement analysis makes these choices on empirical grounds; theory is silent. Combinations of these design choices within NP's structural framework result in 192 models of theoretically grounded candidate ratios for fundamental analysis and security valuation. Our analysis provides evidence on the empirical questions of which combination of design choices performs best, and which design choices matter most.⁵

We analyze the level of disaggregation as a tradeoff between information loss and noise. Less disaggregation increases the risk of information loss. If a ratio's components are not perfectly correlated, each component might exhibit idiosyncratic value-relevant variation that is lost in aggregation. In contrast, greater disaggregation increases the risk of incorporating value-irrelevant noise. If idiosyncratic variation of ratio components cannot be used to trade profitably, either because the information is value-irrelevant or because it is already incorporated in price, there is no gain from disaggregation. Indeed, as discussed in Monahan (2018), research documents that a simple random walk outperforms linear models with a larger (more disaggregated) set of predictors in out-of-sample earnings forecasting.⁶

Including only core items can also be considered as a separate and distinct tradeoff between information loss and noise. Independent of disaggregation as NP apply the concept, managers and analysts provide firm-specific performance measures (pro forma earnings) that exclude certain items

income effects of discontinued operations, gains or losses on the sale of capital investments, write-offs, and restructuring charges. NP's approach is not the only possible one, however. For example, McVay (2006) defines core earnings as operating income before depreciation and special items, defined as "material events that arise from a firm's ongoing, continuing activities, but that are either unusual in nature or infrequent in occurrence – but not both – and must be disclosed as a separate line item as part of income from continuing operations, or in footnotes to the financial statements" (Revsine and Collins 2005, p. 55). Conceptually, we define core items as those arising from a firm's recurring principal business activities, items that are either transitory or largely unrelated to the firm's business model. Empirically, we use NP's variable definitions.

⁵ These investigations are consistent with Yohn's (2020) recommendation that researchers should analyze approaches suggested in textbooks, particularly when those approaches have not been subjected to empirical validation. In the same spirit, NP (p. 110) note that the DuPont decomposition is a "standard textbook" framework for financial statement analysis but "rarely appear[s] in research."

⁶ This idea is consistent with the principle of Ockham's razor. William of Ockham, a 14th century logician, argued that greater model complexity increases the possibility for error. In our context, this principle suggests that simpler models might outperform more complex ones.

designated as non-core; research finds that these measures have predictive ability but may be susceptible to managerial opportunism (Doyle, Lundholm, and Soliman 2003; Bradshaw, Christensen, Gee, and Whipple 2018; Bentley, Christensen, Gee, and Whipple 2018).⁷ It is unclear which approach is preferable for filtering out noise (transitory items): focusing on management/analyst performance measures that presumably include mostly or entirely firm-specific core items or applying consistent rules to all firms' financial statements.⁸ Furthermore, while models including persistent operating items and excluding 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 that managers sometimes manipulate income statement presentation, thereby blurring the core vs. non-core distinction.

A third design choice within NP's framework is the treatment of financing activities, determined by the analyst's assessment of whether a firm's financial structure has valuation implications. Some (Penman 2012; Li, Richardson, and Tuna 2014) view the valuation implications of operating and financing activities as mutually independent, with financial items valued at market (fair) value or book value. In contrast, the agency cost literature suggests that capital structure can affect managers' incentives, inducing an association between operating and financing activities. For example, managers of firms close to bankruptcy might take excessive risks, because of the call option-like payoff structure of equity (Jensen and Meckling 1976). It is thus an empirical question whether past financing activities are informative about future operating activities; we believe our approach is among the first to test directly how the treatment of accounting information about financing activities affects forecasting and trading profitability.

⁷ According to Audit Analytics, 97% of the S&P 500 firms used at least one pro forma metric in their regulatory findings in 2017, up from 59% in 1996 (McKeon 2018).

⁸ Lipe (1986) shows that gross profit; sales, general, and administrative expense; depreciation expense; interest expense; income tax expense; and other Compustat line items have incremental (and differing) predictive ability for future returns. He does not classify line items as core vs. transitory.

Finally, while textbooks recommend using as much past information and forecasting as far in the future as practicable, empirical evidence on both prescriptions is sparse. Indirectly addressing the question of forecast horizon, Nissim (2019) asks which terminal value functional form and forecast horizon/terminal value combination is closest to price in the terminal value year; he reports that a 10-year horizon works best. To the best of our knowledge, no study *directly* examines the choices of forecast horizon and the number of years of past data to include.

Our approach to evaluating the 192 models based on variation in design choices to implement NP's framework adopts the perspective of an equity investor who makes investment decisions using fundamental financial statement analysis and strives to maximize risk-adjusted returns. Black (1986) argues that noise trading causes short-run divergence of stock prices from value while information-based trading causes long-run convergence of stock prices to value. The more price diverges from value, the larger the reward for information-based trading. The implication is that trades based on models that produce a less noisy estimate of value and help to detect divergence of value from price are more profitable.⁹ To implement this idea, we forecast out-of-sample profitability, estimate intrinsic values, calculate value-to-price (VP) ratios, and form hedge portfolios, buying (selling) firms in the highest (lowest) VP decile. We obtain model-specific alpha estimates by regressing value-weighted excess hedge returns on excess market returns, high-minus-low, small-minus-big, and up-minus-down factors (Carhart 1997). Using OLS and quantile regressions, we regress alphas on indicators for model-specific financial statement analysis design choices to evaluate which choices improve model performance. We use alphas to make models of different scales comparable, an important consideration in evaluating the performance of RNOA-based and ROCE-based models.

In preliminary analyses, we confirm many of NP's findings, based on their sample period 1963-

⁹ The absence of an agreed-upon asset pricing model makes identification of mispricing challenging. However, as noted by, for example, Jackwerth and Slavutskaya (2018), asset pricing models can be used to assess relative performance across models incorporating different fundamentals.

1999 and our longer sample period, 1963-2017; both samples include NYSE/AMEX firms, not NASDAQ firms. In extensions of NP's analyses that plot future profitability against combinations of current-period ratios, we show that current-period ratios are individually and interactively associated with current and future profitability in non-linear ways, including S-shaped, U-shaped, and concave patterns. These relations are visible at least 10 years into the future and attenuate over time. We next use DL to estimate these non-linear relations within NP's framework. Following Monahan (2018), we benchmark the resulting one-year-ahead out-of-sample profitability forecasts against a random walk and generally find that the DL-based predictions are more accurate.

In our main analyses, we compute an alpha estimate for each of 192 models derived from combinations of the five financial statement analysis design choices within NP's framework. Hedge portfolios formed based on the value-to-price (VP) ratio obtained from estimating NP's framework with DL achieve an alpha of up to 9.46% annually, which compares favorably to the excess return estimates reported in previous research (e.g., Piotroski 2000). Further tests show that DL's estimation of non-linearities increases alpha by 94% in comparison to linear estimators. We use OLS and quantile regression to analyze the associations between model design-choice features and alpha across the entire alpha distribution and find that performance improves with greater ratio disaggregation and either long forecast horizons paired with a focus on operating activities or short forecast horizons paired with taking financing activities into account. We find weak evidence that models based on core items and more lags of historical information perform better.

To address the possibility that positions taken by the best-performing models concentrate in small, illiquid, costly-to-trade stocks, making the returns we document unachievable, we include only NYSE/AMEX firms and use value-weighted portfolio returns. To probe further, we regress hedge portfolio excess returns on Carhart's (1997) four asset pricing factors and find that returns are unrelated to the market, size, and momentum factors, and they vary positively with the value factor.

Importantly, the returns are driven by the trade's long side. Plots of cumulative buy-and-hold returns for the short and long portfolios demonstrate that hedge portfolio performance derives from a widening spread, not from excessively negative performance of the short portfolio. In sum, the evidence indicates that the returns we document are potentially achievable in actual trades. In additional tests, we find that the model predicts returns incrementally to other factors suggested in prior research. The predictive power is pervasive throughout the sample but pronounced for small, loss-making, financially distressed, and technology firms, indicating that accounting for non-linearities is especially important when forecasting the profitability of these entities.

This paper contributes to research on financial statement analysis and accounting-based valuation in several ways. First, we confirm that current-period financial ratios exhibit complex nonlinear associations with future profitability, and that allowing for these non-linearities provides value estimates to which prices converge (Lee, Myers, and Swaminathan 1999). This evidence extends the work of Dechow, Hutton, and Sloan (1999), who find little support for Ohlson's (1995) assumption that the stochastic residual income process is linear. This finding is significant given the importance of the residual income model for valuation; after Ohlson (1995) reinvigorated the model previously developed by Preinreich (1938) and Edwards and Bell (1961), accounting researchers found that it yields more accurate valuations than discounted cash flow models (Penman and Sougiannis 1998; Frankel and Lee 1998; Francis, Olsson, and Oswald 2000). We further contribute to this literature by answering Dichev's (2020) call for more research on long-run earnings predictions using novel techniques, in particular machine learning, and integrating these forecasts into Ohlson's (1995) fundamentals-based valuation framework.

The second contribution derives from our use of NP's theory-grounded framework to discipline variable selection. Prior research uncovers ratios and accounting signals that predict returns, earnings changes, and analyst forecast revisions (Ou and Penman 1989a, 1989b; Abarbanell and

Bushee 1997; Abarbanell and Bushee 1998; Piotroski 2000). So do we, but our approach differs. Variable selection in this prior research is based on statistical fit, which NP describe (p. 125) as “trawling through the data without structure”; in contrast, NP derive the variables we consider from fundamental principles.¹⁰ We use NP’s framework, as opposed to analyzing an arbitrarily large set of possible predictor ratios chosen based on either intuition (Ou and Penman 1989) or expert judgment (Lev and Thiagarajan 1993). We do so for the reasons put forward by NP (p. 110): their structural approach “not only identifies relevant ratios, but also provides a way of organizing the analysis task,” thereby providing a way to understand the predictive ability or valuation implications of a particular ratio.

Our paper also extends research that applies machine learning to questions in accounting and financial economics.^{11,12} Several papers that use machine learning algorithms to predict earnings are not concerned with using the resulting forecasts as inputs to valuation, do not rely on a framework such as NP’s to discipline predictor selection, and do not derive specific prescriptions for financial statement analysis.¹³ We use machine learning to address the problem of nonlinear relations among the variables in a structural framework, not to search through arbitrarily large sets of possible predictors to arrive at empirically derived ratios. The latter approach might cause an analyst to overlook value-relevant information in distorted accounting numbers (Sloan 2019). Our paper shows how understanding the ways individual pieces of information are conceptually and empirically linked

¹⁰ Fairfield, Sweeney, and Yohn (1996), Fairfield and Yohn (2001), Soliman (2008), and Esplin, Hewitt, Plumlee, and Yohn (2014) evaluate components of NP’s framework by showing how disaggregating ROE into different income statement line items, margins and turnovers, and returns from operation and financing can be used to forecast profitability, return, and analyst forecast revisions. We analyze NP’s framework as a whole.

¹¹ See, e.g., Huerta, Corbacho, and Elkan (2013), Purda and Skillicorn (2015), Frankel, Jennings, and Lee (2016, 2017), Erel, Stern, Tan, and Weisbach (2018), Gu, Kelly, and Xiu (2020), Fu, Du, Guo, Liu, Dong, and Duan (2018), Aubry, Kräussl, Manso, and Spaenjers (2019), Bertomeu, Cheynel, Floyd, and Pan (2019), Ding, Lev, Peng, Sun, and Vasarhelyi (2020), Bao, Ke, Li, Yu, and Zhang (2020), Brown, Crowley, and Elliott (2020), and Avramov, Kaplanski, and Subrahmanyam (2020).

¹² The importance of machine learning in financial markets has also been noted by professional organizations such as the CFA Institute, which began including machine learning in its curriculum in 2019.

¹³ See, e.g., Callen, Kwan, Yip, and Yuan (1996), Hunt, Myers, and Myers (2019), Anand, Brunner, Ikegwu, and Sougiannis (2020), and van Binsbergen, Han, and Lopez-Lira (2020).

within a comprehensive financial statement analysis helps protect investors from fluctuations in accounting numbers induced by the reporting process (Penman 2010, Chapters 4 & 5).

We believe our application of machine learning has two distinctive features. First, in contrast to applications that use statistical fit to arrive at model inputs, our inputs are specified by the valuation theory-based framework developed by NP. We demonstrate how a specific machine learning algorithm, DL, can be used in two related accounting applications implied by NP's framework, profitability forecasting and valuation. In contrast to papers that apply machine learning to returns prediction via large, arbitrary sets of predictors, we use DL to estimate a model tied to NP's structural framework and arrive at a valuation that is used to form a trading strategy. As discussed in Armstrong (2001), complex, atheoretical statistical models tend to include noise, which causes their performance to deteriorate out of sample. Theory-grounded models are less susceptible to this threat as they incorporate an understanding of underlying mechanisms. Second, we focus on understanding what researchers and others including analysts can learn from the application of a machine learning algorithm about the processes underlying valuation and financial statement analysis, going beyond the empirical tasks of profitability prediction and equity valuation. This approach to understanding, that is, learning from artificial intelligence, is inspired by the reasoning of Chess and Go grandmasters who started to learn from the games of the artificial intelligence AlphaZero.¹⁴

2. Methodology

2.1 Structural Accounting-Based Valuation Models

Structural accounting-based valuation models express a firm's equity value as a function of expected accounting outcomes. Manipulating the definition of an expected return ($\rho_w = (V_t^E +$

¹⁴ See <https://arstechnica.com/science/2018/12/move-over-alphago-alphazero-taught-itself-to-play-three-different-games>. In a heavily publicized match, AlphaZero beat the 18-time Go world champion Lee Sedol by 4-1 in 2016. Nineteen years earlier, the artificial intelligence Deep Blue beat the then-reigning Chess world champion Garry Kasparov by 3.5 to 2.5 in 1997. Artificial intelligence has been employed to perform at a world-class level in a range of other games including all 57 Atari video games, StarCraft 2, and Quake 3.

$D_t)/V_{t-1}^E$) under the assumptions of constant expected returns and terminal convergence ($\lim_{T \rightarrow \infty} V_{t+T}^E/\rho_w^T = 0$) yields the discounted dividend model $V_0^E = \sum_{t=1}^T D_t / \rho_w^t$, where ρ_w denotes the (constant) expected or required return to common equity, V^E value to equity holders, D dividends, and all variables in periods following t expected outcomes.

Dividends, however, are hard to forecast as their payment is discretionary, so in practice, most analysts predict accounting earnings. Using the clean surplus relation $CSE_t = CSE_{t-1} + CNI_t - D_t$ and an additional terminal convergence condition ($\lim_{T \rightarrow \infty} CSE_{t+T}/\rho^T = 0$), we reformulate the discounted dividend model into the residual income model $V_0^E = CSE_0 + \sum_{t=1}^T RE_t / \rho_w^t$, where CSE denotes the book value of shareholders' equity, CNI comprehensive net income, and $RE_t = CNI_t - CSE_{t-1} \times (\rho_w - 1)$ residual income (Preinreich 1938; Edwards and Bell 1961; Ohlson 1995).¹⁵ While the residual income model links value to expected accounting outcomes, it provides no guidance on how to forecast those outcomes. While Ohlson (1995) assumes that RE is linked linearly through time to derive further theoretical implications, Dechow et al. (1999) do not find evidence supporting this assumption in annual US data.

2.1.1 Ratio Analysis and Disaggregation

In contrast to Ohlson (1995), NP's framework is agnostic about the functional form of the RE process; instead, it modifies traditional financial statement ratio analysis to decompose residual income into its drivers and relates past to future drivers empirically. Specifically, an alternative way to calculate residual income is $RE = (ROCE - \rho_w + 1) \times CSE_{t-1}$, where $ROCE (= CNI/CSE)$ denotes return on common equity. To facilitate notation, ratios and income statement accounts without subscripts denote period t quantities. This yields a ratio-based formulation of the residual income model:

¹⁵ Both US GAAP and IFRS require firms to display comprehensive income, which satisfies the clean surplus relation.

$$V_0^E = CSE_0 + \sum_{t=1}^{\infty} (ROCE - \rho_w + 1) \times CSE_{t-1} \times \rho_w^{-t}. \quad (1)$$

Figure 1 breaks ROCE into its components in increasing levels of granularity/disaggregation:

Level 1: ROCE = ROTCE × MSR: ROTCE denotes return on total common equity (= (CNI + MII)/(CSE + MI)), MSR minority sharing ratio (= $\frac{CNI/(CNI + MII)}{CSE/(CSE + MI)}$), MII minority (noncontrolling) interest income, and MI minority (noncontrolling) interest.¹⁶

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 × OLSREAD: Sales PM* denotes modified profit margin after consideration of implicit charges on supplier credit (= (Core OI from Sales + io)/Sales), ATO* modified asset turnover (= Sales/OA), OLLEV operating liability leverage (= OL/NOA), OLSREAD the spread between return on operating assets and the implicit interest rate on operating liabilities (= (OI + io)/OA – io/OL), and io the implicit interest charge on operating liabilities. The analysis reveals eight drivers of ROCE: (1) MSR, (2) FLEV, (3) NBC, (4) ATO*, (5) Sales PM*, (6) Other Items/OA, (7) OLLEV, and (8) OLSREAD:

¹⁶ We refer readers who want to learn more about the derivation of NP's decomposition to their section 2.

$$\text{ROCE} = \text{MSR} \times [\text{Sales PM}^* \times \text{ATO}^* + \frac{\text{Other Items}}{\text{OA}} + \text{OLLEV} \times \text{OLSPREAD} + \text{FLEV} \times (\text{RNOA} - \text{NBC})]. \quad (2)$$

Theory is silent as to which of the four levels should be used for prediction, making the choice of disaggregation level an empirical question. As previously explained, models with less disaggregation run the risk of missing information, and models with more disaggregation run the risk of including noise. If idiosyncratic variation of ratio components cannot be used to trade profitably, either because the measures contain too much value-irrelevant information or the relevant information is already incorporated in price (Malkiel and Fama 1970), investors would be better off not disaggregating. Thus, our first hypothesis, stated in alternative form, is:

Hypothesis 1. *Higher-level ratio disaggregation improves model performance.*

2.1.2 Core vs. Transitory Items

NP document time-series and cross-sectional properties of the eight ROCE drivers. While some, such as ATO, are persistent, others, such as RNOA deriving from unusual operating income, show mean reversion. Hence, focusing on persistent components and excluding transitory components of each ratio might enhance forecasting performance by decreasing noise that is irrelevant to prediction. Acknowledging this, NP adjust their final layer of decomposition:

$$\text{ROCE} = \text{MSR} \times [\text{Core Sales PM}^* \times \text{ATO}^* + \frac{\text{Core Other Items}}{\text{OA}} + \frac{\text{UOI}}{\text{OA}} + \text{OLLEV} \times \text{OLSPREAD} + \text{FLEV} \times (\text{Core RNOA} - \text{Core NBC} + \frac{\text{UOI}}{\text{NOA}} - \frac{\text{UFE}}{\text{NFO}})], \quad (3)$$

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 3 identifies eight relatively more persistent

drivers of ROCE: (1) MSR, (2) FLEV, (3) Core NBC, (4) ATO*, (5) Core Sales PM*, (6) Core Other Items/OA, (7) OLLEV, and (8) OLSREAD.

