Dividend Forecasting in the Age of Machine Learning*

Xuesi Wang†

Leonidas G. Barbopoulos‡

Khaladdin Rzayev§

This version
September 12, 2024

*We are grateful for comments and suggestions from Yakov Amihud, Roni Michaely, Andre Santos, Jeffrey Wurgler, and Celim Yildizhan. We are responsible for any remaining errors or omissions.

†University of Edinburgh Business School, The University of Edinburgh, 29 Buccleuch Place, Edinburgh, EH8 9JS, UK. Email: X.Wang-344@sms.ed.ac.uk.

‡University of Edinburgh Business School, The University of Edinburgh, 29 Buccleuch Place, Edinburgh, EH8 9JS, UK. Email: leonidas.barbopoulos@ed.ac.uk. Tel: +44 131 650 8344. Corresponding author.

§University of Edinburgh Business School, The University of Edinburgh, 29 Buccleuch Place, Edinburgh, EH8 9JS, UK. Email: khaladdin.rzayev@ed.ac.uk. Koç University, Istanbul, Turkey, and Systemic Risk Centre, London School of Economics, London, UK.

Dividend Forecasting in the Age of Machine Learning

Abstract

We explore the application of machine learning (ML) techniques for dividend forecasting, demonstrating their superior accuracy over conventional prediction methods. We find that while the accuracy of dividend forecasts generated by analysts diminishes in more complex information structures, our dividend forecasts remain robust, highlighting the effectiveness of ML models in handling complex information domains. Next, we employ the ML–generated dividend forecasts to estimate expected returns and show their superior performance over traditional proxies in both in-sample and out-of-sample evaluations. Overall, our results highlight the importance of ML in dividend forecasting, contributing to the ongoing debate on whether ML can help design effective corporate finance policies and affect investment decisions.

Keywords: Dividend Forecast; Machine Learning; Expected Return; Payout Policy; Firm Complexity.

JEL Classification Codes: G11, G12, G14.

"Thirty years ago the best fund manager was the one with the best intuition... Now those who take a "scientific approach", using machines, data and AI, can have an edge."

David Siegel, co-chairman of Two Sigma (*The Economist, October 5th, 2019*)

1. Introduction

Recent advancements in financial technology and machine learning (ML), together with access to vast amounts of data, have revolutionized information production dynamics in financial ecosystem, profoundly influencing both financial decision-making and economic growth (Chalfin et al., 2016; Farboodi et al., 2019; Farboodi and Veldkamp, 2020; Dessaint et al., 2023; Dugast and Foucault, 2023). Scholars continue to argue that technological developments have significantly reduced the marginal cost associated with gathering and processing information, enabling entities to economically benefit from more information.[1] Consequently, stock prices have become more efficient, incorporating a broader array of information, thereby improving market reactions to firm events and reducing biases among market participants, resulting in lower forecasting errors (e.g., Verrecchia, 1982; Kim and Verrecchia, 1994). The extensive gathering and processing of information not only reduces the uncertainty faced by investors but also stimulates their efforts to acquire and act upon information pertaining to various fundamental aspects of firms (Goldstein and Yang, 2015). However, previous research, including studies on the role of ML in forecasting earnings, has overlooked the ability of ML to accurately predict dividends.

Distributing profits to shareholders via dividends plays a crucial role in the economic landscape, signaling financial vulnerability, influencing capital allocation decisions, enhancing shareholder wealth, as well as addressing investor preferences and taxation considerations. Therefore, dividend forecasts are a key element in capital allocation decisions, with investors demanding detailed insights into these forecasts. This is underscored by the substantial increase in the proportion of dividend-paying firms for which analysts also generate dividend forecasts, skyrocketing from 3% in 2001 to

[1]For instance, advancements in computational power and the implementation of ML algorithms empower investors to outperform traditional empirical asset pricing models (Gu et al., 2020), make use of highly unstructured textual data (Bybee et al., 2020), and gain an advantage over slower traders (Van Kervel and Menkveld, 2019). Technological innovations also accelerate information processing, leading to greater economies of scale (Chen et al., 2020).

nearly 90% in 2021, alongside an increase in the average number of analysts providing such forecasts from 1 to 5 over the same period, as shown in Figure (1). This trend highlights the importance of understanding the influence of technological innovations on dividend forecasting, a key scientific inquiry that this paper addresses. Two likely outcomes stand out. One possibility is the higher forecast precision, achieved by leveraging richer firm-level data and processing it at a much faster rate. Conversely, ML methods utilizing vast and diverse datasets could introduce more noise and distortions, potentially diminishing the accuracy of dividend forecasts (Dugast and Foucault, 2018). Determining which of these opposing effects dominates remains an open empirical question with significant implications for the allocation of resources and risk.

Previous research has also accepted dividend forecasts as an important input in financial decision-making and identified three primary approaches to dividend forecasting: (a) rely on historical realized dividend payouts, (b) rely on earnings forecasts multiplied by the payout ratio, and (c) consider analysts' dividend forecasts. The first two approaches, as discussed by Lintner (1956) and Brav et al. (2005), grow on assumptions that may not always hold true. They assume that historical dividends, expected earnings, and the payout ratio contain adequate information for accurately predicting future dividends. Additionally, they assume a linear relationship between expected dividends and their determinants, as well as a stable, long-term payout ratio. Nevertheless, empirical evidence offers limited support for these assumptions, particularly given that payout ratios are often estimated rather than directly observed. As illustrated in Figure (2), Compustat data reveals a significant upward trend in the proportion of firms that do not maintain a constant Dividend-Per-Share (DPS) compared to the previous fiscal year, rising from 60% to 85% since the 2000s. Additionally, we calculate the deviation in dividend payments from the target level implied by Earnings-Per-Share (EPS) and the Target Payout Ratio (TPR), following Leary and Michaely (2011). Our results indicate a similar upward trend, with the median absolute deviation increasing from 0.15 to 0.35. Both findings highlight an increased instability in dividend payouts since the 2000s, relative to both the previous year's level and the target based on earnings. Furthermore, we estimate dividend sensitivity to past dividends and earnings, using the approach of Leary and Michaely (2011). As shown in Figure (3), the slightly increased Speed of Adjustment (SOA) since the 2000s suggests a

reduced sensitivity to past dividend levels, while the consistently low sensitivity to earnings reflects the diminished role of earnings in determining dividends. The overall decline in the fitted R-square, from 0.5 to 0.3, further indicates the reduced explanatory power of past dividends and earnings in predicting dividend changes.

The third approach, centered on analysts' dividend forecasts, has garnered considerable attention in recent years. De La O and Myers (2021) utilize dividend forecasts from I/B/E/S to underscore the comparative significance of subjective cash flow news versus discount rate news. Bilinski and Bradshaw (2022) delve deeper into the informativeness and precision of analyst forecasts regarding dividends, demonstrating their substantive value (Francis and Philbrick, 1993; Bartov et al., 2002). However, Hoitash et al. (2021) and Barinov et al. (2022) reveal a diminishing accuracy and informative capacity of analyst forecasts, particularly when predicting dividends for companies operating within opaque or intricate informational landscapes.

The above discussion on the existing three approaches in generating dividend forecast all reveal the problems of relying limited information supported by empirical evidence. Consequently, a valid inquiry arises: can we enhance the accuracy of dividend forecasts by incorporating a broader spectrum of data and leveraging sophisticated ML methodologies?

Our answer to this question is, in short, yes. Based on the Mean Squared Error (MSE) decomposition framework developed by de Silva and Thesmar (2021), we begin by theoretically showing that combining dividend forecasts using both analyst consensus and forecasts generated from publicly available information outperform dividend forecasts solely based on Lintner's model and analyst consensus. Subsequently, we leverage a comprehensive dataset encompassing analyst predictions, financial statements, stock market data, and macroeconomic indicators. This is achieved through the application of three tree-based supervised learning regression models: (a) random forest, (b) gradient boost trees, and (c) extreme-gradient boost trees.