Empirically, it is not clear whether including items labeled as persistent/core and excluding items labeled as transitory/non-core (i.e., UOI/OA, UFE/NFO) improves forecasting. There are at least two considerations. First, management's reporting decisions may result in transitory income items that nonetheless have predictive ability. Penman and Zhang (2002) argue that under conservative accounting, management can create earnings reserves by taking actions that generate future-period (accounting) benefits while decreasing current-period income; for example, managers can increase future accounting performance by recording a current-period impairment loss. The impairment loss, classified as transitory (non-core), would be relevant for predicting future earnings. The second consideration derives from the view that the distinction between transitory (non-core) and persistent (core) income items is at least partly firm-specific, arising from features of the firm's business model. The empirical measures used by NP and in this paper are based on Compustat data definitions applied to all entities, which may result in an imperfect separation of core from non-core items for at least some firms. Thus, whether focusing on core items improves forecasts is an empirical question that motivates our second hypothesis, also stated in alternative form:

Hypothesis 2. *Focusing on core items improves model performance.*

2.1.3 Value of Operating Activities

NP propose a model simplification based on US GAAP and IFRS that require recognition or disclosure of market (fair) values of Net Financial Obligations (NFOs). If NFO fair value equals fundamental value, equation 1 can be simplified as follows:

$$V_0^E = \text{NOA}_0 - \text{NFO}_0 + \sum_{t=1}^{\infty} (\text{RNOA} - \rho_W + 1) \times \text{NOA}_{t-1} \times \rho_W^{-t}, \quad (4)$$

where NOA denotes value of net operating assets, NFO value of net financial obligations, ρ_W

weighted average cost of capital (WACC) ($\rho_W = \rho_W \times \frac{V_0^E}{(V_0^E + NFO_0)} + \rho_D \times \frac{(1-\tau) \times NFO_0}{(V_0^E + NFO_0)}$), ρ_D cost of debt, and τ marginal tax rate. Expression 4 reduces the forecasting inputs to the five drivers of RNOA attributable to common shareholders: (1) ATO*, (2) Sales PM*, (3) Other Items/OA, (4) OLLEV, and (5) OLSREAD.

As in the case of the core vs. noncore distinction, whether the simplified form of the residual income model in Expression 4 improves forecasting performance is *ex ante* unclear. The agency cost literature questions the independence of operating and financing activities, as capital structure can affect managers' incentives. As previously noted, managers of (highly) levered firms close to bankruptcy might engage in excessive risk-taking to benefit shareholders who face limited downside and unlimited upside given the call option-like payoff structure of equity (Jensen and Meckling 1976). Hence, our third hypothesis, stated in alternative form:

Hypothesis 3. *Focusing on operating activities improves model performance.*

2.1.4 Terminal Values

Equations 1 and 4 require indefinite-horizon forecasts, an infeasible task. For practicality, analysts assume that firms eventually reach steady-state residual income growth to estimate terminal value (Penman 2012; Nissim 2019). Applying a terminal growth rate g to equations 1 and 4 yields:

$$V_0^E = CSE_0 + \sum_{t=1}^T \frac{(ROCE_t - \rho_W + 1) \times CSE_{t-1}}{\rho_W^t} + \frac{(ROCE_{T+1} - \rho_W + 1) \times CSE_T}{\rho_W^T \times (\rho_W - g)} \quad (5)$$

and

$$V_0^E = NOA_0 - NFO_0 + \sum_{t=1}^T \frac{(RNOA_t - \rho_W + 1) \times NOA_{t-1}}{\rho_W^t} + \frac{(RNOA_{T+1} - \rho_W + 1) \times NOA_T}{\rho_W^T \times (\rho_W - g)}. \quad (6)$$

The appropriate choice of T is *ex ante* unclear. While one-, five- and 10-year horizons are common in practice, Nissim (2019) finds that a 10-year horizon yields the terminal value estimate closest to observed price in the terminal year. This motivates our fourth hypothesis, stated in alternative form:

Hypothesis 4. *Expanding the prediction horizon improves model performance.*

Lastly, using longer-past outcomes as predictors increases the likelihood that (1) a firm's activities have changed sufficiently to reduce the information to noise, or (2) the information in the reports is already incorporated in price. In the words of Black (1986), “[t]rading on that kind of information will be just like trading on noise.” Hence, our fifth and final hypothesis, stated in alternative form:

Hypothesis 5. *Using more predictor lags improves model performance.*

We acknowledge that, as a practical matter, a researcher or analyst must make other choices in designing a financial statement analysis, including the reinvestment rate, discount rates ρ_w and ρ_W , and growth rate g . Because the number of models to be estimated grows multiplicatively with the number of choices, we focus only on the five design choices we believe to be most important.¹⁷ In our empirical implementation, we invoke Miller and Modigliani (1961) principles and set the reinvestment rate to one. Given the lack of an agreed-upon asset pricing model, we follow NP's agnostic approach and set the discount rate to the one-year Treasury bill yield plus 5%. In line with historical GDP figures, we assume a 2% growth rate (e.g., Penman and Sougiannis 1998).

Expressions 5 and 6 link value to ratios, and financial statement analysis links past realizations to future realizations of ratios. Because both links or relations are non-linear, it is necessary to use methods/tools that can accommodate complex nonlinear associations to apply financial statement

¹⁷ Testing Hypotheses 1 to 5 requires estimation of 4 (number of disaggregation levels) \times 2 (core vs. non-core) \times 2 (ROCE vs. RNOA) \times 3 (1-, 5-, and 10-year horizons) \times 4 (0, 1, 3, and 5 ratio lags) \times 29 (number of years for which we estimate each model) = 5,568 models.

analysis in forecasting and value estimation. The next section introduces such a tool: Deep Learning, a special case of a neural network machine learning algorithm.

2.2 Deep Learning

A neural network generalizes simpler 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) non-linear 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 or $(1 + e^x)^{-1}$, respectively.

A neural network organizes relations among inputs into layers ($j = 1, 2, \dots, J$). Each layer j , including the input layer $j = 0$ (the independent variables) and output layer $j = J$ (the dependent variables), consists of a series of neurons. Layers between the input and output layers are called hidden layers. A neural network with one (multiple) hidden layer(s) is referred to as shallow (deep). Deep neural networks are also referred to as Deep Learning (DL) algorithms. While the connections between layers can be set arbitrarily, the most common implementations are sequential models that fully connect each neuron in the preceding layer to each neuron in the subsequent 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 and output layers, constant activation functions, and a fixed number of neurons ($i = 1, 2, \dots, I$) in each hidden layer.¹⁸ By combining the activation functions with a series of layers in this way, the neural

¹⁸ As is standard in the recent literature, we use rectified linear units activation functions (ReLUs), $h_{\theta}(x) = \max(0, x)$, which make computation more efficient relative to other possible choices such as sigmoid or hyperbolic tangent activation functions (Glorot, Bordes, and Bengio 2011).

network can model complex non-linear relations without the need to specify a functional form.

Figure 2 Panels A to C compare three special cases of neural networks: a single-layered network with one predictor and one outcome variable; a single-layered network with eight predictors and one outcome variable; and a multi-layered model with two hidden layers, 10 neurons per layer, eight predictors, and five outcome variables. Circles symbolize neurons, and lines indicate connections among neurons. The Panel A model resembles the random walk model recommended by Watts and Leftwich (1977), who find that it outperforms more complicated earnings prediction schemes in time-series regressions. The Panel B model resembles the model of Hou, Van Dijk, and Zhang (2012), who argue that earnings prediction can be enhanced by adding additional predictors such as accruals and the book-to-market ratio in a pooled cross-sectional model. The Panel C model resembles a more general neural network with the potential to model the non-linearities and interactive effects prevalent in NP's framework by adding hidden layers. The figure highlights the neural network's ability to jointly and simultaneously predict multiple outcome variables, which, as NP argue, can be useful when the predictions are mutually dependent, as in the case of earnings prediction. The number of equations describing a layer in a fully connected network is determined by the number of neurons in the layer and its preceding layer. Thus, in standard notation, layer j is modeled as:

$$\begin{aligned}
 a_1^{(j)} &= h_{\theta}(\theta_{10}^{(j-1)} + \sum_{i=1}^{I_{j-1}} \theta_{1i}^{(j-1)} a_i^{(j-1)}) \\
 a_2^{(j)} &= h_{\theta}(\theta_{20}^{(j-1)} + \sum_{i=1}^{I_{j-1}} \theta_{2i}^{(j-1)} a_i^{(j-1)}) \\
 &\vdots \\
 a_{I_j}^{(j)} &= h_{\theta}(\theta_{I_j 0}^{(j-1)} + \sum_{i=1}^{I_{j-1}} \theta_{I_j i}^{(j-1)} a_i^{(j-1)})
 \end{aligned} \tag{7}$$

where $a_i^{(j)}$ denotes neuron i in hidden layer j , h_{θ} the ReLU activation function, and $\theta_{ij}^{(j-1)}$ the slope coefficient connecting neuron i in layer $j - 1$ to neuron i_j in layer j . The system of equations is more compactly expressed in matrix notation:

$$a_i^{(j)} = h_{\theta}(\theta^{(j-1)} a^{(j-1)}) \quad (8)$$

such that $\dim(\theta^{(j-1)}) = 1 \times (I + 1)$ and $\dim(a^{(j-1)}) = I \times 1$ where $\dim(X)$ returns the dimension of matrix X . To illustrate, for the model depicted in Figure 2 Panel A, $J = 2$, $I_1 = 1$, and $I_2 = 1$; in Figure 2 Panel B, $J = 2$, $I_1 = 8$, and $I_2 = 1$; and in Figure 2 Panel C, $J = 4$, $I_1 = 8$, $I_2 = I_3 = 10$, and $I_4 = 5$.

Neural networks are estimated iteratively through minimizing a loss function $g_{\theta}(a^{(j)}, \hat{a}^{(j)})$ where hat symbols indicate predicted outcomes. Following Easton, Kapons, Kelly, and Neuhierl (2020), who find that minimizing mean absolute deviation yields better results for fundamental forecasting than minimizing squared deviation, we set $g_{\theta}(a^{(j)}, \hat{a}^{(j)}) = |a^{(j)} - \hat{a}^{(j)}|$. To specify the process through which the minimum is found, the researcher must choose an optimizer. Following the recent literature, we use the Adam optimizer, which has been shown to return local minima more efficiently and more precisely than other popular optimization algorithms (Kingma and Ba 2014).

As mentioned previously, neural networks can estimate arbitrarily complex functions among independent and dependent variables. While in principle one could achieve the same outcome by including higher-order polynomials and interaction terms in linear regression models, the dimensionality of this problem quickly makes estimation infeasible. If a model has 10 independent variables, a simple linear regression 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 $1 + 10 + 10 + 10 + 29! = 8.84 \times e^{30}$. Once the number of parameters exceeds the number of observations in the dataset, the model is either not estimable via Ordinary Least Squares (OLS) or the estimates will behave poorly (Huber 1973). In contrast, neural networks allow researchers to capture higher-order and interactive relations without explicitly specifying them. The combination

of hidden layers connected through non-linear 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 non-linear relationships between the past and future value-determining fundamentals discussed in the previous section.¹⁹ Indeed, as shown in Figure 3, the popularity (as measured by web queries analyzed via Google Trends) of neural networks more generally and DL more specifically has increased substantially relative to other prominent machine learning algorithms previously used in the accounting literature, such as Lasso Regressions and Random Forests.

We implement the neural network using Google's Tensorflow API. We choose a fully connected deep neural network architecture with five hidden layers and 500 neurons per layer, a structure that allows the model to estimate complex, non-linear relationships among variables.²⁰ To address overfitting concerns, we add the regularization term $\lambda \sum_{i,j,k} (\theta_{i,j}^k)^2$ to the value function. This practice biases parameter estimates towards zero, thereby mitigating the influence of noise on our estimates (Friedman, Hastie, and Tibshirani 2001). We estimate each neural network over 50 epochs using learning (α) and regularization (λ) rates of 10^{-3} and 10^{-5} . This standard architecture is able to capture complex non-linearities while also being simple enough to be estimated in six to seven

¹⁹ For an illustration of how a neural network estimates arbitrarily complex higher-order and interactive relationships among variables, see playground.tensorflow.org.

²⁰ While there is no agreed-upon method to choose a neural network's numbers of neurons and hidden layers, we use a simplification of the formulas discussed in Hagan, Demuth, Beale, and De Jesús (1996) to choose the number of neurons in the model: $N_h = 0.5 \times N_s / (N_i + N_o) = 0.5 \times 55972 / (9 + 1) \approx 2,799$, where N_h denotes the recommended upper bound for the number of neurons, N_i the number of variables in the input layer, and N_o the number of variables in the output layer. Bengio (2009) argues that neural networks need multiple hidden layers to represent the high levels of abstraction required for many artificial intelligence tasks. While traditional architectures were confined to one to three hidden layers, He, Zhang, Ren, and Sun (2016) propose an architecture with up to 1000 hidden layers. Given the discussion above and the computational challenges discussed in previous footnotes, the architecture of five hidden layers and 500 neurons per layer that we choose for this paper strikes a compromise to allow computation of complex relationships while maintaining computational feasibility.

minutes with a single GPU and 15 GB of RAM on Google Cloud's AI Platform.²¹

2.3 Model Evaluation

Because NP's framework is designed to aid in equity valuation, and assuming the investor wishes to maximize risk-adjusted return, we compare model performance using future portfolio returns. We measure portfolio performance as alphas from a Jensen, Black, and Scholes (1972) time-series asset pricing model. We calculate each model's alpha as follows:

1. Calculate a firm's Value to Price (VP) ratio. Value is computed using the accounting-based valuation models described in section 2.1. Valuation inputs are obtained using the deep neural networks under the financial statement analysis design choices described in section 2.2. To ensure that all information is available at the time of portfolio formation in year t , forecasting models are estimated from year $t - \text{Leads} - 9$ to year $t - \text{Leads}$, where Leads is the number of ROCE or RNOA leads the model is designed to forecast. The estimation period of 10 years is chosen in line with Hou et al. (2012).²² Figure 4 shows the timeline. To ensure that differences in model performance are driven by choice of fundamentals and not by differences in sample composition, we condition the estimation sample into firm-year observations for which 10 leads and five lags

²¹ Given that we estimate 192 (number of different design choice combinations) $\times 29$ (number of years for which we estimate each model) = $5,568$ models, more complex model architectures become prohibitively costly to estimate. However, while we estimate all $5,568$ models to analyze the effects of fundamental selection, researchers or practitioners applying the model in real time can use the findings in this paper to select the appropriate model for their needs. Thus, they have to estimate only one model, which makes more complex architectures feasible. We use the scikit-optimize Bayesian optimization API (scikit-optimize.github.io) to find the optimal architecture for an RNOA-based, level 4 disaggregation model that uses five years of lagged information, forecasts 10 years of RNOA, and focuses on core items over a grid of $\mathbb{Z} \in [0,10]$ for the number of hidden layers, $\mathbb{Z} \in [1,1000]$ for the number of neurons per layer, $\mathbb{Z} \in [1,1000]$ for the number of epochs, and $\mathbb{R} \in [10^{-12}, 10^{-1}]$ for the learning and regularization rates. We pool all observations and, as is common in the literature, split the sample into an estimation sample consisting of 90% of all observations and a validation sample consisting of the remaining 10% of observations. We estimate the model using the estimation sample and apply the model to forecast RNOA for the validation sample. Performance is measured as mean absolute RNOA forecast error. The search yields an optimal architecture that consists of 10 hidden layers, 1000 neurons per layer, 100 epochs, and learning and regularization rates of 4.27×10^{-4} and 1.99×10^{-10} .

²² If the parameters to be estimated vary predictably by industry, estimating the model by industry would improve forecasting. However, Damodaran (2007) argues that such variation is not predicted by valuation theory, and Fairfield, Ramnath, and Yohn (2009) do not find that estimating earnings forecasting models by industry yields more accurate predictions.

for each predictor variable used in a level 4 disaggregation model are available.

2. Following Fama and French (1993), at the end of June in year $t + 1$, form decile portfolios using the VP ratios based on accounting data for fiscal year t .²³ The first portfolios are formed in 1988 (1963 plus a maximum lead of 10 years plus a maximum lag of five years plus a sample estimation period of 10 years).
3. Calculate future annual value-weighted portfolio returns from July in year $t + 1$ to June in year $t + 2$.²⁴ The portfolio holding period is shown in Figure 4.²⁵
4. Calculate excess returns by subtracting returns on the risk-free security and the lowest decile portfolio from the return on the highest decile portfolio.
5. Regress excess return on the risk factors proposed in Carhart (1997).²⁶ Compute model alpha as the regression intercept estimate.

This approach has two benefits. First, it allows the researcher to compare models based on ROCE with models based on RNOA. Given their differing distributional properties as shown in Tables 1 and 3, it is unclear how to analyze the relative performance of ROCE-based and RNOA-based models in other ways, such as comparing absolute forecast error or bias (Hou et al. 2012; Evans, Njoroge, and Yong 2017). Second, if earnings outcomes are influenced by unforecastable macroeconomic shocks (Ball, Sadka, and Sadka 2009; Bonsall, Bozanic, and Fischer 2013), earnings expectations – not earnings realizations – determine stock prices. Because of the lack of a generally agreed-upon asset pricing

²³ Using median, terciles or quintiles instead of deciles does not affect our inferences.

²⁴ As noted by Loughran and Ritter (2000), using value-weighted instead of equal-weighted returns helps to ensure that results are not driven by small, illiquid stocks that are costly to trade. Tests based on value-weighted returns hence offer less power to detect mispricing.

²⁵ Our results are robust to using monthly instead of annual returns.

²⁶ The more recent Fama and French (2015) five factor model adds a factor for profitability, precisely the variable this paper is concerned with. To avoid multi-collinearity issues, we thus use the Carhart (1997) four factor model and note that the goal of this paper is not to document a novel asset pricing anomaly, but to understand the decision usefulness of different financial statement analysis choices. However, we redo our analysis using alternative factor models including the CAPM, the Fama and French (1993) three factor model, the Fama and French (2015) five factor model, and a simple unadjusted excess return model. Our inferences remain unchanged. We are unable to determine whether the models' abnormal return performance is due to mispricing or whether they identify another risk factor not captured by various factor models.

model, we are agnostic about individual models' alpha magnitudes and, as proposed by Jackwerth and Slavutskaya (2018), we focus on comparing the *relative* alpha magnitudes across models to test the hypotheses presented in section 2.1.

3. Data and Descriptive Evidence

3.1 Data

Following NP, we use annual Compustat data from 1963 to 2017 for NYSE and AMEX firms. We retain observations with five lags for all required variables and non-negative values for CSE, NOA, OA, and OL at fiscal-year beginning and end to ensure that our results are not driven by different sample compositions across models and meaningful values for computed ratios. Stock return data and asset pricing factor data come from CRSP and Ken French's website, respectively.

3.2 Summary Statistics

Table 1 Panel A presents descriptive statistics for price and valuation anchors. Following Fama and French (1993), we measure price as the firm's market value of equity on the last day of June in the year succeeding the fiscal year of portfolio formation. Price's mean and median exceed those of CSE and NOA, indicating that the market, on average, expects firms to yield positive future residual income. For most firms NFO is positive, meaning that financial obligations exceed financial assets. However, the first percentile of the NFO distribution is negative; some firms' financial assets exceed their financial obligations.