Tree–based methods ensemble algorithms offer two advantages over traditional methods, such as linear forecasting when constructing dividend forecasts. First, they accommodate nonlinear relationships between explanatory and forecasted variables, a feature often overlooked by traditional

econometric forecasting models (Gu et al., 2020; van Binsbergen et al., 2022). Second, the tree–based ensemble algorithm ensures an unbiased forecast, even with a large number of explanatory variables. This is achieved by leveraging all available relevant information to estimate the expected dividend, addressing a common challenge in econometric modeling highlighted by Kelly et al. (2022), where traditional OLS estimators may become unreliable with an abundance of explanatory variables.[2]

We find the superior performance of machine-learning dividend forecasts compared to those generated by analysts, as well as forecasts relying solely on historical dividends or those derived from earnings forecasts and payout ratios, in terms of accuracy. This superiority of ML forecasts becomes particularly pronounced over long-term horizons. These findings suggest that ML techniques are adept at analyzing a wider array of information compared to alternative methods. Furthermore, our analysis reveals that analysts tend to produce less accurate dividend forecasts for firms with more intricate information structures, whereas ML maintains consistently high accuracy levels. This underscores the economic benefits of employing ML in forecasting, especially for firms with complex information structures.

As an application, we use ML–generated dividend forecasts to compute expected returns via the Dividend Discount Model, often denoted as the implied cost of capital (ICC) in the literature. Following this, we engage in a comparative analysis–"horse race"–to compare our expected returns with benchmark metrics. These benchmarks stem from earnings forecasts, sourced from both analyst projections and ML algorithms, adjusted by an estimated plowback ratio (Pástor et al., 2008; Li et al., 2013). Across the timeframe spanning January 2005 to December 2021, we evaluate the predictive capacity of equal-weighted and value-weighted monthly returns on the broader market, as represented by the S&P 500 index. Our findings indicate that the ML–generated dividend–centric expected return consistently outperforms those derived from earnings adjusted by the estimated

[2]While non-parametric random forest regression facilitates data-driven forecasting and allows for the inclusion of numerous variables, our selection of relevant variables is informed by both theoretical and empirical studies on dividend payout determinants. This intentional selection, as opposed to arbitrary feature inclusion, is based on the understanding that dividend payout is primarily influenced by a firm's management policy. Thus, forming accurate expectations requires monitoring the relevant information that reflects managerial decision-making dynamics. For instance, we incorporate specific firm actions such as mergers and acquisitions (Zwiebel, 1996), tax payments (Yagan, 2015), share repurchases (Floyd et al., 2015), and the valuation of investment projects (Miller and Rock, 1985) into our predictors (please refer to the Online Appendix B.1). Consequently, the feature importance derived from the tree-based regression models also serves to validate the significance of these firm actions in determining dividend payouts.

payout ratio. This heightened predictability of returns persists even when contrasted with widely employed valuation metrics, business cycle indicators, both in- and out-of-sample periods, and when adopting a portfolio–based approach.

Our research contributes to two main arrays of studies. First, it aligns with the current trend of utilizing ML methods to shape projections of a company's future fundamentals, particularly focusing on earnings. van Binsbergen et al. (2022) use random forests to tackle nonlinearity in earnings predictions, establishing a statistically optimal benchmark for real-time firm-level earnings expectations. Additionally, de Silva and Thesmar (2021) conducts a thorough examination of earnings forecast accuracy, showcasing how ML techniques can help alleviate the bias and noise typically present in analyst forecasts or traditional econometric tools.