Table 1 Panel B replicates and extends NP's Table 1; although the table includes 17 additional years, the numbers are generally similar, with a few exceptions. First, and consistent with the results in Givoly and Hayn (2000), the number of loss firms is increasing and profitability is decreasing over time. Mean ROCE in the extended sample is 1.2% lower, and its standard deviation is 3.8% higher. RNOA's distribution, on the other hand, changed little. This difference could be due to increases in financial leverage, as documented in Schularick and Taylor (2012). FLEV's mean is 9% higher, and its

standard deviation increased from 0.87 to 1.14 in the more recent sample. NBC and ATO are similar over time, while sales PM's mean is 1% higher and its standard deviation is 5% higher than the corresponding NP statistics. Core items are on average close to their non-core counterparts. Unusual items such as Core Other Items/OA, UOI/NOA, and UFE/NFO have means close to zero, with sizable 1st and 99th percentiles. OLLEV has a larger mean and a similar median in the more recent sample. It appears that some firms have greater operating leverage after 2000 than previously. In contrast, OLSPEAD's distribution does not change much.

Table 2 presents correlations for selected ratios, with Pearson (Spearman) correlations above (below) the diagonal. Several correlations are economically and statistically significant, suggesting the usefulness of NP's framework for prediction. ROCE and ROTCE are close to perfectly correlated. RNOA, however, exhibits idiosyncratic variation. Its Pearson (Spearman) correlations with ROCE and ROTCE are 0.78 (0.90) and 0.77 (0.90), respectively. Sales PM is more strongly correlated with contemporaneous non-core profitability measures than Core Sales PM, while the opposite holds for Core RNOA. While Fairfield and Yohn (2001) find that turnovers are better profitability predictors than margins, in Table 2 margins are more strongly correlated with contemporaneous profitability than turnovers. OLLEV correlates more strongly with RNOA than with ROCE, a potential by-product of cross-sectional variation in financial leverage confounding the univariate relation between OLLEV and ROCE. Lastly, financial leverage and borrowing cost are positively correlated, albeit not strongly. Firms that take on higher financial risk have to compensate their lenders with higher interest payments (Penman, Reggiani, Richardson, and Tuna 2017).

3.3 Non-Linearities

Figure 5 Panels A to H present evidence on the relation between future profitability and contemporaneous ratios, by plotting median portfolio ROCE in periods t to $t+10$ by ROCE, FLEV, SPREAD, ATO, Sales PM, OLLEV, OLSPEAD, and RNOA decile in period t . Except for OLLEV,

the plots provide visual evidence of non-linear 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. These functional forms are visible over 10 years, attenuating over time. Figure 6 Panels A to D present examples of interactive relations across ratios in predicting future ROCE. The surfaces obtained from plotting OLLEV and ATO, FLEV and ATO, SPREAD and Sales PM, and OLSPREAD and Sales PM on the X and Y axes and lead 1 (Panels A and B) and lead 5 (Panels C and D) ROCE on the Z axis exhibit curvatures that are visually different from the straight plane one would observe under linear, non-interactive relations. In sum, the visual evidence in Figures 5 and 6 suggests non-linearities in the dynamic relations across several fundamental ratios and subsequent profitability.

It would be difficult to specify these (visually) nonlinear functional forms in a linear model based on accounting or financial statement analysis intuition, which makes flexible machine learning algorithms the appropriate estimation tool. While other algorithms such as Random Forests or Gradient Boosted Trees can handle non-linearities, we focus on DL because it is readily able to approximate any functional form and, as evidenced by Figure 3, because it is widely used in practice (Hornik et al. 1989; Schmidhuber 2015; Huang, Jin, Gao, Thung, and Shen 2016). We acknowledge that in our finite sample, other algorithms or some combination of algorithms might produce more accurate profitability forecasts. However, our main concern is the estimation of non-linear relations within NP's framework to derive implications for financial statement analysis, not analyzing different machine learning algorithms to determine which one yields the most accurate predictions.²⁷

4. Results

4.1 Forecast Errors

We first analyze whether estimating NP's framework via DL yields more accurate out-of-

²⁷ Examples of papers that take an algorithm-comparison approach include Hunt et al. (2019) and Anand et al. (2020).

sample profitability forecasts than a random walk. As discussed in Monahan (2018), a random walk tends to yield more accurate predictions than other more complex earnings forecasting models, which justifies its use as a benchmark in our setting. Table 3 Panels A and B, which show one-year-ahead means of average absolute ROCE and RNOA out-of-sample forecast errors for each of our 192 models and comparisons to random walk forecast errors, indicate that most of the models outperform the random walk.

4.2 Alphas

Table 4 reports Carhart (1997) alphas computed as discussed in section 2.3. Panel A (Panel B) presents results for ROCE- (RNOA-) based models. Following Newey and West (1986), we compute heteroscedasticity and autocorrelation robust standard errors with a lag order of $4 \times (29/100)^{2/9} \approx 3$. For both matrices, alphas increase from the upper left to the lower right of each panel; that is, higher levels of ratio disaggregation, focusing on core items, expanding the forecast horizon, and using more lagged information are all positively correlated with alpha. Low disaggregation models that incorporate few leads and lags perform poorly. Alphas of ROCE models are smaller in absolute magnitude, are significantly positive in only five out of 96 cases, and are less dispersed than their RNOA counterparts. While RNOA-based alphas range from -9.61% to 9.46%, the ROCE-based alphas range from -3.23% to 5.29%. However, alphas for 10-year forecast horizon RNOA-based models are generally higher than their ROCE-based counterparts. In contrast, RNOA-based alphas forecasting one year ahead are significantly negative, while those based on ROCE are not statistically different from zero.²⁸ Together,

²⁸ We compare the characteristics of firms in the Table 4 Panel B short- and long-horizon model portfolios to determine why alphas for short-horizon models are negative. Value estimates in short-horizon models largely derive from NOA, the valuation anchor. In consequence, these models tend to focus on capital intensive firms with high financial leverage, low asset turnovers, and declining profitability. In contrast, long-horizon value estimates derive largely from increases in residual operating income. Short-horizon models' reliance on the valuation anchor results from the research design choice to apply the same terminal growth rate to all firms. As a result, short-horizon models that do not take account of the growth trajectory tend to underestimate the value of firms with strong growth prospects. The long-horizon models avoid these problems by expanding the forecast horizon. We do not attempt to estimate firm-specific terminal growth rates both because it is unclear how to measure these growth rates and because the appropriate year for the terminal value calculation likely differs by firm.

these findings suggest that independent consideration of operating and financing activities improves (deteriorates) model performance for long (short) forecast horizons. In the short run, financing activities are informative about future operating performance; in the long run, with increasing model complexity, this relation turns into noise.

Table 4 Panel A (Panel B) shows alphas obtained from ROCE-based (RNOA-based) models under different financial statement analysis design choices. Evaluating the effects of different financial statement analysis design choices individually runs the risk of confusing the effect of one design choice for that of another. In contrast, regression coefficients measure the effects of an individual financial statement analysis design choice holding others constant. To capture as much as practicable of the information contained in the alpha distributions shown in Table 4 Panels A and B, we use quantile regressions in addition to OLS regressions. OLS results, which estimate the effect of analysis-design choices on the conditional mean of the absolute forecast error distribution, are shown in column 1. The results of quantile regressions, which estimate the effect of financial statement analysis choice on the conditional 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles of the absolute forecast error distribution, are shown in columns 2 to 8. Robust White (1980) t-statistics are presented in parentheses.²⁹

Table 4 Panel C tests Hypotheses 1 to 5 by regressing alpha on indicators for the financial statement analysis design choices the model is based on. Specifically, *Level 2 (3, 4)* is an indicator that the profitability forecasting model uses the level 2 (3, 4) disaggregation illustrated in Figure 1. *Core* is an indicator that the model excludes transitory items from the ratios used for prediction. *Leads 5 (10)*

²⁹ We are running regressions on the outputs of a series of machine learning models that vary only in their inputs. We therefore know the true functional form (which can be observed in Panels A and B) that generates these results, but we cannot run a regression with this functional form because the number of observations would always be equal to the number of independent variables to estimate. The regressions in Table 4 Panel C are meant to serve as linear approximations, which facilitates interpretation of our results. Because we are using a linear approximation, the quantile regressions help to uncover potentially deceptive conditional mean results. For example, a specific financial statement analysis design choice could make the best models better and the worst models worse, with a resulting mean effect of zero. Our presentation allows the reader to see the effect of including an input on different parts of the alpha distribution.

is an indicator that the model predicts profitability 5 (10) years ahead. *Lags 1 (3, 5)* is an indicator that the model uses ratios from the current and preceding 1 (3, 5) years.

As discussed before, the focus on alphas allows us to directly compare the performance of RNOA-based and ROCE-based models. To do so, we pool the alphas in Table 4 Panels A and B into one regression and include *ROCE*, an indicator for ROCE-based models (i.e., an indicator that the valuation model is based on equation (6) as one of the equity value drivers. Given the strong interactive relation between RNOA-based models and forecasts horizon apparent in Table 4 Panel B, we add interaction terms between *ROCE* and *Leads 5* and *Leads 10* to avoid model misspecification.

4.2.1 Hypothesis 1: Disaggregation Level

We find evidence that level 2 to 4 disaggregation improves model performance on average and across all percentiles relative to level 1 disaggregation. The slope coefficients generally increase from level 2 disaggregation, which incorporates financial leverage, to level 3 disaggregation, which incorporates margins and turnovers, to level 4 disaggregation, which incorporates operating leverage. The economic magnitude of these effects is large. On average, level 2, 3, and 4 disaggregation increases alpha by 2.3, 2.4, and 2.8%.

4.2.2 Hypothesis 2: Core Items

Core's slope coefficient is significantly positive for the 50th percentile but not significant at conventional levels for other models. Thus, inconsistent with Hypothesis 2 and at least some conventional wisdom, we do not find strong evidence that focusing on core items improves trading profitability. Although there are potential statistical and accounting-based explanations (including those discussed in section 2.1.2) of why excluding non-core items might not enhance model performance, examining these explanations is beyond the scope of this paper.

4.2.3 Hypothesis 3: Operating Activities

ROCE's slope coefficients are strongly positive and significant across all models. On average,

ROCE-based models yield alphas 9.3% higher than RNOA-based models when the forecast horizon is one year. In contrast, its interactions with *Lead 5* and *Lead 10* are significantly negative and of larger absolute magnitude than *ROCE*'s main effect. On average, RNOA-based models yield alphas that are 2.5% ($= 11.8 - 9.3$) and 2.1% ($= 11.4 - 9.3$) higher than those of ROCE-based models when the forecast horizon is 10 years and five years, respectively; 9.3% lower when the forecast horizon is one year. We interpret these results as follows: First, analysts who forecast over short horizons should not focus only on operating activities, taking the value of financing activities as given. They are better off using ROCE-based models. Second, analysts who forecast over longer horizons are better off focusing on operating activities and taking the value of financing activities as given. This result suggests that the Miller and Modigliani (1961) principles invoked in the derivation of equation (4) to abstract away from potential financing frictions are appropriate (inappropriate) for valuation models based on long (short) forecast horizons. It also provides empirical support for the advice frequently given in valuation textbooks to forecast as far in the future as possible when using such models (e.g., Penman 2012). A potential *ex post* explanation is that financing-induced agency-cost issues are linked to firm value in the short run but not in the long run. However, further examination of this conjecture is beyond the scope of this paper, the primary purpose of which is to evaluate financial statement analysis design choices within NP's structural framework.

To summarize, in tests of Hypothesis 3, we find an interaction with the design choice of forecast horizon: focusing on operating activities increases alpha based on forecasts over long horizons, but if the forecasting horizon is short, an analyst should account for the effects of financing activities.

4.2.4 Hypothesis 4: Forecast Horizon

On average and across all percentiles, we find evidence that increasing forecast horizon improves model performance for RNOA-based models, with larger effects for longer forecast

horizons. F-tests reject the equality of the 5-leads and 10-leads slope coefficients at the 1% level across all percentiles. The effect of the forecast horizon is much smaller for ROCE-based models. As the results of the tests of Hypothesis 3 hint, we find that the positive main effect of different horizons is approximately offset by the negative interactive effect with *ROCE*. In terms of economic magnitude, on average, forecasting 10 years rather than one year ahead increases alpha by 13.3% for RNOA-based models and by 1.5% (= 13.3 – 11.8) for ROCE-based models. To summarize, tests of Hypothesis 4 show that increasing forecast horizon improves model performance, especially for RNOA-based models.

4.2.5 Hypothesis 5: Historical Information

Despite textbooks' recommendations to use as much historical information as feasible to form forecasts, we find at best limited evidence that using more historical information improves model performance. Using last-period financial information combined with current-period information increases alpha by 0.9, 0.9, and 0.5% for the 50th, 75th, and 95th percentiles. All other slope coefficients are insignificant at conventional levels.

4.2.6 Summary

We find evidence supporting Hypothesis 1, which points to the benefits of disaggregation. We do not find evidence supporting Hypothesis 2, which suggests that the core vs. non-core distinction is not crucial for the financial statement analyses we consider. We find evidence for an interaction between the design choices analyzed by tests of Hypotheses 3 (operating vs. financing activities) and 4 (short vs. long forecast horizons). When the forecast horizon is long, focusing on operating activities improves model performance; when the forecast horizon is short, joint consideration of operating and financing activities improves model performance. We find limited and weak evidence that using more historical information increases alpha.

4.3 Additional Tests

4.3.1 Importance of Non-Linearities for Financial Statement Analysis

We explore the interaction between inherent nonlinearities in the inputs to NP's framework and the five financial statement analysis design choices we consider. Figure 1 and equations 2 and 3 reveal multiplicative relations between certain ratios and profitability, with the number of such relations increasing in the level of disaggregation. Further, visual evidence in Figures 5 and 6 indicates non-linear associations between ratios and profitability that attenuate over a 10-year horizon. This evidence suggests that consideration of non-linearities is important for the choices of forecast horizon and the number of historical periods to include in the short run, but decreasingly important in the long run.

To provide evidence on the importance of considering non-linearities, we re-estimate the 192 models analyzed in Table 4 using OLS and compute alpha for each model as previously described. As in Table 4, we regress the alpha estimates of all 384 estimates (192 DL estimates plus 192 OLS estimates) on fundamental indicators using OLS and quantile regressions, including interaction terms between each fundamental indicator and an indicator that the model is estimated via DL, *DL*. Table 5 presents the results. Main effects are suppressed to enhance readability.

We find that consideration of non-linearities improves average performance by 2.4% for level 3 disaggregation, with mixed evidence for level 2 and 4 disaggregations. As Figure 1 shows, this result may be due to the multiplicative relation between Sales PM and ATO. However, for level 2 disaggregation, the interactive effect with *DL* is positive only for the 75th percentile; for level 4, it is negative for the 10th, 75th, 90th and 95th percentiles.

The interaction terms between *DL* and *ROCE*, and *DL* and *Lead 5* and *Lead 10* load positively, but the triple interaction between *DL*, *ROCE*, and *Lead 10* loads negatively on average. This suggests that non-linearities are important for short-horizon models that account for financing activities and

long-horizon models that focus on operating activities. Considering non-linearities increases the alphas of ROCE-based models that forecast one year ahead and RNOA-based models that forecast five years ahead by 1.3% and 2.2%, respectively.

We find little evidence that consideration of non-linearities is important for models using more historical information. Only the interaction between *DL* and *Lag 1* loads positively for the 50th and 75th percentiles of the alpha distribution, with magnitudes of 0.9% and 0.7%, respectively. All other interaction terms are statistically insignificant. We find no evidence that consideration of non-linearities is important for the effect of core activities.

In sum, we find evidence that consideration of non-linearities affects model performance for all financial statement analysis design choices except for focusing on core items. The results are most pronounced for RNOA-based models with longer forecasting horizons.

4.3.2 Trading Cost

We have designed the trading strategies we describe to reduce or eliminate certain obstacles to implementation: including NYSE/AMEX stocks and excluding NASDAQ stocks, conditioning on firms with at least five lags of fundamental data, using value-weighted returns, and taking positions in June of the following year to ensure that all accounting information is available. Nevertheless, the alphas presented might be unrealistic if most of the returns come from short positions.³⁰ Short-selling is expensive, and shares to short might be unavailable in the slow-moving over-the-counter short market (Lee and So 2015). To evaluate this possibility, we probe the performance of the model with the highest alpha, i.e., a five-year horizon RNOA-based level 3 disaggregation model without a focus on core items and incorporating only the ratios available from the most recent financial statements. We first estimate the Carhart (1997) time-series regression for this model to understand which factors drive its performance; results for equal and value-weighted portfolios are presented in Table 6

³⁰ Engle, Ferstenberg, and Russell (2012) estimate overall one-way transaction cost for NYSE stocks at nine basis points.

Columns 1 and 2. The model's hedge portfolio returns do not move with the market factor ($Mkt-Rf$) and move positively with the value factor (HML); performance is not driven by small stocks, as is evident from the insignificant association with the size factor (SMB). The model's performance does not vary with the momentum factor (UMD) proposed in Jegadeesh and Titman (1993). The portfolio does not focus on stocks whose price dramatically increased or decreased recently.

Next, Figure 7 Panel A, which plots the cumulative long and short portfolio returns over 1988 to 2017 along with the cumulative CRSP market return over the same period, provides visual evidence that the portfolio returns are driven by the long side. Indeed, while the cumulative return difference between the two portfolios increases over time, there is no evidence of extreme negative returns for the short portfolio. Further, the return difference between the short and long portfolios increases over time, and the long portfolio return exceeds the short portfolio return in 22 years and falls short in 7 years (1989, 1995, 1999, 2011, 2013, 2015, 2017). The cumulative return of the long portfolio exceeds that of the short portfolio every year, providing additional support for the view that the returns are not driven by risk factors. The cumulative return on the short portfolio initially equals that of the CRSP market portfolio but falls short starting in 2006. We conclude that the hedge portfolio returns appear to be driven by long positions in medium-sized firms trading on major exchanges, providing support for the practicability of the trading strategy.

4.3.3 Comparison to Other Estimators

DL is more computationally expensive than traditional estimators such as OLS or LAD; while computing power has substantially increased, estimating the deep neural networks analyzed in this paper remains time consuming. For comparison, while modern hardware computes the OLS and LAD estimators in seconds, the DL estimator used in this paper requires six to seven minutes of computing

time.³¹ In this subsection, we provide evidence that DL provides benefits as compared to the traditional OLS and LAD estimators, helping to justify DL's incremental computing cost.

Table 6 Columns 3 to 6 present the results of estimating the full model via OLS and LAD for equal-weighted and value-weighted returns. Alphas are insignificant across all models. For value-weighted returns, DL yields 94% ($= 0.095/0.049 - 1$) higher returns than the linear LAD estimator. Figure 7 Panel B plots the cumulative long and short portfolio returns for portfolios based on OLS. While the cumulative return of the long portfolio exceeds that of the short portfolio over 1989-2017, that return is close to the return on the market portfolio. In sum, the DL estimator consistently outperforms OLS and LAD estimators, supporting the view that consideration of non-linearities is important when estimating NP's model, even at the cost of additional computing time.

4.3.4 Panel Data Regressions

We extend our analysis by applying an alternative method, panel data regressions, to test the ability of the VP measure to predict stock returns.³² Table 7 column 1 presents the results of regressing one-year-ahead returns on VP decile ranks (*VP Rank*) computed using the best-performing model and controls including log market value of equity (*Size*), market-to-book ratio (*MtB*), and current-year stock return (*Ret*). We control for firm and year fixed effects. Standard errors are clustered at the firm level. Consistent with Fama and French (1992), *Size*'s slope coefficient is negative. *Ret*'s negative slope coefficient replicates the results of Jegadeesh and Titman's (1993) Table 7 showing that return momentum reverses in the following year. However, *MtB*'s slope coefficient is not significantly different from zero, possibly because of the strong positive relation between the portfolio return and

³¹ While it may take six to seven minutes to train the model, the model generates a prediction in less than a second once the weights are established. The total DL estimation time, therefore, is not prohibitive to generating a trading strategy at scale.