While ML techniques have gained significant attention in recent financial economics research for estimating earnings expectations, to our knowledge, the exploration of dividend forecasting remains relatively limited, if not absent. Our study aims to address this gap. This focus is crucial for two reasons. First, the increase in the proportion of dividend-paying firms since the early 2000s (Michaely and Moin, 2022) highlights a shift towards dividends as a preferred method for investment returns and hence an amplified demand for accurate dividend forecasts, as also shown in our Figure (1) (Bilinski and Bradshaw, 2022). Second, dividends and earnings forecasts are considered equally vital in the literature, with dividends offering additional insights into a firm's value beyond what earnings data can provide (Fama and French, 1998; Kaplan and Pérez-Cavazos, 2022). Consistent with this, Kane et al. (1984) and Ely and Mande (1996) suggest that investors often use the information on earnings and dividends together, which further affects stock prices through their combined impact. This suggests that information derived from dividend forecasts remains a significant input to asset pricing models and investment decisions, even after considering earnings forecasts. Hence, the real-time and statistically optimized dividend payout forecasts presented in our study, utilizing an ML approach, can offer investors a valuable benchmark to consider, complementing information from analyst forecasts and traditional linear models or earnings forecasts.

Our research also contributes to the literature by illuminating a key factor that enhances the accuracy of ML in the forecasting of dividends. Specifically, we observe that ML–based dividend fore-

casts are less susceptible to the complexities of firms compared to forecasts computed by analysts. This underscores the ability of machines or algorithms to process more intricate sets of information effectively. This finding suggests that the economic benefits of employing ML algorithms in forecasting fundamental data are particularly evident when dealing with complex information structures. Moreover, similar observations extend to other areas of the financial sector, such as the credit market. As emphasized by Fuster et al. (2022) and Berg et al. (2022), the adoption of advanced statistical technologies leads to increased adaptability and better utilization of information, resulting in more accurate assessments of creditworthiness. With the exponential growth of data generated by various business stakeholders, projected to reach 175 zetabytes by 2025,[3] we anticipate a growing role for ML techniques in the future.

Lastly, developing a forward-looking return measure has long been a central theme in financial economics. We introduce an expected return measure derived from the dividend discount model, integrating our ML–driven dividend forecasts as a key input. By comparing it against established benchmarks (Pástor et al., 2008; Li et al., 2013; Lee et al., 2020), we illustrate the superior predictive power of our expected return measure regarding future realized returns. It is important to note that while the computation of this novel expected return stands as a significant application of our ML–based dividend forecasts, our primary emphasis lies in the dividend forecast measure itself.

The paper is structured as follows: Section (2) builds a theoretical framework of the study. Section (3) provides an overview of the data and empirical methodology used to construct our machine learning dividend expectations and presents the corresponding results. Section (4) explores how firm complexity impacts the accuracy of dividend forecasts. Section (5) discusses the calculation of our expected return measure and examines its empirical performance. Finally, Section (6) provides the paper's conclusion.

2. Theoretical Framework

Several arrays of studies in the literature help us to understand how ML tools could impact dividend pay decisions, whether dividends could be predicted more accurately, the channels through which

[3] https://www2.deloitte.com/cy/en/pages/technology/articles/data-grown-big-value.html

those effects may occur, and where ML tools can have an edge relative to conventional methods. The related studies also include those that analyze the relative efficacy of ML methods in assimilating corporate information, studies that rely on ML to predict earnings, as well as studies that analyze the deficiencies originating from ML and big data.