³² Inconsistent with the notion that only systematic risk is priced, Daniel and Titman (1997), Daniel, Grinblatt, Titman, and Wermers (1997), and Hirshleifer, Hou, and Teoh (2012) find that firm-specific market-to-book ratios and accruals predict returns better than their factor mimicking portfolios constructed following Jensen et al. (1972).

the value factor documented in Table 6. *VP Rank*'s slope coefficient is statistically ($p < 0.01$) and economically significant. Ceteris paribus, a stock in the highest VP decile experiences a ($0.5\% \times 9 = 4.5\%$) higher annual return than a stock in the lowest VP decile. Column 2 tests whether this finding is robust to the inclusion of two additional variables that have been shown to predict returns: F-Score (*F-Score*) and Accruals (ΔNOA) measured as change in NOA scaled by average total assets (Piotroski 2000; Richardson, Sloan, Soliman, and Tuna 2005). Prior literature reports that firms with stronger fundamentals and lower accruals have higher future returns. We find, however, that *VP Rank*'s slope coefficient is unaffected by adding *F-Score* and ΔNOA .

We next examine which types of firms' valuations benefit most from considering non-linearities. Specifically, we test whether *VP Rank*'s predictive power is pronounced for small firms, loss-making firms, technology firms, and firms in financial distress. Small, loss-making and financially distressed firms might be exposed to more bankruptcy risk; however, limited liability protects equity investors through the abandonment option, resulting in a non-linear payoff structure (Hayn 1995; Fama and French 1993). Further, as shown in Barth, Li, and McClure (2019), the relation between value and fundamentals requires accounting for non-linearities, especially for technology, loss-making, and small young firms. Therefore, models that account for non-linearities might provide more accurate valuations for these entities.

Table 7 Columns 3 to 6 present the results of this analysis. Loss firms (*Loss*) are firms with negative ROCE; technology firms (*Tech*) are firms in the Computers, Electronic Equipment, and Measuring and Control Equipment Fama and French (1997) 48 industries; and financially distressed firms are those with Z-scores below 1.8 (*Z-Score*; Altman 1968). The results indicate that *VP Rank*'s predictive power is most pronounced for loss-making firms and technology firms; the effect is 0.6% ($t = 1.83$) and 1.5% ($t = 2.18$) larger for valuing loss-making and technology firms, respectively.

To summarize, we find that our results based on hedge portfolio regressions are robust to the

use of panel data methods. *VP Rank* remains a significant predictor of returns even after we control for several other factors documented in the accounting and finance literatures. *VP Rank*'s predictive ability for returns is pervasive throughout the sample but most pronounced for loss-making and technology firms.

5. Conclusion

We estimate the structural valuation framework proposed but not fully evaluated by Nissim and Penman (NP, 2001). NP acknowledge that the estimation methods available to them were insufficient for the non-linear structure of their model; we address this issue by applying Deep Learning, a well-known machine learning algorithm that can capture arbitrarily complex relationships across variables to estimate the non-linear information dynamics in NP's model and forecast firms' future earnings to arrive at an estimate of a firm's value. We also explore the effects of variation in several financial statement analysis design choices that analysts must make on empirical rather than theoretical grounds. We take the perspective of an equity investor who wishes to maximize risk-adjusted returns, and we test how different financial statement analysis design choices within NP's structural framework affect the performance of value-to-price-based trading strategies. We find that more disaggregation, combined with focusing on operating activities and expanding the forecast horizon, improves model performance; we find little evidence that incorporating more historical information and focusing on core items improves performance. Consideration of non-linearities improves model performance throughout the sample, especially for small, loss-making, technology, and financially distressed firms.

Our findings are subject to several caveats. First, restricting predictors to those theoretically derived by NP might exclude valuation-useful information. For example, Easton et al. (2020) show that Hou et al.'s (2012) empirically derived profitability predictors have incremental predictive power over a random walk when using the LAD estimator. Second, while our assumptions of constant

discount, growth, and reinvestment rates facilitate the attribution of model performance to fundamentals, models might perform even better after each of these rates is carefully calibrated. Third, we test model performance for firms with at least five years of lagged data for all predictors. If information dynamics for firms with more-limited historical data take a different functional form, value estimates for these firms obtained using the algorithms presented in this paper might perform poorly. Fourth, restricting the estimation sample to 10 years (Hou et al. 2012) might not be optimal for DL algorithms; while traditional estimators' marginal performance decreases in sample size, DL's marginal performance decreases at a slower rate.³³ Fifth, we use a standard architecture of the neural network (five hidden layers with 500 neurons per layer and learning and regularization rates of 10^{-3} and 10^{-5}), but one could apply Bayesian Optimization algorithms to identify optimal architectures *ex ante*. Finally, given the absence of an agreed-upon asset pricing model, it is unclear whether the alpha magnitudes we report are due to risk or mispricing (Roll 1977).

³³ Ng, A. Why DL is Taking Off. <https://www.coursera.org/lecture/neural-networks-deep-learning/why-is-deep-learning-taking-off-praGm>. October 17, 2018.

Appendix A: Variable Definitions

Variable	Definition
ATO	Asset turnover: Sales/NOA
ATO*	Modified asset turnover: Sales/OA
Core NBC	Core net borrowing cost: Core NFE/NFO
Core NFE	Core net financial expense: Compustat: $(XINT - IDIT) \times (1 - MTR) + DVP$
Core OI from Sales	Core operating income from sales: OI from Sales – UOI
Core Other Items	Core other items: Other Items – UOI
Core RNOA	Core return on net operating assets: Core OI from Sales/NOA
Core Sales PM	Core sales profit margin: Core OI from Sales/Sales
Core Sales PM*	Modified core sales profit margin: $(\text{Core OI from Sales} + io)/\text{Sales}$
Core SPREAD	Core financial leverage spread: Core RNOA – Core NBC
CSE	Common equity: Compustat (average): CEQ + TSTKP – DVPA
FLEV	Financial leverage: NFO/CSE
io	Implicit interest charge on operating liabilities: $R_f \times (OL - TXDITC)$
MIB	Minority interest book value: Compustat (average): MIB
MII	Minority interest income: Compustat: MII
MSA	Marketable security adjustment: Compustat: MSA
MSR	Minority sharing ratio: $(CNI/(CNI + MII)) \times (CSE/(CSE + MIB))^{-1}$ (if CNI, MII, CSE, MIB ≥ 0 , otherwise set to 1)
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).
NBC	Net borrowing cost: NFE/NFO
NFE	Core net financial expense: Core NFE – Δ MSA
NFO	Net financial obligations: Compustat (average): $(DLC + DLTT + PSTK - TSTKP + DVPA) - (CHE + IVAO)$
NOA	Net operating assets: NFO + CSE + MIB
OA	Operating assets: Compustat (average) AT – CHE – IVAO
OI	Operating income: NFE + CNI + MII
OI from Sales	OI – Other Items
OL	Operating liabilities (average): OA – NOA
OLLEV	Operating liability leverage: OL/NOA
OLSPREAD	Operating leverage spread: $(OI + io)/OA - io/OL$
Other Items	Compustat: ESUB

P	Price at the end of the June in year following the fiscal year: Compustat: $PRCCF \times CSHO$
R_f	Risk-free rate: One-year Treasury bill yield during the year
RNOA	Return on net operating assets: OI/NOA
ROCE	Return on common equity: $NOA/CSE \times RNOA - NFO/CSE \times NBC$
ROTCE	Return on total common equity: $(CNI + MII)/(CSE + MIB)$
Sales	Compustat: SALE
Sales PM	Sales profit margin: $OI \text{ from Sales}/Sales$
Sales PM*	Modified sales profit margin: $(OI \text{ from Sales} + io)/Sales$
SPREAD	Financial leverage spread: $RNOA - NBC$
TXDITC	Deferred taxes and investment tax credit: Compustat: TXDITC
UFE	Unusual financial expense: Compustat: ΔMSA
UOI	Unusual operating income: Compustat: $(NOPI + SPI) \times (1 - MTR) - ESUB + XIDO + \Delta RECTA$

Δ denotes change over the fiscal year and (average) average over fiscal year beginning and ending values.

References

- Abarbanell, J. S., and B. J. Bushee. 1997. Fundamental Analysis, Future Earnings, and Stock Prices. *Journal of Accounting Research* 35 (1):1-24.
- Abarbanell, J. S., and B. J. Bushee. 1998. Abnormal Returns to a Fundamental Analysis Strategy. *The Accounting Review*:19-45.
- Altman, E. I. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance* 23 (4):589-609.
- Anand, V., R. Brunner, K. Ikegwa, and T. Sougiannis. 2020. Predicting Profitability Using Machine Learning. Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3466478.
- Armstrong, J. S. 2001. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Vol. 30: Springer Science & Business Media.
- Aubry, M., R. Kräussl, G. Manso, and C. Spaenjers. 2019. Machine Learning, Human Experts, and the Valuation of Real Assets. No. 635 Center for Financial Studies working paper. Available at <https://EconPapers.repec.org/RePEc:zbw:cfswo:635>.
- Avramov, D., G. Kaplanski, and A. Subrahmanyam. 2020. Post-Fundamentals Drift in Stock Prices: A Regression Regularization Perspective. Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3507512.
- Ball, R., G. Sadka, and R. Sadka. 2009. Aggregate Earnings and Asset Prices. *Journal of Accounting Research* 47 (5):1097-1133.
- Bao, Y., B. Ke, B. Li, Y. J. Yu, and J. Zhang. 2020. Detecting Accounting Fraud in Publicly Traded Us Firms Using a Machine Learning Approach. *Journal of Accounting Research* 58 (1):199-235.
- Barnea, A., J. Ronen, and S. Sadan. 1976. Classificatory Smoothing of Income with Extraordinary Items. *The Accounting Review* 51 (1):110-122.
- Barth, M. E., K. Li, and C. McClure. 2019. Evolution in Value Relevance of Accounting Information. Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2933197.
- Bengio, Y. 2009. Learning Deep Architectures for Ai. *Foundations and Trends in Machine Learning* 2 (1):1-127.
- Bentley, J. W., T. E. Christensen, K. H. Gee, and B. C. Whipple. 2018. Disentangling Managers' and Analysts' Non-GAAP Reporting. *Journal of Accounting Research* 56 (4):1039-1081.
- Bertomeu, J., E. Cheynel, E. Floyd, and W. Pan. 2019. Using Machine Learning to Detect Misstatements. *Review of Accounting Studies (Forthcoming)*.
- Black, F. 1986. Noise. *The Journal of Finance* 41 (3):528-543.
- 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.
- Bradshaw, M. T., T. E. Christensen, K. H. Gee, and B. C. Whipple. 2018. Analysts' GAAP Earnings Forecasts and Their Implications for Accounting Research. *Journal of Accounting and Economics*.
- Brown, N. C., R. M. Crowley, and W. B. Elliott. 2020. What Are You Saying? Using Topic to Detect Financial Misreporting. *Journal of Accounting Research* 58 (1):237-291.
- 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.
- Carhart, M. M. 1997. On Persistence in Mutual Fund Performance. *The Journal of Finance* 52 (1):57-82.
- Cybenko, G. 1989. Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals and Systems* 2 (4):303-314.
- Damodaran, A. 2007. Valuation Approaches and Metrics: A Survey of the Theory and Evidence. *Foundations and Trends in Finance* 1 (8):693-784.

- Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. *The Journal of Finance* 52 (3):1035-1058.
- Daniel, K., and S. Titman. 1997. Evidence on the Characteristics of Cross Sectional Variation in Stock Returns. *The Journal of Finance* 52 (1):1-33.
- Dechow, P. M., A. P. Hutton, and R. G. Sloan. 1999. An Empirical Assessment of the Residual Income Valuation Model. *Journal of Accounting and Economics* 26 (1):1-34.
- Dichev, I. D. 2020. 50 Years of Capital Markets Research in Accounting: Achievements So Far and Opportunities Ahead. Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3633489.
- Ding, K., B. Lev, X. Peng, T. Sun, and M. A. Vasarhelyi. 2020. Machine Learning Improves Accounting Estimates: Evidence from Insurance Payments. *Review of Accounting Studies*.
- Doyle, J. T., R. J. Lundholm, and M. T. Soliman. 2003. The Predictive Value of Expenses Excluded from Pro Forma Earnings. *Review of Accounting Studies* 8 (2-3):145-174.
- Dye, R. A. 2002. Classifications Manipulation and Nash Accounting Standards. *Journal of Accounting Research* 40 (4):1125-1162.
- Easton, P. D., M. Kapons, P. Kelly, and A. Neuhierl. 2020. Beating a Random Walk. Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3040354.
- Edwards, E., and P. Bell. 1961. *The Theory and Management of Business Income*: University of California Press, Berkeley.
- Engle, R., R. Ferstenberg, and J. Russell. 2012. Measuring and Modeling Execution Cost and Risk. *Journal of Portfolio Management* 38:14-28.
- Erel, I., L. H. Stern, C. Tan, and M. S. Weisbach 2018. Selecting Directors Using Machine Learning. NBER working paper 24435.
- 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.
- Evans, M. E., K. Njoroge, and K. O. Yong. 2017. An Examination of the Statistical Significance and Economic Relevance of Profitability and Earnings Forecasts from Models and Analysts. *Contemporary Accounting Research*.
- Fairfield, P. M., S. Ramnath, and T. L. Yohn. 2009. Do Industry - Level Analyses Improve Forecasts of Financial Performance? *Journal of Accounting Research* 47 (1):147-178.
- Fairfield, P. M., R. J. Sweeney, and T. L. Yohn. 1996. Accounting Classification and the Predictive Content of Earnings. *The Accounting Review* 71 (3):337-355.
- 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.
- Fama, E. F., and K. R. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33 (1):3-56.
- . 1997. Industry Costs of Equity. *Journal of Financial Economics* 43 (2):153-193.
- Fama, E. F., and K. R. French. 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116 (1):1-22.
- Francis, J., P. Olsson, and D. R. Oswald. 2000. Comparing the Accuracy and Explainability of Dividend, Free Cash Flow, and Abnormal Earnings Equity Value Estimates. *Journal of Accounting Research* 38 (1):45-70.
- Frankel, R., J. Jennings, and J. Lee. 2016. Using Unstructured and Qualitative Disclosures to Explain Accruals. *Journal of Accounting and Economics* 62 (2-3):209-227.
- Frankel, R., and C. M. C. Lee. 1998. Accounting Valuation, Market Expectation, and Cross-Sectional Stock Returns. *Journal of Accounting and Economics* 25 (3):283-319.

- Frankel, R. M., J. N. Jennings, and J. A. Lee. 2020. Using Natural Language Processing to Assess Text Usefulness to Readers: The Case of Conference Calls and Earnings Prediction. Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3095754.
- Friedman, J., T. Hastie, and R. Tibshirani. 2001. *The Elements of Statistical Learning*. Vol. 1: Springer series in statistics New York, NY, USA.
- Fu, X., J. Du, Y. Guo, M. Liu, T. Dong, and X. Duan. 2018. A Machine Learning Framework for Stock Selection. Working paper. Available at <https://arxiv.org/pdf/1806.01743.pdf>.
- Givoly, D., and C. Hayn. 2000. The Changing Time-Series Properties of Earnings, Cash Flows and Accruals: Has Financial Reporting Become More Conservative? *Journal of Accounting and Economics* 29 (3):287-320.
- 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.
- Glorot, X., A. Bordes, and Y. Bengio. 2011. Deep Sparse Rectifier Neural Networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 315-323.
- Graham, J. R. 2000. How Big Are the Tax Benefits of Debt? *The Journal of Finance* 55 (5):1901-1941.
- Gu, S., B. T. Kelly, and D. Xiu. 2020. Empirical Asset Pricing Via Machine Learning. *Review of Financial Studies* 33 (5):2223-2273.
- Hagan, M. T., H. B. Demuth, M. H. Beale, and O. De Jesús. 1996. *Neural Network Design*. Vol. 20: Pws Pub. Boston.
- Hayn, C. 1995. The Information Content of Losses. *Journal of Accounting and Economics* 20 (2):125-153.
- He, K., X. Zhang, S. Ren, and J. Sun. 2016. Identity Mappings in Deep Residual Networks. In *European conference on computer vision*, 630-645.
- Hirshleifer, D., K. Hou, and S. H. Teoh. 2012. The Accrual Anomaly: Risk or Mispricing? *Management Science* 58 (2):320-335.
- 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.
- Huang, L., Y. Jin, Y. Gao, K.-H. Thung, and D. Shen. 2016. Longitudinal Clinical Score Prediction in Alzheimer's Disease with Soft-Split Sparse Regression Based Random Forest. *Neurobiology of Aging* 46:180-191.
- Huber, P. J. 1973. Robust Regression: Asymptotics, Conjectures and Monte Carlo. *The Annals of Statistics* 1 (5):799-821.
- Huerta, R., F. Corbacho, and C. Elkan. 2013. Nonlinear Support Vector Machines Can Systematically Identify Stocks with High and Low Future Returns. *Algorithmic Finance* 2 (1):45-58.
- 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>.
- Jackwerth, J. C., and A. Slavutskaya. 2018. Relative Alpha. Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2439145.
- Jegadeesh, N., and S. Titman. 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance* 48 (1):65-91.
- Jensen, M. C., F. Black, and M. S. Scholes. 1972. *The Capital Asset Pricing Model: Some Empirical Tests*: Praeger Publishers, Inc.
- Jensen, M. C., and W. H. Meckling. 1976. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics* 3 (4):305-360.

- 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.
- Lee, C. M. C., J. Myers, and B. Swaminathan. 1999. What Is the Intrinsic Value of the Dow? *The Journal of Finance* 54 (5):1693-1741.
- Lee, C. M. C., and E. C. So. 2015. Alphanomics: The Informational Underpinnings of Market Efficiency. *Foundations and Trends in Accounting* 9 (2--3):59-258.
- Li, N., S. Richardson, and I. Tuna. 2014. Macro to Micro: Country Exposures, Firm Fundamentals and Stock Returns. *Journal of Accounting and Economics* 58 (1):1-20.
- Loughran, T., and J. R. Ritter. 2000. Uniformly Least Powerful Tests of Market Efficiency. *Journal of Financial Economics* 55 (3):361-389.
- Malkiel, B. G., and E. F. Fama. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance* 25 (2):383-417.
- McKeon, J. 2018. Long-Term Trends in Non-GAAP Disclosures: A Three-Year Overview: Audit Analytics.
- McVay, S. E. 2006. Earnings Management Using Classification Shifting: An Examination of Core Earnings and Special Items. *The Accounting Review* 81 (3):501-531.
- Miller, M. H., and F. Modigliani. 1961. Dividend Policy, Growth, and the Valuation of Shares. *The Journal of Business* 34 (4):411-433.
- Monahan, S. J. 2018. Financial Statement Analysis and Earnings Forecasting. *Foundations and Trends in Accounting* 12 (2):105-215.
- Myers, S. C. 1984. The Capital Structure Puzzle. *The Journal of Finance* 39 (3):574-592.
- Newey, W. K., and K. D. West. 1986. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation-Consistent Covariance Matrix: National Bureau of Economic Research Cambridge, Mass., USA.
- Nissim, D. 2019. Terminal Value. Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3095564.
- Nissim, D., and S. H. Penman. 2001. Ratio Analysis and Equity Valuation: From Research to Practice. *Review of Accounting Studies* 6 (1):109-154.
- Ohlson, J. A. 1995. Earnings, Book Values, and Dividends in Equity Valuation. *Contemporary Accounting Research* 11 (2):661-687.
- Ou, J. A., and S. H. Penman. 1989a. Accounting Measurement, Price-Earnings Ratio, and the Information Content of Security Prices. *Journal of Accounting Research* 27 (Supplement):111-144.
- . 1989b. Financial Statement Analysis and the Prediction of Stock Returns. *Journal of Accounting and Economics* 11 (4):295-329.
- 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., F. Reggiani, S. A. Richardson, and I. Tuna. 2017. A Framework for Identifying Accounting Characteristics for Asset Pricing Models, with an Evaluation of Book-to-Price. *European Financial Management* 24 (4):488-520.
- Penman, S. H., and T. Sougiannis. 1998. A Comparison of Dividend, Cash Flow, and Earnings Approaches to Equity Valuation. *Contemporary Accounting Research* 15 (3):343-383.
- Penman, S. H., and X.-J. Zhang. 2002. Accounting Conservatism, the Quality of Earnings, and Stock Returns. *The Accounting Review* 77 (2):237-264.
- Piotroski, J. D. 2000. Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers. *Journal of Accounting Research* 38 (Supplement):1-41.

- Preinreich, G. A. D. 1938. Annual Survey of Economic Theory: The Theory of Depreciation. *Econometrica: Journal of the Econometric Society*:219-241.
- Purda, L., and D. Skillicorn. 2015. Accounting Variables, Deception, and a Bag of Words: Assessing the Tools of Fraud Detection. *Contemporary Accounting Research* 32 (3):1193-1223.
- Revsine, L., and D. Collins. 2005. *Financial Reporting & Analysis*: Upper Saddle River, NJ: Prentice Hall.
- Richardson, S. A., R. G. Sloan, M. T. Soliman, and I. Tuna. 2005. Accrual Reliability, Earnings Persistence and Stock Prices. *Journal of Accounting and Economics* 39 (3):437-485.
- Roll, R. 1977. A Critique of the Asset Pricing Theory's Tests Part I: On Past and Potential Testability of the Theory. *Journal of Financial Economics* 4 (2):129-176.
- Schmidhuber, J. 2015. Deep Learning in Neural Networks: An Overview. *Neural Networks* 61:85-117.
- Schularick, M., and A. M. Taylor. 2012. Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008. *The American Economic Review* 102 (2):1029-1061.
- 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.
- 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. 2020. Man Versus Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases. NBER working paper 27843.
- Watts, R. L., and R. W. Leftwich. 1977. The Time Series of Annual Accounting Earnings. *Journal of Accounting Research* 15 (2):253-271.
- White, H. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica: Journal of the Econometric Society*:817-838.
- Yohn, T. L. 2020. Research on the Use of Financial Statement Information for Forecasting Profitability. *Accounting and Finance* 60: 3163-3181.

Figure 1. Nissim and Penman (2001) Analysis of ROCE

This figure depicts the analysis of profitability developed in Nissim and Penman (2001). All variables are defined in Appendix A.

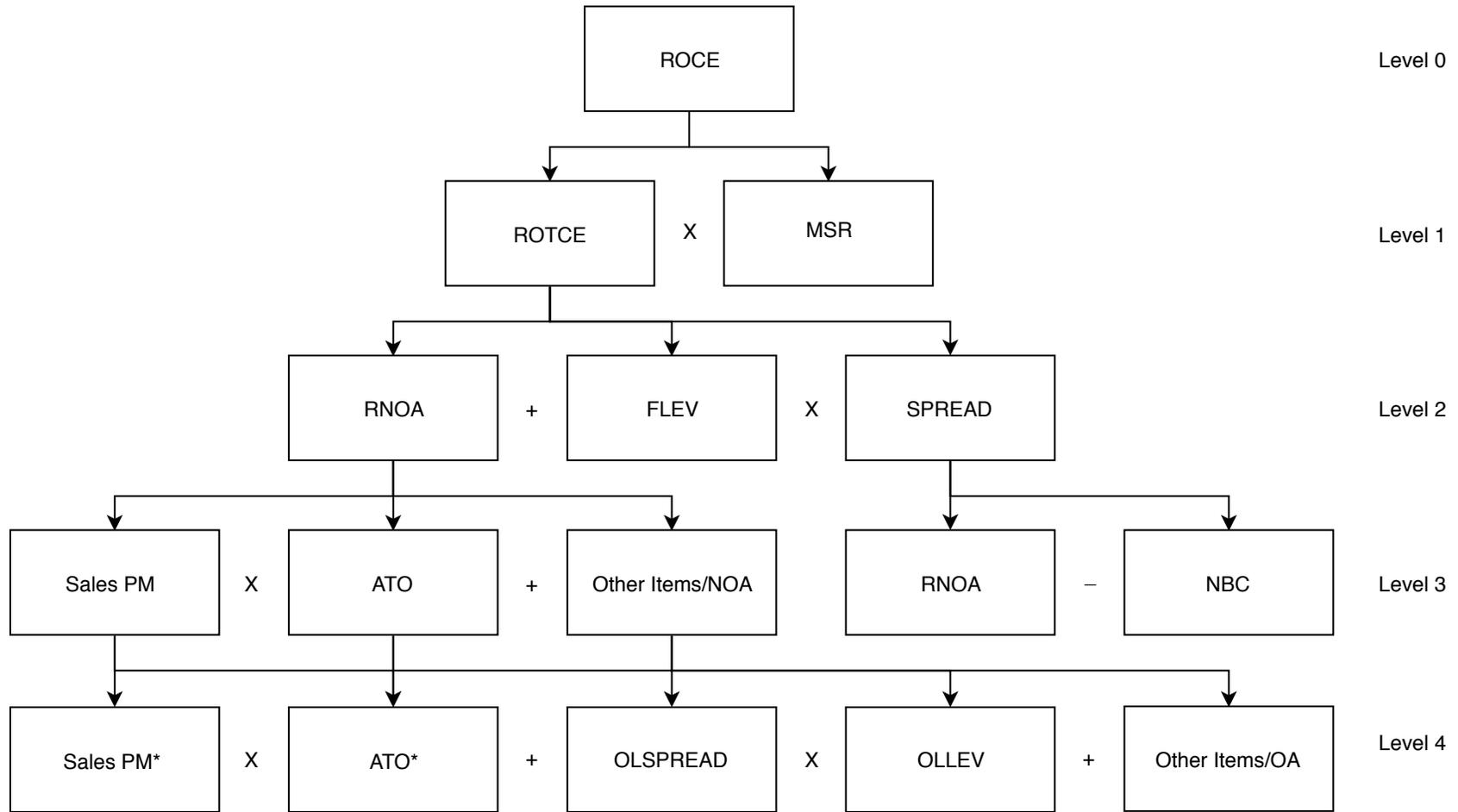
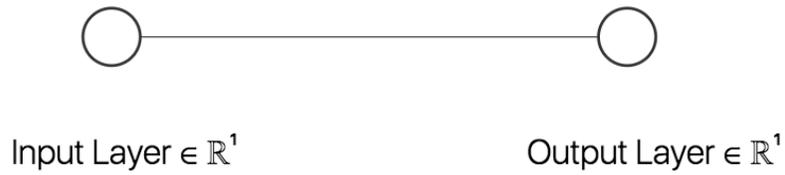


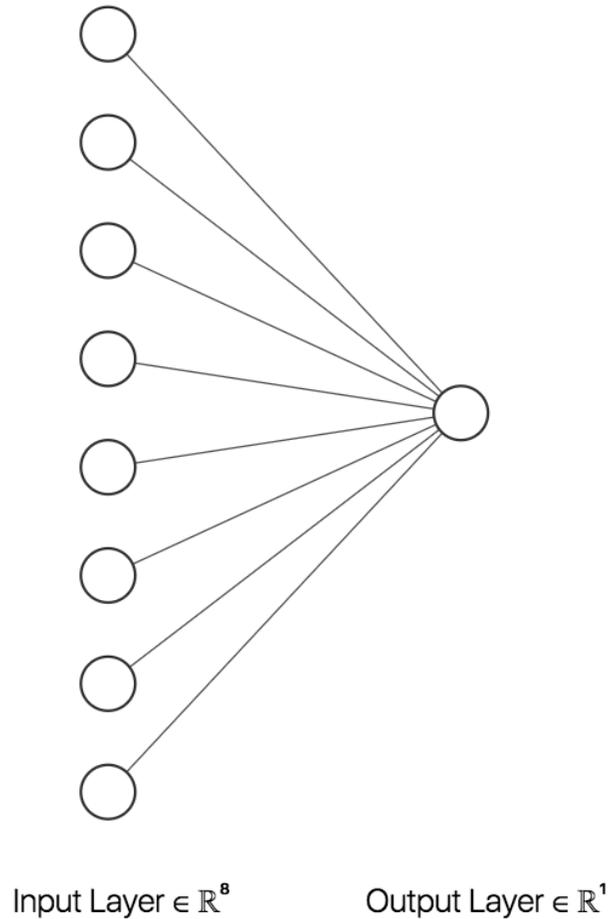
Figure 2. Estimator Comparison

This figure's Panels A, B, and C graphically present the architecture of a single-layered neural network with a single predictor, a single-layered neural network with eight predictors, a multi-layered neural network (a fully connected sequential multi-layered neural network with one input layer with eight neurons, two hidden layers with 10 neurons each, and one output layer with 5 neurons).

Panel A. Single-layered Neural Network with a Single Predictor



Panel B. Single-layered Neural Network with Multiple Predictors



Panel C. Multi-layered Neural Network

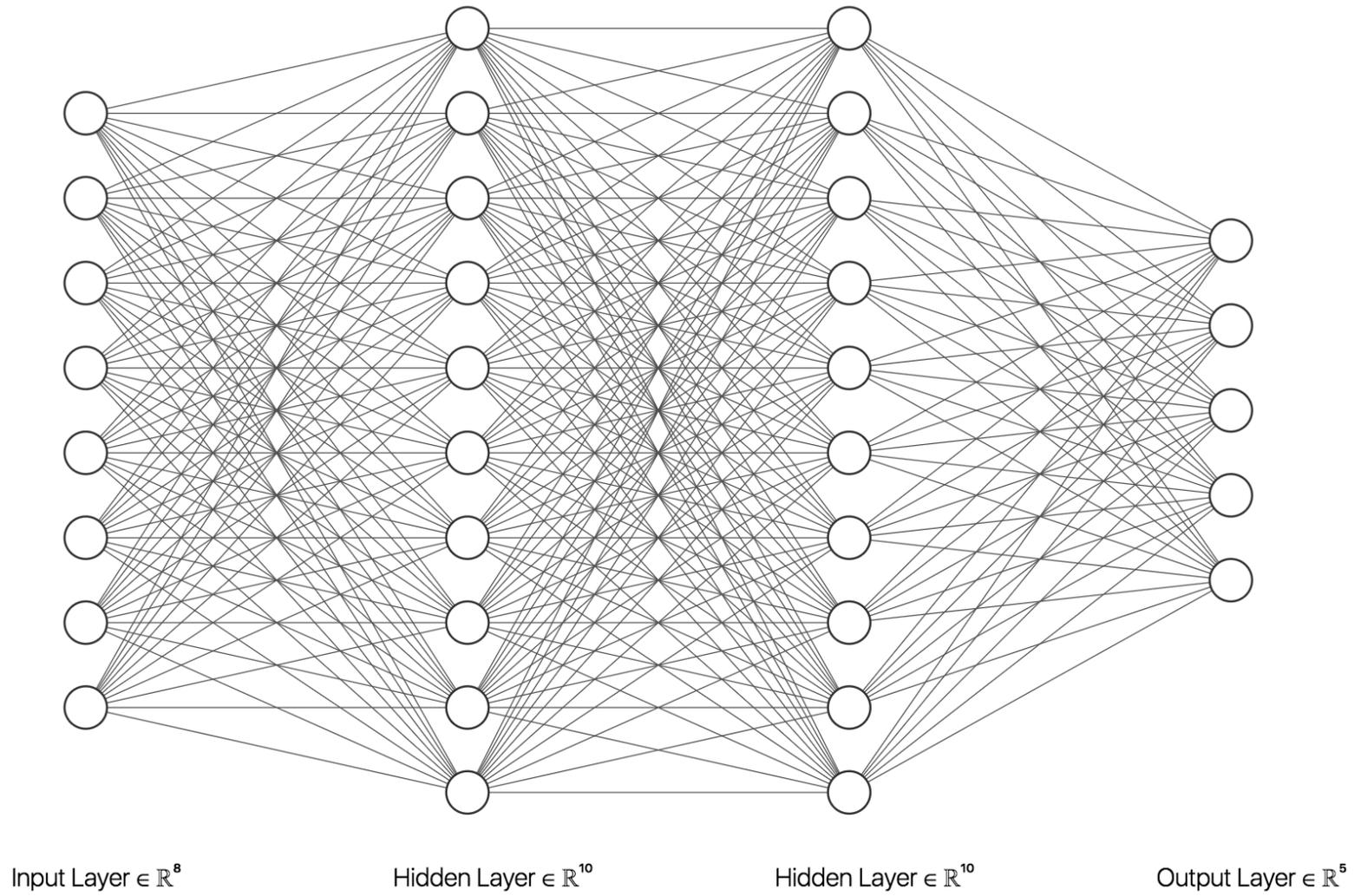


Figure 3. Prominence of Different Machine Learning Algorithms.

This figure depicts the prominence of different machine learning algorithms as measured by web queries analyzed via the web tool Google Trends over the 2004 to 2020 period.

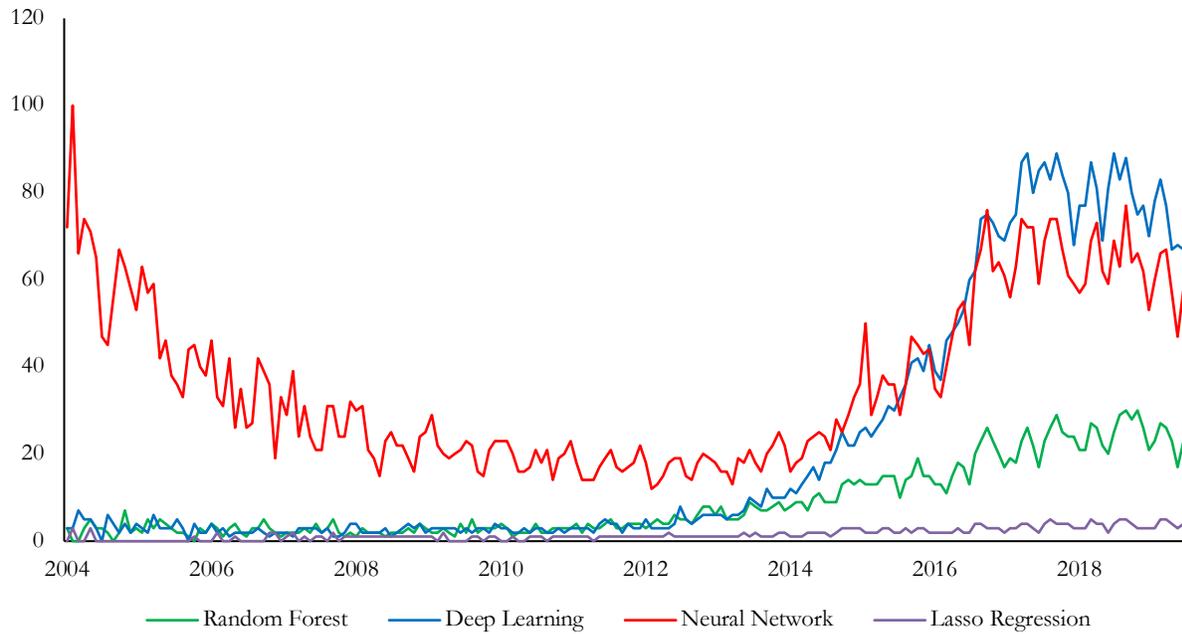


Figure 4. Model Estimation and Forecasting Timeline

This figure depicts the model estimation and forecasting timeline. The sample starts in 1963.

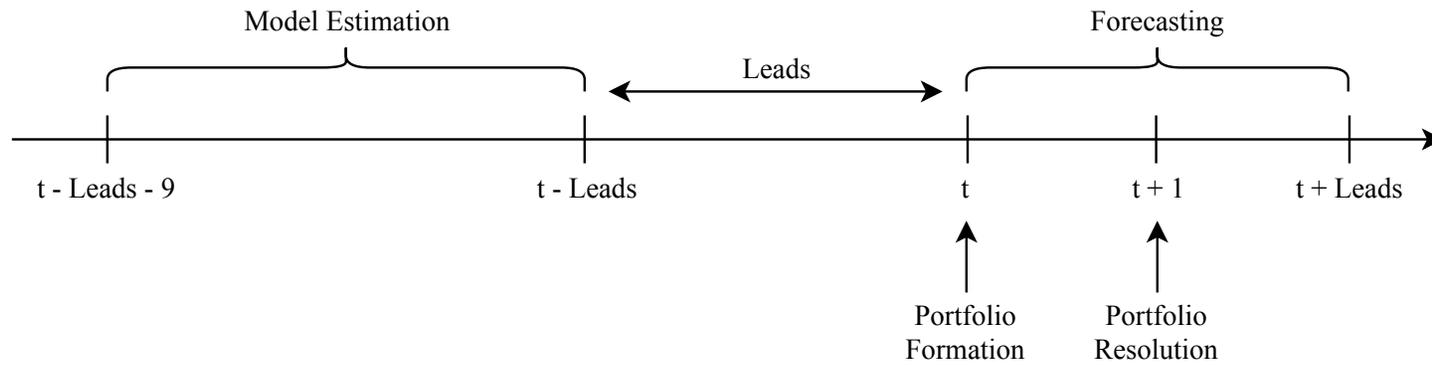
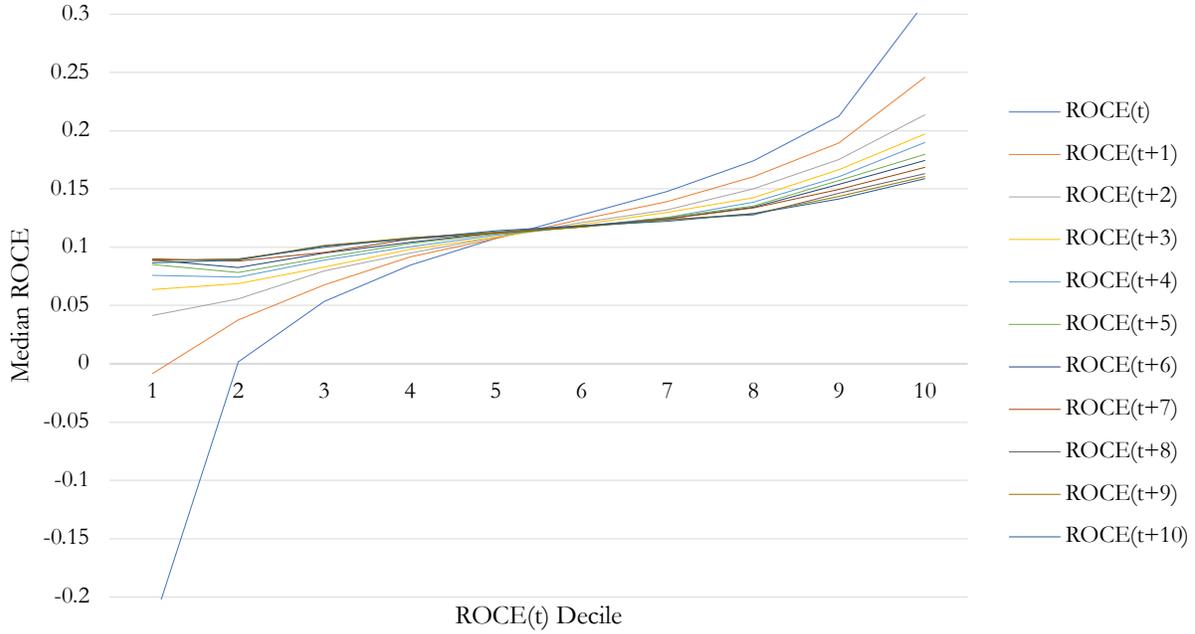