2.1. What affects a firm's decision to pay dividends?

The inquiry into what factors influence a firm's decision to distribute profits through dividends, or to maintain smooth dividend payments, has long puzzled scholars in finance (Fama and French, 2001; Baker and Wurgler, 2004; Leary and Michaely, 2011). Fama and French (2001) study the characteristics of dividend-paying firms, revealing that shifts in these characteristics play a crucial role in the declining prevalence of dividend payers. The authors also note a diminishing tendency among firms to distribute dividends, given their evolving attributes. Building on this, Baker and Wurgler (2004) observe that the inclination to pay dividends rises (declines) when a measure of the stock market dividend premium is positive (negative). Similarly, Leary and Michaely (2011) identifies factors such as firm size, dividend yield, earnings volatility, and institutional ownership as significant determinants of dividend smoothing behaviors. For instance, they find that younger, smaller firms with volatile earnings tend to smooth dividends less, while cash-rich firms with low growth prospects and substantial institutional ownership exhibit more pronounced smoothing tendencies. Allen et al. (2000) argue that institutional investors, like pension funds, often view dividend-paying firms as more desirable due to tax considerations favoring shareholders. Consequently, the dividend allocation decisions made by a firm's executives directly impact the strategies of investors and portfolio managers. The growing demand for dividend forecasts further underscores the economic significance of this phenomenon. Figure (1) illustrates a notable rise in the proportion of dividend-paying firms for which analysts produce dividend forecasts, from 3% in 2001 to nearly 90% in 2021, alongside an increase in the average number of analysts providing such forecasts, from 1 to 5 over the same period. This surge in demand for dividend forecasts gains additional significance when viewed through the lens of Benamar et al. (2021) findings, suggesting that pre-event information demand can serve as a proxy for investor uncertainty and influence the market impact of significant events.

2.2. ML and cost of information

Recent research on the impact of ML in financial markets suggests that the cost of information analysis has decreased significantly with the integration of ML technologies. For instance, ML enables greater economies of scale in information processing (Chen et al., 2020). Moreover, advancements in computational capabilities coupled with the utilization of ML algorithms empower investors to surpass traditional empirical asset pricing models (Gu et al., 2020). This allows for the utilization of highly unstructured textual data (Bybee et al., 2020) and the ability to anticipate and capitalize on market movements ahead of slower traders (Van Kervel and Menkveld, 2019). Conversely, less sophisticated investors, such as retail traders, often lack access to advanced technical resources and analytical expertise. Consequently, they rely more heavily on public and social media sources, using manual efforts rather than automated strategies. Early studies, such as Huberman and Regev (2001), suggest that certain corporate disclosures may impact a stock's price only after gaining attention in higher-circulation press. More recent research by Pedersen (2022) offers theoretical insights into how discord among retail investors on social networks can lead to irrational trading behaviors, characterized by increased turnover, heightened volatility, and price momentum, followed by reversals.

2.3. How could more information lead to better accuracy?

Based on the preceding discussion regarding the determinants of a firm's dividend payout policy and the superiority of ML methods in information processing, an important question arises: Is the incorporation of complex information likely to result in more accurate forecasts? According to conventional dividend smoothing theories, the answer to this question is no, given that future dividends can be derived simply based on the information of historical dividends, target payout ratio, and future earnings (Lintner, 1956; Brav et al., 2005). However, the current rich information environment introduces uncertainties to this question, as investors increasingly have the ability to obtain information from various sources. We now present a direct comparison among the structures of conditional dividend expectation respectively formed by a classical parsimonious model, analyst forecast, and a combined forecast that incorporates all available information.

As a benchmark forecast, we can write the conditional expectation for future dividends obtained from the Lintner (1956)'s model:

$$\mathbb{E}_t^L(D_{t+1}|\Omega_t^1) = (1-\gamma)D_t + \gamma\mathbb{E}_t(E_{t+1}|\Omega_t^1)\mathbb{E}_t(\theta_{t+1}|\Omega_t^1), \tag{1}$$

Eq. (1) provides a benchmark expectation of future dividends through a linear function of (i) the current dividend level (D_t), (ii) expected future earnings $\mathbb{E}_t(E_{t+1}|\Omega_t^1)$, (iii) the expected payout ratio ($\mathbb{E}_t(\theta_{t+1}|\Omega_t^1)$) and, (iv) the SoA ($\gamma$) that is implicit in the payout policy. The information set Ω_t^1 encompasses the available information at time t that has predictability for future earnings and payout ratio. [4]

This benchmark forecast sheds light on firms' dividend payout policies from two distinct perspectives. First, it suggests that firms tend to maintain a consistent dividend payout level over time. Second, it highlights that a firm's dividend payout is closely linked to its expected earnings and payout ratio. When $\gamma = 0$, dividend smoothing is maximized, and the firm is expected to maintain a stable dividend equivalent to its current payout level. By contrast, with γ approaching 1, the firm is expected to adjust its dividend solely based on expected earnings and the payout ratio, allowing for greater flexibility in dividend adjustments.