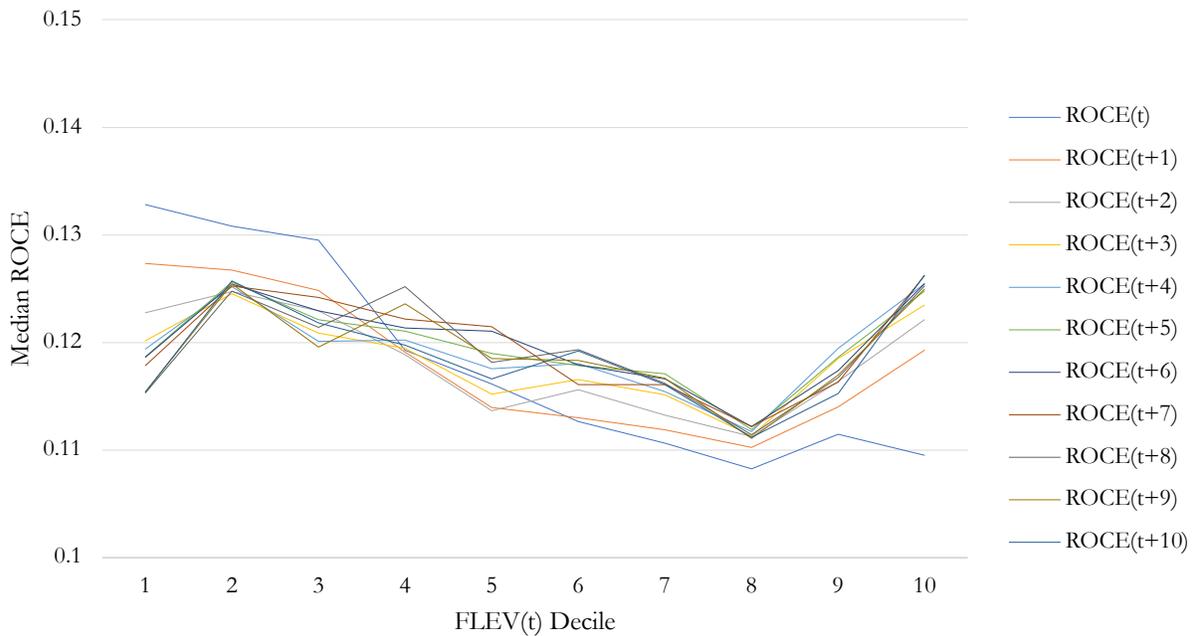
Figure 5. Univariate Time-Series Plots

This figure's Panels A to H present median portfolio ROCE in periods t to t+10 by ROCE, FLEV, SPREAD, ATO, Sales PM, OLLEV, OLSPREAD, and RNOA decile in period t. All variables are defined in Appendix A.

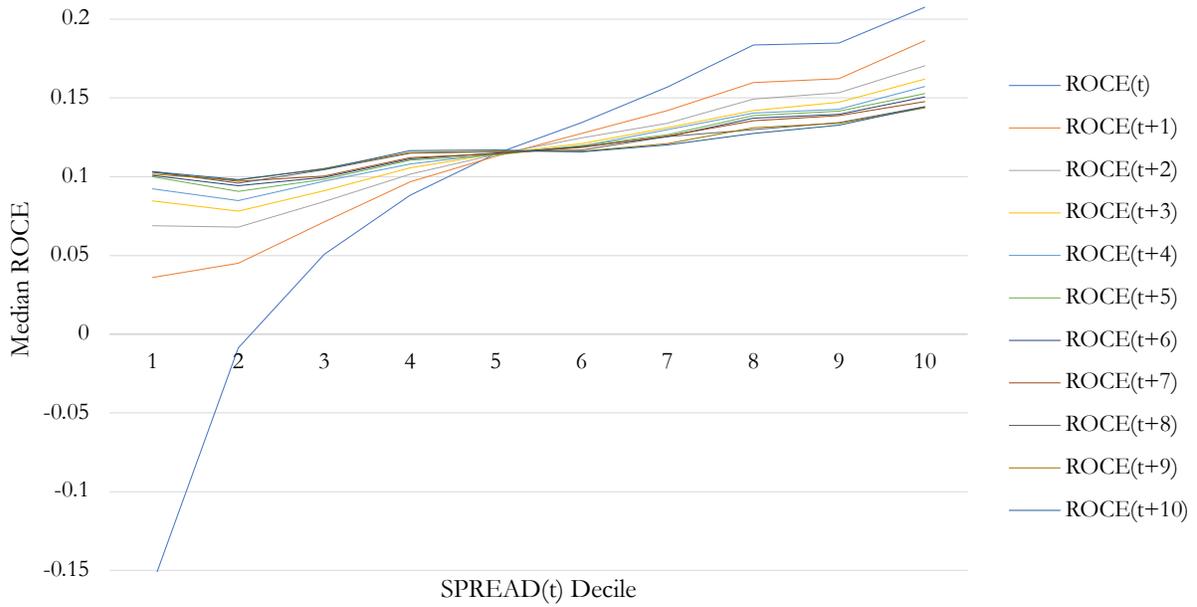
Panel A. Future ROCE by Current ROCE Decile



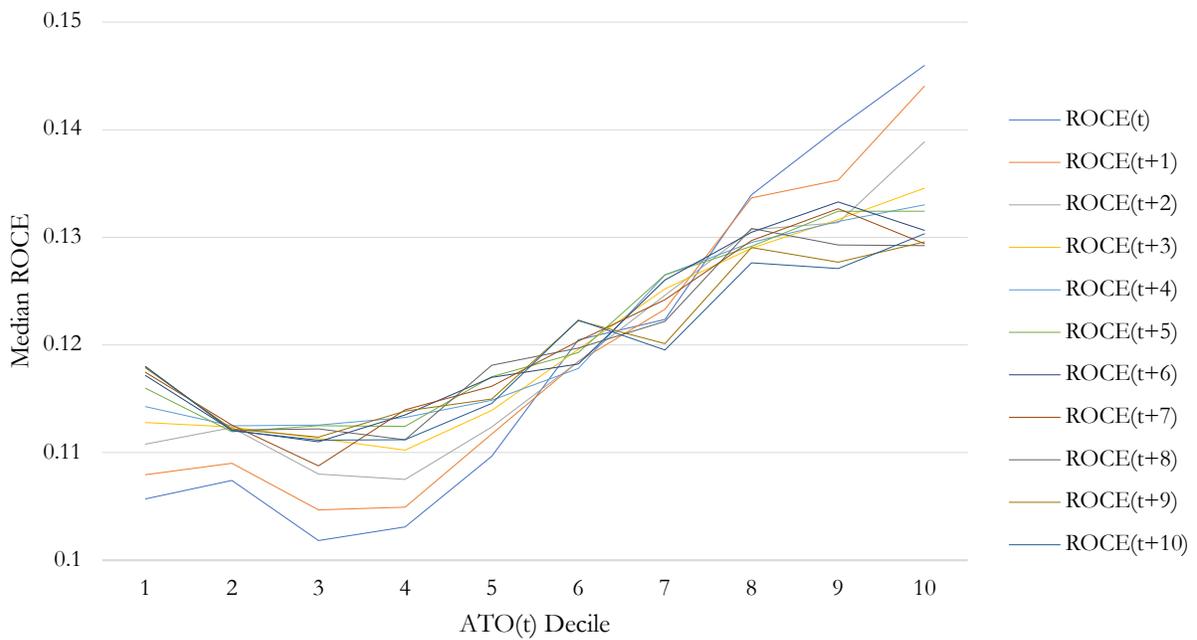
Panel B. Future ROCE by Current FLEV Decile



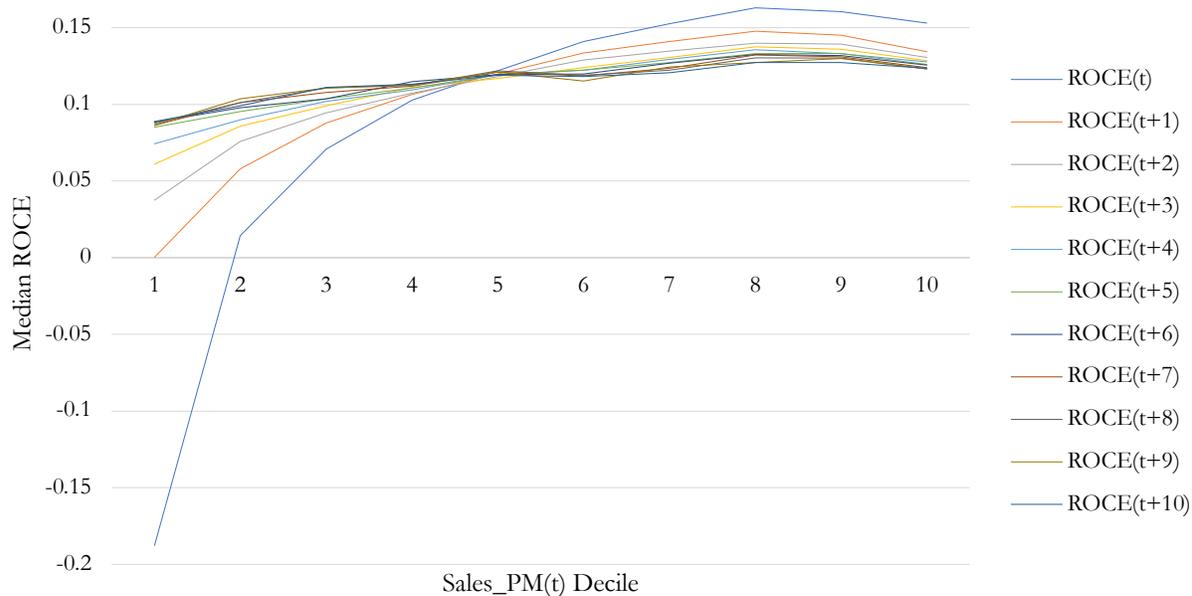
Panel C. Future ROCE by Current SPREAD Decile



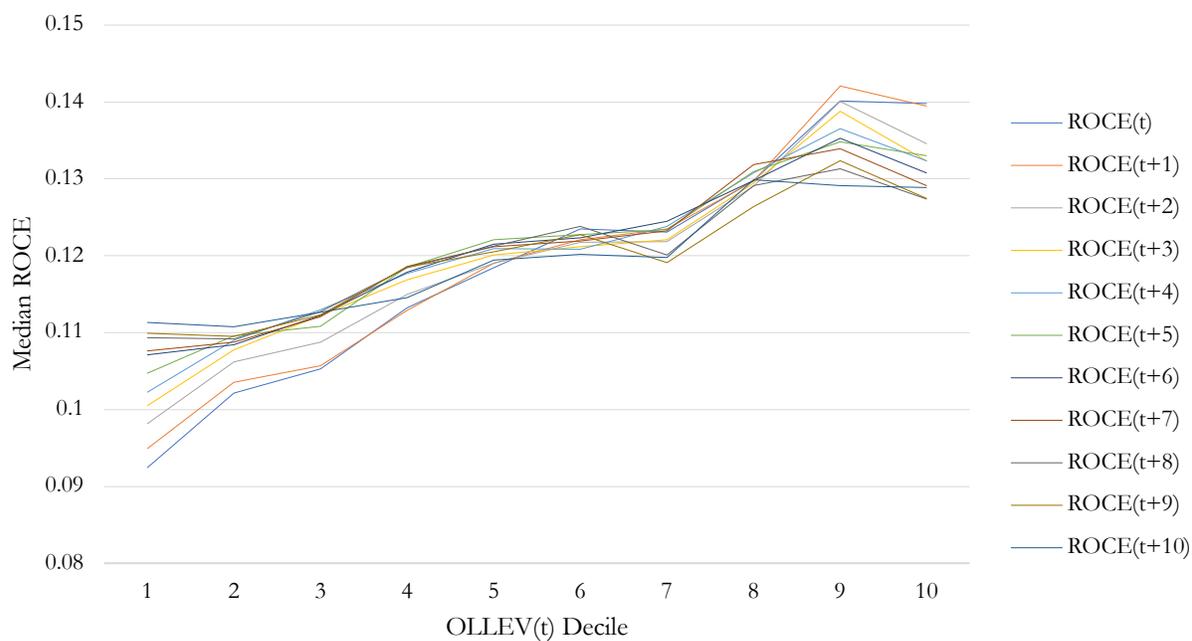
Panel D. Future ROCE by Current ATO Decile



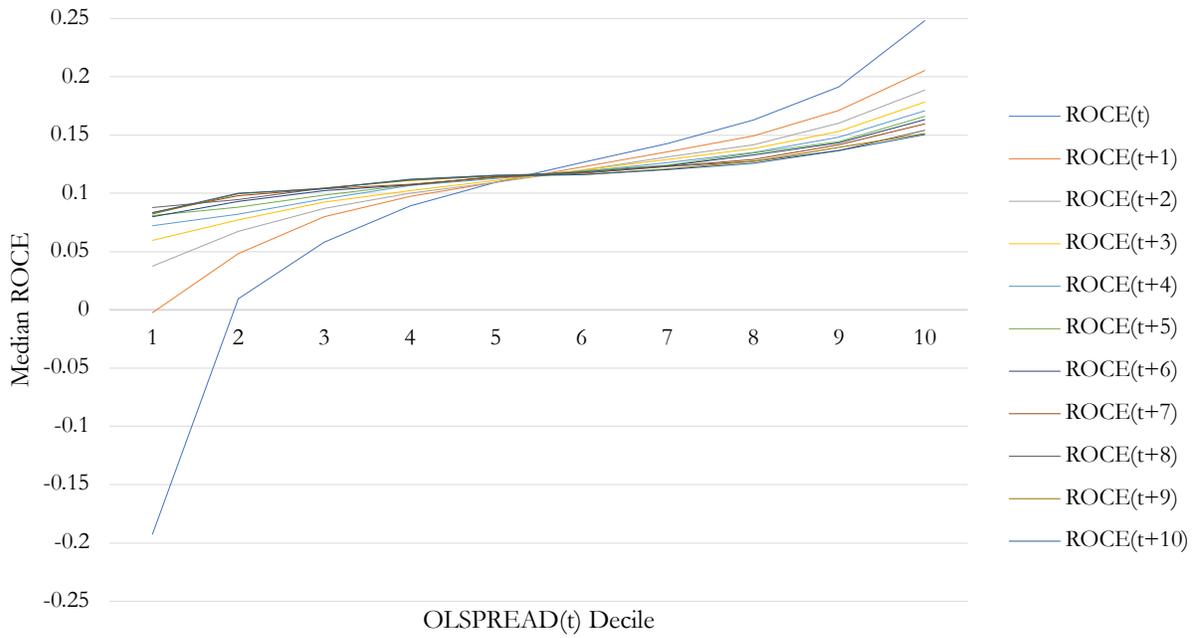
Panel E. Future ROCE by Current Sales PM Decile



Panel F. Future ROCE by Current OLLEV Decile



Panel G. Future ROCE by Current OLSPREAD Decile



Panel H. Future ROCE by Current RNOA Decile

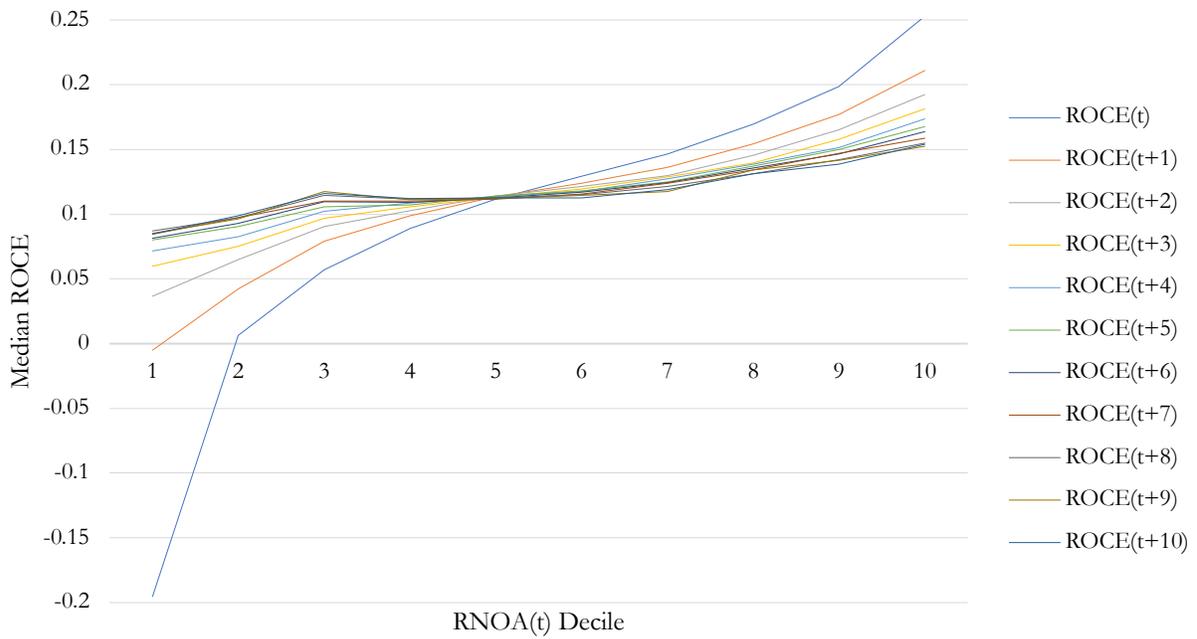
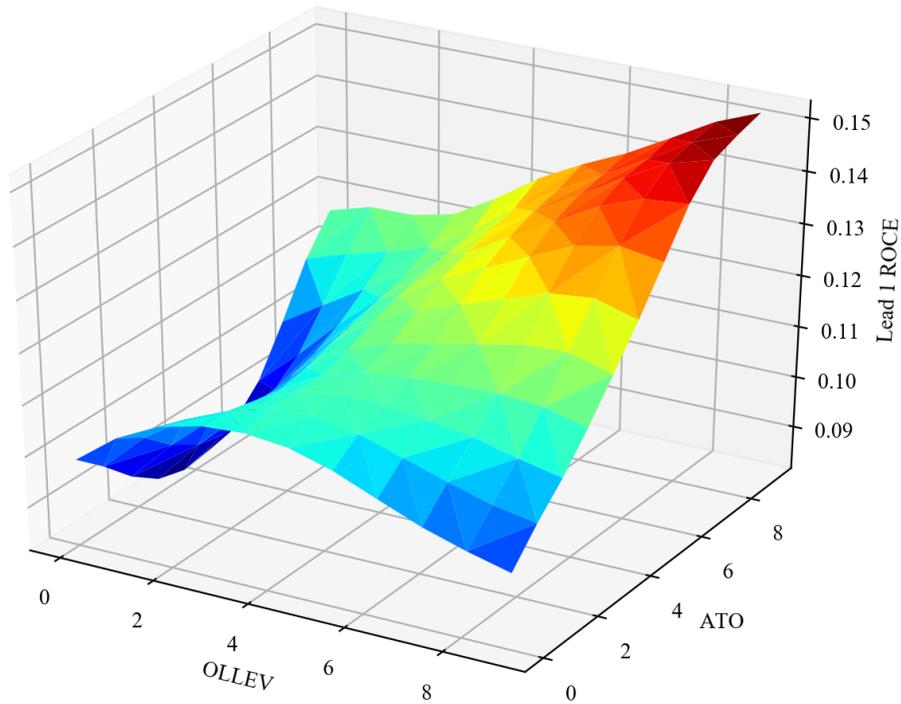


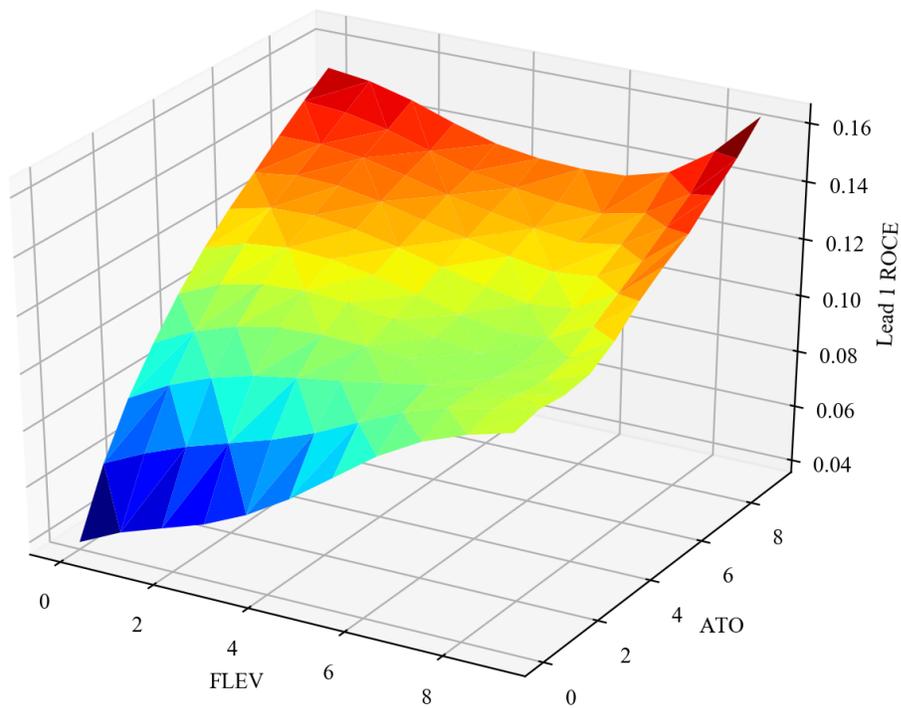
Figure 6. Illustrating Interactive Relationships across Variables in ROCE Prediction

This figure's Panels A to D illustrate examples of non-linear, interactive relationships across various ratios and future ROCE. All variables are defined in Appendix A.

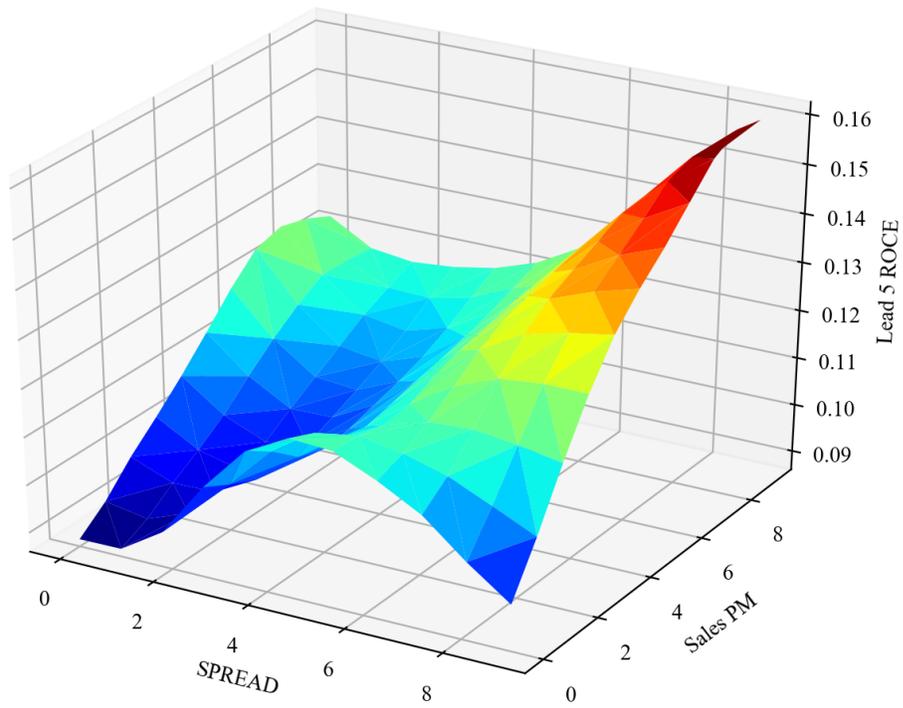
Panel A. OLLEV, ATO, and Lead 1 ROCE



Panel B. FLEV, ATO, and Lead 1 ROCE



Panel C. SPREAD, Sales PM, and Lead 5 ROCE



Panel D. OLSPREAD, Sales PM, and Lead 5 ROCE

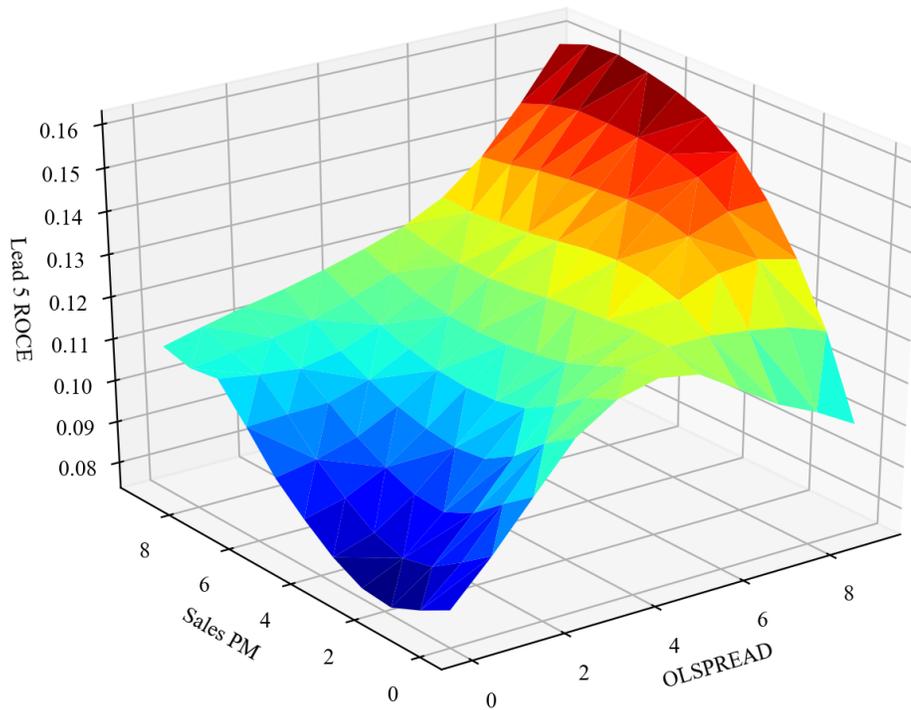
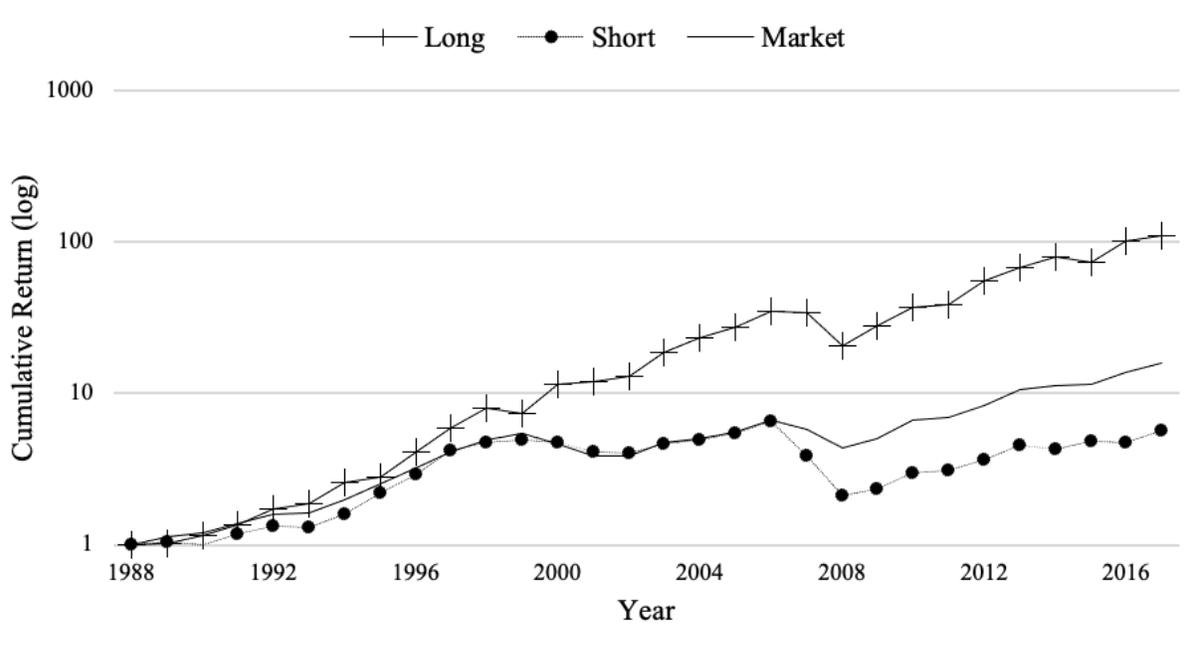


Figure 7. Cumulative Portfolio Long and Short Position Returns

This figure compares the performance of models based on Deep Learning profitability forecasts to that of models based on OLS and LAD forecasts. Panels A and B plot the log of cumulative returns for the long and short portfolios formed based on value-price ratios in the preceding year estimated via Deep Learning and OLS and cumulative value-weighted CRSP market return. Value price ratios are based on market prices at the end of June and value estimates from a residual income model based on forecasts for 5 RNOA leads obtained from Deep Learning and OLS models that do not focus on core items and incorporate information only from the most recent financial statements.

Panel A. Deep Learning



Panel B. OLS

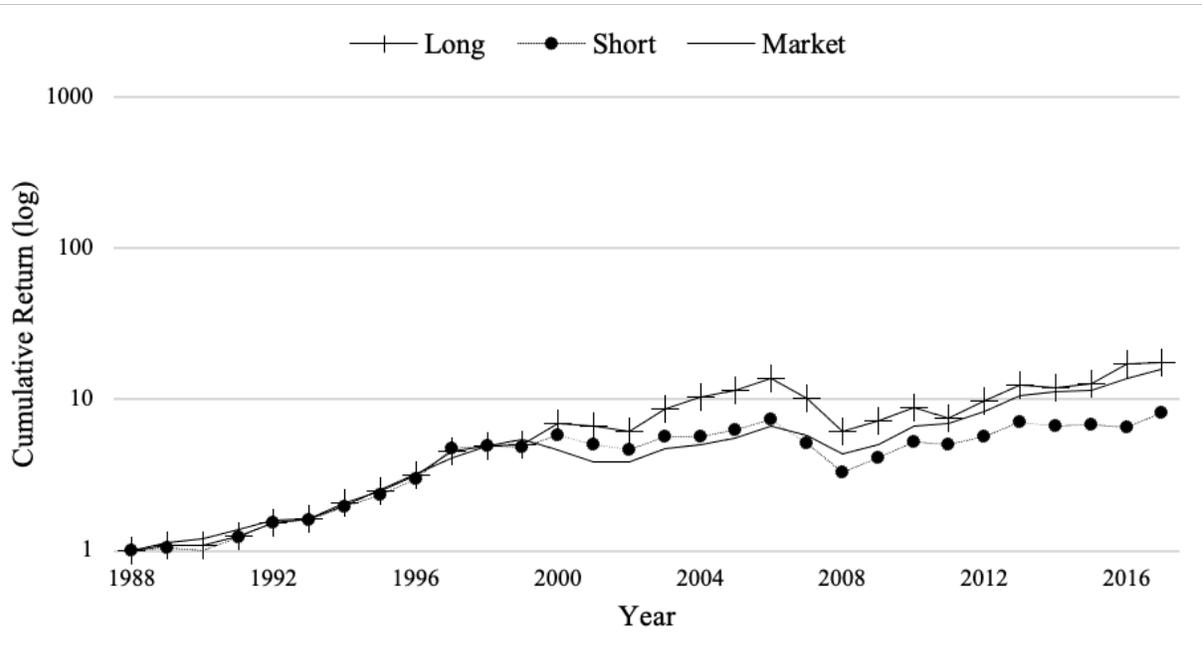


Table 1. Ratio Descriptive Statistics

This table's Panels A and B present descriptive statistics for valuation anchors and ratios. All ratios are winsorized at the 1st and 99th percentiles. All variables are defined in Appendix A.

Panel A. Anchors

Variable	N	Mean	StD	P1	P25	Median	P75	P99
P	55,972	4,093	17,305	3	65	394	1,948	68,935
aCSE	55,972	1,569	6,836	4	55	248	962	21,354
aNOA	55,972	2,939	17,195	4	76	366	1,510	37,994
aNFO	55,972	1,343	14,130	-1,138	4	64	477	16,110

Panel B. Ratios

Variable	N	Mean	StD	P1	P25	Median	P75	P99
ROCE	55,972	0.09	0.20	-0.83	0.05	0.12	0.17	0.58
MSR	55,972	1.00	0.03	0.83	1.00	1.00	1.00	1.08
ROTCE	55,972	0.09	0.21	-0.81	0.05	0.12	0.17	0.56
RNOA	55,972	0.10	0.15	-0.44	0.05	0.09	0.15	0.63
FLEV	55,972	0.69	1.14	-0.68	0.07	0.44	0.96	6.54
SPREAD	55,972	0.05	0.25	-0.84	-0.01	0.04	0.10	1.06
Sales PM	55,972	0.06	0.11	-0.36	0.02	0.05	0.10	0.35
ATO	55,972	2.35	2.27	0.20	1.11	1.87	2.79	12.17
Other Items/NOA	55,972	0.00	0.01	-0.01	0.00	0.00	0.00	0.06
NBC	55,972	0.05	0.12	-0.56	0.03	0.05	0.07	0.64
Sales PM*	55,972	0.07	0.12	-0.35	0.03	0.06	0.11	0.45
ATO*	55,972	1.45	0.97	0.10	0.79	1.31	1.85	5.52
OLLEV	55,972	0.58	0.94	0.08	0.27	0.40	0.60	4.04
OLSPREAD	55,972	0.02	0.09	-0.35	-0.01	0.03	0.06	0.28
Other Items/OA	55,972	0.00	0.01	-0.01	0.00	0.00	0.00	0.04
Core Sales PM	55,972	0.06	0.10	-0.28	0.03	0.05	0.09	0.31
Core Other Items/NOA	55,972	0.00	0.06	-0.20	-0.01	0.00	0.01	0.26
Core RNOA	55,972	0.10	0.13	-0.32	0.05	0.09	0.14	0.55
Core NBC	55,972	0.05	0.12	-0.53	0.03	0.05	0.07	0.61
Core SPREAD	55,972	0.05	0.23	-0.74	-0.01	0.03	0.09	0.95
Core Sales PM*	55,972	0.07	0.10	-0.27	0.04	0.06	0.10	0.40
Core Other Items/OA	55,972	0.00	0.04	-0.11	-0.01	0.00	0.01	0.15
UOI/NOA	55,972	0.00	0.06	-0.24	-0.01	0.00	0.01	0.20
UFE/NFO	55,972	0.00	0.01	-0.04	0.00	0.00	0.00	0.04

Table 2. Ratio Correlation Matrix

This table presents correlations among selected fundamental ratios. All ratios are winsorized at the 1st and 99th percentiles. Pearson (Spearman) correlations are above (below) the diagonal. * indicates statistical significance at the 5% level. All variables are defined in Appendix A.

Variable	ROCE	ROTCE	RNOA	FLEV	SPREAD	Sales PM	ATO	NBC	OLLEV	OLSPREAD	Core Sales PM	Core RNOA
ROCE	1.00	0.99*	0.78*	-0.14*	0.52*	0.65*	0.09*	-0.03*	0.03*	0.80*	0.55*	0.67*
ROTCE	0.99*	1.00	0.77*	-0.11*	0.52*	0.64*	0.10*	-0.03*	0.04*	0.79*	0.53*	0.66*
RNOA	0.90*	0.90*	1.00	-0.20*	0.69*	0.62*	0.24*	-0.06*	0.17*	0.90*	0.53*	0.89*
FLEV	-0.11*	-0.11*	-0.38*	1.00	-0.16*	0.08*	-0.25*	0.06*	-0.13*	-0.16*	0.10*	-0.20*
SPREAD	0.75*	0.75*	0.80*	-0.35*	1.00	0.42*	0.18*	-0.71*	0.15*	0.62*	0.37*	0.62*
Sales PM	0.65*	0.65*	0.60*	0.04*	0.53*	1.00	-0.13*	-0.02*	0.00	0.71*	0.88*	0.53*
ATO	0.19*	0.19*	0.33*	-0.42*	0.22*	-0.41*	1.00	-0.03*	0.38*	0.10*	-0.16*	0.26*
NBC	-0.07*	-0.07*	-0.08*	0.25*	-0.52*	-0.12*	0.03*	1.00	-0.06*	-0.06*	-0.03*	-0.06*
OLLEV	0.17*	0.17*	0.25*	-0.24*	0.17*	-0.12*	0.49*	-0.01*	1.00	0.04*	0.05*	0.22*
OLSPREAD	0.79*	0.80*	0.85*	-0.30*	0.74*	0.67*	0.12*	-0.18*	0.17*	1.00	0.62*	0.80*
Core Sales PM	0.56*	0.56*	0.49*	0.06*	0.45*	0.89*	-0.45*	-0.13*	-0.11*	0.60*	1.00	0.59*
Core RNOA	0.79*	0.80*	0.89*	-0.37*	0.71*	0.50*	0.35*	-0.09*	0.28*	0.77*	0.57*	1.00
Core NBC	-0.07*	-0.07*	-0.08*	0.26*	-0.51*	-0.12*	0.03*	0.97*	-0.02*	-0.18*	-0.14*	-0.09*
Core SPREAD	0.65*	0.65*	0.69*	-0.34*	0.90*	0.45*	0.23*	-0.55*	0.20*	0.66*	0.50*	0.78*
UOI/NOA	0.33*	0.33*	0.37*	-0.13*	0.29*	0.31*	0.06*	-0.01	-0.02*	0.28*	0.01*	0.06*
UFE/NFO	-0.02*	-0.02*	-0.01	-0.01*	-0.06*	-0.01*	0.01	0.09*	0.00	-0.01*	-0.01	0.00

Table 3. Mean Absolute Forecast Errors

This table examines model performance for different equity value drivers. Panel A (Panel B) presents out of sample 1 year ahead mean absolute forecast errors for ROCE- (RNOA-) based models. † indicates that the model has a lower mean absolute forecast error than a random walk. The sample mean random walk forecast error for ROCE (RNOA) equals 0.1182 (0.0847).