On the other hand, the structure of dividend expectations provided by analysts, and those derived from a combined information set that includes both analysts' forecasts and other related state variables can be respectively denoted as $\mathbb{E}_t^{AL}(D_{t+1}|\Omega_t^2)$ and $\mathbb{E}_t^{Com}(D_{t+1}|\Omega_t^3)$, where $\Omega_t^1, \Omega_t^2 \subseteq \Omega_t^3$. Without loss of generality, the information set Ω_t^2 can contain both publicly available information of state variables and also private information originating either from the analyst's personal source of information (e.g., private communication with management (Brown et al., 2015)), or generated by her own skills used in understanding the firm's payout policy (So, 2013). For the information set Ω_t^3, it combines both the consensus analyst forecast and also all other potential related state variables.

In our Online appendix A, we adopt and extend an MSE decomposition framework introduced by de Silva and Thesmar (2021) and demonstrate how dividend forecast accuracy can benefit from

[4] In the classical Lintner's model, θ is traditionally regarded as a constant. However, in practice, θ is frequently estimated by leveraging historical payout ratios to address potential small-sample bias, a concept elaborated by Leary and Michaely (2011). Consequently, in this context, we treat θ as time-varying.

an information set with a more complex structure, i.e., Ω_t^3. We also explain why forecasts based solely on Lintner's model or analysts may have a large MSE. Importantly, our findings do not rely on strict assumptions regarding the form of analyst beliefs or the true DGP for dividends, and they do not necessarily suggest that Lintner's or analyst forecasts underperform in all situations. Rather, our goal is to demonstrate how dividend forecasts can potentially benefit from richer information. This perspective aligns with the findings of van Binsbergen et al. (2022), which suggest that incorporating additional features in the optimal estimator reduces errors because irrelevant information is automatically filtered out.

2.4. Shortcomings of ML and negative effects

ML tools, however, are not free of problems. Both academics and practitioners have raised concerns that computationally intensive black-box algorithms may end up identifying spurious correlations or data biases, ultimately underperforming conventional methods in the long run (Hulbert, 2020). Scholars have also argued that ML methods produce strategies that appear profitable, but load on difficult-to-arbitrage stocks and often fail when trading costs and other frictions are taken into consideration (Avramov et al., 2022). Moreover, ML tools are limited in their capacity to process soft information, compared to analysts (Bradshaw et al., 2020). Along these lines, Dugast and Foucault (2018) illustrate a trade-off between the speed and accuracy of decision-making and forecasting accuracy.

Another potential downside that has been explored in the algorithmic trading literature, including Weller (2018), is that automated models may have an adverse effect on stock price informativeness, despite their importance for transmitting available information into prices. For example, Lee and Watts (2021) find that when the SEC's "Tick Size Pilot Programe" increased the tick size and reduced algorithmic trading, pre-announcement stock returns were better able to predict the news of the upcoming earnings releases, consistent with improved price discovery. From a different angle, Baldauf and Mollner (2020) find that faster speeds enable HFT to be more successful at order anticipation, which can have a negative effect on information production due to informed traders having less time to trade before HFT reacts.

Undoubtedly, ML models rely on fast processing of big data may have little or no impact on stocks without large quantities of data. Begenau et al. (2018), and in a similar vein, Farboodi et al. (2022), investigate whether investors price firm news more efficiently with big data. They find that large firms benefit disproportionately from supplying more data, which enables investors to produce more accurate forecasts and reduce uncertainty. Furthermore, while the automated system may substantially escalate information acquisition and awareness, its effect on information integration may be minimal. For example, Blankespoor et al. (2018) show that although algorithmic information processing boosts disclosing firms' liquidity and trading volume, it does not affect the price discovery process.