Panel A. ROCE

	Core 0, Level 1	Core 0, Level 2	Core 0, Level 3	Core 0, Level 4	Core 1, Level 1	Core 1, Level 2	Core 1, Level 3	Core 1, Level 4
Leads 1, Lags 0	0.1095†	0.1105†	0.1118†	0.1125†	0.1095†	0.1046†	0.1084†	0.1068†
Leads 1, Lags 1	0.1089†	0.1105†	0.1132†	0.1130†	0.1086†	0.1053†	0.1104†	0.1104†
Leads 1, Lags 3	0.1084†	0.1128†	0.1157†	0.1152†	0.1085†	0.1079†	0.1128†	0.1116†
Leads 1, Lags 5	0.1093†	0.1156†	0.1190	0.1179†	0.1093†	0.1102†	0.1165†	0.1150†
Leads 5, Lags 0	0.1100†	0.1114†	0.1132†	0.1143†	0.1100†	0.1059†	0.1087†	0.1082†
Leads 5, Lags 1	0.1088†	0.1113†	0.1134†	0.1135†	0.1088†	0.1066†	0.1086†	0.1089†
Leads 5, Lags 3	0.1088†	0.1125†	0.1152†	0.1145†	0.1088†	0.1069†	0.1115†	0.1118†
Leads 5, Lags 5	0.1100†	0.1149†	0.1185	0.1185	0.1100†	0.1090†	0.1150†	0.1151†
Leads 10, Lags 0	0.1108†	0.1117†	0.1147†	0.1153†	0.1108†	0.1071†	0.1100†	0.1089†
Leads 10, Lags 1	0.1097†	0.1113†	0.1147†	0.1148†	0.1098†	0.1068†	0.1106†	0.1110†
Leads 10, Lags 3	0.1087†	0.1116†	0.1161†	0.1163†	0.1086†	0.1075†	0.1134†	0.1138†
Leads 10, Lags 5	0.1095†	0.1146†	0.1191	0.1201	0.1096†	0.1094†	0.1164†	0.1161†

Panel B. RNOA

	Core 0, Level 1	Core 0, Level 2	Core 0, Level 3	Core 0, Level 4	Core 1, Level 1	Core 1, Level 2	Core 1, Level 3	Core 1, Level 4
Leads 1, Lags 0	0.0915	0.0789†	0.0793†	0.0801†	0.0915	0.0751†	0.0783†	0.0772†
Leads 1, Lags 1	0.0909	0.0788†	0.0797†	0.0798†	0.1201	0.0786†	0.0835†	0.0784†
Leads 1, Lags 3	0.0905	0.0803†	0.0823†	0.0819†	0.1095†	0.0758†	0.0785†	0.0801†
Leads 1, Lags 5	0.0909	0.0823†	0.0848	0.0846†	0.1086†	0.0763†	0.0783†	0.0821†
Leads 5, Lags 0	0.0915	0.0797†	0.0810†	0.0820†	0.1085†	0.0767†	0.0799†	0.0788†
Leads 5, Lags 1	0.0912	0.0792†	0.0805†	0.0809†	0.1093†	0.0784†	0.0830†	0.0788†
Leads 5, Lags 3	0.0910	0.0804†	0.0815†	0.0813†	0.1100†	0.0763†	0.0786†	0.0796†
Leads 5, Lags 5	0.0913	0.0815†	0.0847	0.0840†	0.1088†	0.0763†	0.0791†	0.0818†
Leads 10, Lags 0	0.0921	0.0794†	0.0817†	0.0817†	0.1088†	0.0771†	0.0808†	0.0783†
Leads 10, Lags 1	0.0914	0.0791†	0.0815†	0.0815†	0.1100†	0.0779†	0.0834†	0.0790†
Leads 10, Lags 3	0.0911	0.0794†	0.0825†	0.0825†	0.1108†	0.0783†	0.0772†	0.0807†
Leads 10, Lags 5	0.0915	0.0810†	0.0843†	0.0855	0.1098†	0.0787†	0.0784†	0.0833†

Table 4. Model Alphas

This table examines model performance for different equity value drivers. Panel A (Panel B) presents Carhart (1997) alphas for ROCE- (RNOA-) based models. Standard errors are computed following Newey and West (1987) with a lag order of 3. Panel C Model 1 (2, 3, 4, 5, 6, 7, 8) presents OLS (5th, 10th, 25th, 50th, 75th, 90th, 95th percentile quantile) results of regressing the Carhart (1997) model alphas presented in Table 3 on indicators for different equity value drivers. Robust t-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10% level.

Panel A. ROCE								
	Core 0, Level 1	Core 0, Level 2	Core 0, Level 3	Core 0, Level 4	Core 1, Level 1	Core 1, Level 2	Core 1, Level 3	Core 1, Level 4
Leads 1, Lags 0	-0.0048	0.0085	-0.0031	-0.0088	-0.0043	0.0066	-0.0001	0.0090
Leads 1, Lags 1	0.0009	0.0172	0.0026	0.0041	0.0007	0.0026	-0.0016	0.0021
Leads 1, Lags 3	0.0026	0.0039	-0.0011	0.0059	0.0031	0.0113	0.0119	0.0153
Leads 1, Lags 5	0.0022	0.0041	0.0099	0.0042	0.0021	0.0138	0.0150	0.0053
Leads 5, Lags 0	-0.0316*	-0.0010	0.0386	-0.0218	-0.0317*	0.0154	0.0151	0.0436
Leads 5, Lags 1	-0.0155	-0.0027	0.0025	0.0514**	-0.0144	0.0336	0.0240	0.0190
Leads 5, Lags 3	-0.0067	-0.0323	-0.0085	0.0170	-0.0086	-0.0259	-0.0053	0.0389
Leads 5, Lags 5	-0.0013	-0.0255	-0.0071	0.0378	-0.0004	0.0433	0.0139	0.0119
Leads 10, Lags 0	-0.0113	0.0156	0.0041	0.0347	-0.0159	0.0380	0.0299	0.0431**
Leads 10, Lags 1	0.0246	0.0529**	0.0351	0.0477*	0.0132	0.0233	0.0340	0.0416**
Leads 10, Lags 3	0.0323	-0.0068	-0.0028	-0.0109	0.0346	0.0210	0.0170	0.0464*
Leads 10, Lags 5	0.0347	0.0095	0.0092	0.0228	0.0338	0.0028	-0.0097	-0.0055

Panel B. RNOA								
	Core 0, Level 1	Core 0, Level 2	Core 0, Level 3	Core 0, Level 4	Core 1, Level 1	Core 1, Level 2	Core 1, Level 3	Core 1, Level 4
Leads 1, Lags 0	-0.0844***	-0.0879***	-0.0844***	-0.0942***	-0.0846***	-0.0924***	-0.0896***	-0.0961***
Leads 1, Lags 1	-0.0843***	-0.0864***	-0.0872***	-0.0934***	-0.0844***	-0.0944***	-0.0919***	-0.0924***
Leads 1, Lags 3	-0.0867***	-0.0817***	-0.0844***	-0.0926***	-0.0862***	-0.0902***	-0.0936***	-0.0924***
Leads 1, Lags 5	-0.0923***	-0.0919***	-0.0870***	-0.0904***	-0.0924***	-0.0867***	-0.0839***	-0.0872***
Leads 5, Lags 0	-0.0237*	0.0559***	0.0946***	0.0535**	-0.0236*	0.0425**	0.0353**	0.0405*
Leads 5, Lags 1	-0.0132	0.0314	0.0389**	0.0588***	-0.0131	0.0595***	0.0443**	0.0455***
Leads 5, Lags 3	-0.0138	0.0048	0.0167	0.0196	-0.0151	0.0420*	0.0056	0.0500**
Leads 5, Lags 5	-0.0304	0.0165	0.0587*	0.0433	-0.0229	0.0508	0.0239	0.0526*
Leads 10, Lags 0	-0.0172	0.0392	0.0609***	0.0790***	-0.0180	0.0495**	0.0641***	0.0484*
Leads 10, Lags 1	0.0026	0.0765***	0.0589***	0.0291	-0.0037	0.0638***	0.0768***	0.0834***
Leads 10, Lags 3	0.0105	0.0755***	0.0509***	0.0678***	0.0061	0.0700***	0.0587***	0.0605***
Leads 10, Lags 5	0.0047	0.0532**	0.0512***	0.0545**	0.0065	0.0384*	0.0733***	0.0410**

Panel C. Alpha Determinant Model

Quantile	(1) Mean	(2) P5	(3) P10	(4) P25	(5) P50	(6) P75	(7) P90	(8) P95
Level 2	0.023*** (5.475)	0.036*** (5.893)	0.030*** (6.041)	0.030*** (10.873)	0.017** (2.053)	0.010*** (2.605)	0.010*** (4.222)	0.010*** (10.620)
Level 3	0.024*** (5.730)	0.037*** (6.485)	0.032*** (6.746)	0.030*** (10.637)	0.019** (2.346)	0.007** (2.170)	0.007** (2.581)	0.012*** (3.980)
Level 4	0.028*** (6.620)	0.038*** (6.255)	0.029*** (6.102)	0.032*** (9.101)	0.020** (2.348)	0.012*** (3.872)	0.012*** (7.113)	0.012*** (14.933)
Core	0.003 (1.046)	0.001 (0.280)	-0.001 (-0.632)	0.001 (0.449)	0.006** (2.018)	0.001 (0.713)	0.001 (0.391)	-0.000 (-0.185)
ROCE	0.093*** (29.019)	0.092*** (12.895)	0.092*** (26.669)	0.095*** (48.586)	0.099*** (32.622)	0.093*** (65.565)	0.090*** (38.793)	0.089*** (109.408)
Leads 5	0.115*** (23.851)	0.111*** (12.860)	0.100*** (23.679)	0.108*** (23.442)	0.124*** (18.419)	0.128*** (28.487)	0.131*** (47.010)	0.133*** (6.866)
Leads 10	0.133*** (29.692)	0.123*** (21.804)	0.123*** (40.377)	0.129*** (25.580)	0.144*** (23.010)	0.148*** (25.928)	0.152*** (74.005)	0.151*** (61.394)
Lags 1	0.006 (1.536)	0.008 (1.275)	0.001 (0.393)	-0.000 (-0.104)	0.009** (2.477)	0.009** (2.547)	0.002 (0.957)	0.005** (2.209)
Lags 3	-0.002 (-0.408)	-0.004 (-0.547)	-0.001 (-0.213)	-0.002 (-0.888)	0.001 (0.284)	0.003 (0.956)	-0.002 (-0.692)	-0.002 (-1.604)
Lags 5	-0.000 (-0.022)	-0.002 (-0.254)	0.002 (0.396)	-0.000 (-0.103)	0.004 (1.150)	0.003 (1.001)	-0.001 (-0.609)	-0.002 (-1.323)
ROCE × Leads 5	-0.114*** (-17.869)	-0.124*** (-12.654)	-0.119*** (-14.297)	-0.114*** (-17.601)	-0.125*** (-15.643)	-0.120*** (-13.596)	-0.104*** (-21.920)	-0.104*** (-5.307)
ROCE × Leads 10	-0.118*** (-18.927)	-0.121*** (-13.240)	-0.126*** (-25.334)	-0.122*** (-16.472)	-0.131*** (-16.020)	-0.120*** (-18.221)	-0.121*** (-36.093)	-0.120*** (-41.570)
Constant	-0.110*** (-21.294)	-0.140*** (-17.655)	-0.124*** (-20.705)	-0.122*** (-42.366)	-0.113*** (-13.149)	-0.094*** (-22.455)	-0.085*** (-30.890)	-0.084*** (-59.418)
Observations	192	192	192	192	192	192	192	192
Pseudo R ²	0.851	0.686	0.700	0.678	0.596	0.596	0.610	0.618

Table 5. Importance of Non-Linearities for Alpha Determinant Model

This table examines the importance of non-linearities for different equity value drivers. Model 1 (2, 3, 4, 5, 6, 7) presents 5th (10th, 25th, 50th, 75th, 90th, 95th) percentile quantile results of regressing the Carhart (1997) model alphas presented in Table 3 on indicators for the model's equity value drivers, an indicator that the model was estimated using Deep Learning (DL) rather than OLS, and interactions between the Deep Learning and equity value driver indicators. Robust t-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10% level.

Variables	(1) Mean	(2) Q5	(3) Q10	(4) Q25	(5) Q50	(6) Q75	(7) Q90	(8) Q95
DL × Level 2	0.005 (0.768)	0.009 (0.620)	-0.008 (-1.176)	-0.006 (-0.501)	0.014 (1.519)	0.009** (2.355)	0.003 (0.618)	-0.003 (-0.208)
DL × Level 3	0.025*** (3.805)	0.016* (1.741)	0.013** (2.055)	0.003 (0.216)	0.030*** (3.465)	0.012** (2.554)	0.007* (1.742)	0.009 (1.211)
DL × Level 4	-0.010 (-1.413)	0.002 (0.134)	-0.016** (-2.174)	-0.012 (-0.942)	0.007 (0.661)	-0.013** (-2.351)	-0.028*** (-3.724)	-0.038** (-2.213)
DL × Core	-0.002 (-0.567)	0.001 (0.259)	-0.002 (-0.528)	-0.001 (-0.317)	0.003 (1.105)	-0.002 (-0.710)	0.001 (0.315)	-0.000 (-0.031)
DL × ROCE	0.013* (1.895)	0.017 (1.177)	-0.005 (-0.694)	0.008 (1.238)	0.015 (1.634)	0.014** (2.076)	0.020*** (3.058)	0.011 (0.433)
DL × Leads 5	0.022*** (2.701)	0.037*** (3.648)	0.026*** (4.246)	0.034*** (7.226)	0.038*** (2.785)	0.012 (1.373)	0.007 (0.900)	0.013 (0.624)
DL × Leads 10	0.004 (0.751)	0.007 (0.839)	-0.001 (-0.127)	0.002 (0.388)	0.009** (2.065)	0.007* (1.849)	0.007 (1.536)	0.006 (0.378)
DL × Lags 1	-0.005 (-0.750)	0.000 (0.031)	0.001 (0.080)	-0.001 (-0.150)	-0.002 (-0.293)	0.001 (0.161)	0.001 (0.263)	-0.002 (-0.143)
DL × Lags 3	-0.007 (-1.044)	-0.004 (-0.431)	0.001 (0.125)	0.001 (0.172)	0.000 (0.066)	-0.004 (-1.327)	-0.004 (-1.015)	-0.007 (-0.374)
DL × Lags 5	0.010* (1.885)	0.004 (0.504)	0.005 (0.741)	0.010*** (2.771)	0.015*** (4.813)	0.012*** (5.152)	0.006 (1.530)	0.006 (0.515)
DL × ROCE × Leads 5	0.000 (0.015)	0.004 (0.313)	0.012 (1.008)	0.008 (0.755)	-0.008 (-0.730)	-0.003 (-0.228)	0.010 (1.281)	0.019 (0.663)
DL × ROCE × Leads 10	-0.022** (-2.087)	-0.008 (-0.366)	-0.019** (-2.103)	-0.021 (-1.419)	-0.047*** (-2.944)	-0.011 (-1.152)	-0.003 (-0.330)	-0.009 (-0.330)
Observations	384	384	384	384	384	384	384	384
Pseudo R2	0.811	0.644	0.659	0.633	0.561	0.598	0.624	0.638
Main Effects	YES	YES	YES	YES	YES	YES	YES	YES

Table 6. Estimator Comparison

This table examines performance across different profitability forecasting models. Models 1 to 6 regress value-weighted (VW) and equal-weighted (EW) hedge portfolio excess returns on asset pricing factors following Carhart (1997). Hedge portfolios are formed based on value-to-price ratio deciles. Value-to-price ratios are computed using market prices at the end of June and value estimates from a residual income model based on forecasts for 5 RNOA leads. RNOA forecasts in models 1 and 2 (3 and 4, 5, and 6) are obtained from a deep neural network (OLS, LAD) model without a focus on core items and 0 lags of all level 3 disaggregation variables. Newey and West (1987) t-statistics with a lag order of 3 are reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10% level.

Model	(1) NN VW	(2) NN EW	(3) OLS VW	(4) OLS EW	(5) LAD VW	(6) LAD EW
Mkt-Rf	-0.164 (-0.689)	0.174 (0.706)	0.013 (0.117)	-0.303 (-1.242)	-0.041 (-0.154)	-0.124 (-0.552)
HML	0.534*** (4.624)	0.524** (2.615)	0.209* (1.969)	0.085 (0.582)	0.301* (1.901)	0.125 (0.812)
SMB	0.170 (0.416)	1.330* (1.959)	0.278 (1.288)	0.655 (1.015)	0.927*** (2.933)	1.021 (1.277)
UMD	-0.153 (-0.617)	-0.085 (-0.392)	-0.114 (-0.715)	0.080 (0.424)	-0.007 (-0.043)	0.019 (0.092)
Alpha	0.095*** (4.108)	0.076** (2.564)	0.000 (0.010)	0.056 (1.142)	0.049 (1.704)	0.062 (1.688)
Observations	29	29	29	29	29	29
NW SEs	3 lags	3 lags	3 lags	3 lags	3 lags	3 lags

Table 7. Panel Data Regressions

This table examines cross-sectional determinants of returns. Models 1 to 6 regress market adjusted returns on value-to-price ratio deciles ranks (VP Rank) and controls. Value-to-price ratios are computed using market prices at the end of June and value estimates from a residual income model based on forecasts for 5 RNOA leads. Standard errors are clustered by firm. Robust t-statistics are reported in parentheses. ***, **, and * denote statistical significance at the 1, 5, and 10% level.

Model	(1) Ret _{t+1}	(2) Ret _{t+1}	(3) Ret _{t+1}	(4) Ret _{t+1}	(5) Ret _{t+1}	(6) Ret _{t+1}
VP Rank	0.005*** (3.24)	0.005*** (3.32)	0.021*** (3.19)	0.003** (2.23)	0.009*** (3.62)	0.004** (2.37)
Size	-0.197*** (-19.89)	-0.192*** (-19.82)	-0.186*** (-17.97)	-0.197*** (-19.56)	-0.191*** (-19.62)	-0.197*** (-19.87)
MtB	0.003 (0.28)	0.001 (0.13)	0.004 (0.36)	0.003 (0.29)	0.008 (0.71)	0.002 (0.23)
Ret _t	-0.039*** (-3.77)	-0.041*** (-3.96)	-0.038*** (-3.70)	-0.039*** (-3.79)	-0.042*** (-4.04)	-0.038*** (-3.67)
F-Score		0.003* (1.80)				
ΔNOA		-0.159*** (-2.79)				
VP Rank × Size			-0.002*** (-2.83)			
Loss				-0.033** (-2.00)		
VP Rank × Loss				0.006* (1.85)		
Z-Score					-0.241 (-1.30)	
VP Rank × Z-Score					-0.084** (-2.07)	
VP Rank × Tech						0.015** (2.18)
Observations	30,129	30,129	30,129	30,129	29,606	30,129
Adj. R ²	0.154	0.154	0.154	0.154	0.155	0.154
Fixed Effects	Firm & Year					