3. Forecasting of Dividends Using ML

We begin by outlining our approach for constructing real-time dividend forecasts using predictive regressions based on machine-learning. We employ three nonlinear and non-parametric tree-based models: Random Forest (RF), Gradient Boost Trees (GB), and Extreme Gradient Boosting Trees (XGB). We then provide summary statistics and conduct a comparative analysis of the accuracy of our one- two- and three-quarter-ahead, as well as one- two- and three-year-ahead dividend forecasts against those based on Lintner's model (i.e., lagged actual dividends or expected earnings multiplied by the estimated payout ratio) and analyst forecasts of dividends (reported in I/B/E/S). In short, we find that our machine-learning dividend forecasts consistently outperform alternative methods used in the literature.

3.1. Data for DPS forecast

For our empirical exercise, we rely on data from four primary sources. We collect monthly analyst forecasts of dividend per share (DPS) for all U.S. firms available in I/B/E/S summary file for the period spanning January 2002 to December 2022. To mitigate the influence of extreme forecasts, we use the median values as the consensus forecast. We also obtain realized DPS and earnings per share (EPS) data from the I/B/E/S actuals file. It is worth noting that we opt for I/B/E/S data rather Compustat for actual DPS and EPS due to differences in accounting standards followed by I/B/E/S,

as suggested by van Binsbergen et al. (2022). We also obtain quarterly firm fundamentals from Compustat, monthly stock prices and returns from CRSP, and monthly macro-economic indicators from the Federal Reserve Economic Data (FRED). We enhance the efficiency of DPS forecasting while retaining a vast pool of observations by prioritizing fundamental variables with minimal missing data known to correlate with corporate payout policy, as demonstrated in previous research (Allen and Michaely, 2003; Farre-Mensa et al., 2014). Furthermore, we compute various financial ratios from these selected variables. In total, we employ a total of 80 explanatory variables with different frequencies to generate DPS forecasts across different time horizons. Table (**1**) provides a brief overview of our three sets of predictors (analyst forecasts, firm fundamentals, and macro variables). For a comprehensive list of all the variables and dataset details, please refer to Online Appendix B.1.

To be retained in our sample, a firm must have non-missing analyst forecasts of DPS, future realized DPS, and stock prices for each forecasting month. Additionally, we apply winsorization at the 1% and 99% levels and standardize all variables, following James et al. (2013). In instances where quarterly fundamentals of firms exhibit missing values (which comprise up to 5% of the total sample), we impute these values using industry-specific medians. Industries are categorized based on the Fama-French 30 industry portfolios.[5]

3.2. Forecast formation

We employ a rolling window approach to partition the data chronologically into 'training' and 'test' sets, as in van Binsbergen et al. (2022). Besides, it is also essential to ensure that the dependent variables included in the training set are accessible at the time of forecasting to limit any look-ahead bias. To achieve this, we introduce gap years with the same duration as the forecast horizon between the training and testing periods. As a result, each window comprises three sub-periods: the model formation year, the gap year(s), and the forecast formation year.[6]

[5] At certain time points, the sample may not encompass all industry categories when utilizing more intricate classification methods, such as those involving 48 or 49 industries, due to a limited number of observations. It's noteworthy that the various classification methods do not yield significant effects on our forecast results.

[6] We opted for a one-year training set to maximize the forecast's time horizon coverage. The overall forecast results remain stable when we experiment with two or three-year training periods. For example, suppose we are in 2004 and aim to predict DPS in 2005 (one-year-ahead). In this case, we would train our model using independent variables from 2002 and a dependent variable from 2003 to estimate the model. We then set 2003 as the gap year, and input the independent variables of 2004 to generate the forecasts of DPS in 2005. The forecast formation for two-year and

As analysts typically provide their forecasts on a monthly basis, often before the end of each month, we ensure that our forecast is built after the announcement of the analyst forecast. For other predictors updated quarterly or annually, we use their lagged values and ensure that this information is available at each forecast-time point. To optimize our model, we employ a 5-fold cross-validation procedure and fine-tune the hyperparameters within each rolling estimation window using the parameter ranges suggested by de Silva and Thesmar (2021).[7] The Online Appendix B.3 reports detailed information about the selected hyperparameter ranges and categories for each supervised learning method. Each rolling window is therefore constructed independently, and all forecasts are conditional and out-of-sample by design:

$$E_t^{ML}(DPS_{i,t+\tau}) = ML\Big[X_{i,t}^{AL}, X_{i,t}^{Fin}, X_t^{Macro}\Big],$$

where $X_{i,t}^{Fin}$ and X_t^{Macro} refer to firm fundamental and macro market variables, respectively. ML encompasses each of our three nonlinear supervised learning methods: RF, GB, and XGB.

In essence, the ensemble regression models are constructed to mitigate overfitting by aggregating predictions from multiple independently generated trees. For the baseline model RF, each tree in the forest is built from a random sample of the data, and it makes a prediction by averaging the outputs of all the individual trees. This approach combines the simplicity of decision trees with flexibility, reducing the risk of overfitting and improving accuracy by averaging multiple deep decision trees.

Unlike RF, which builds trees independently, GB builds trees sequentially, with each tree designed to correct errors made by its predecessor, focusing on minimizing residuals. This iterative process continues until a predefined number of trees is reached, or a specified level of performance is achieved. XGB, an enhanced version of GB, extends the principles of gradient boosting by incorporating advanced regularization (L1 and L2), which prevents overfitting by penalizing complex models. Furthermore, XGB improves upon traditional GB by optimizing computational resources, effectively handling sparse data, and allowing for the weighting of misclassified instances, thus refining the model's accuracy and execution speed.

three-year-ahead forecasts follows a similar pattern.

[7] We also select optimal hyperparameters using a split-sample approach, similar to the method employed by van Binsbergen et al. (2022). Our results remain robust.

In summary, both GB and XGB use boosting techniques with a focus on sequential tree construction to enhance forecast accuracy. However, XGB sets itself apart by incorporating regularization and weighted sampling, making its optimization process more stable and robust. The flowchart in Figure (4) below provides a brief comparison of the working processes of the three tree-based regression models. For a more detailed explanation of the algorithms, please refer to Online Appendix B.2.

3.3. Forecast results and evaluation

We train the ML regression model starting in January 2002, following the formation process mentioned above. Eventually, we collect a dataset containing 534,255 monthly one-year-ahead DPS forecasts for 7,213 unique firms and 316,083 monthly three-year-ahead forecasts for 5,047 unique firms.

Figure (5) illustrates an example of a regression decision tree used to extract one-year-ahead DPS forecasts from the RF Regression model. To simplify the structure, we limited the tree's depth to 3 levels. Starting from the root node, the initial selected feature is the analyst forecast of DPS, and the splitting is based on a threshold value of 1.0. The data then proceeds to subsequent nodes, each representing a further branch of the feature space, guided by additional feature thresholds. This sequential partitioning continues until a terminal leaf node is reached.

Table (2) presents a summary of statistics and accuracy evaluations for all the forecasts. In the top panel, we observe that the analyst forecast (AF) closely approximates the actual value ($Actual$) for short-term forecasts (quarters ahead) but tends to deviate as the forecast horizon extends to 3 years. The forecasts generated based on the dividend target on earnings using the long-term payout ratio are denoted as $EPS * payout$. To calculate a valid payout ratio, we replace the negative earnings with 0.06 $\times$ total assets as in Hou et al. (2012) and Guay et al. (2011). The payout ratio is estimated as the sum of past DPS and repurchases divided by EPS, with this ratio constrained to fall within the range of 0 to 1. Interestingly, forecasts produced by this approach consistently diverge the most from the actual values across various forecasting horizons. Compared to these two commonly used approaches, dividend forecasts generated using three ML methods, namely RF, GB,

and XGB, exhibit more consistent mean values that closely match the actual ones, which are also less influenced by the forecast horizon's length.

To further assess the overall quality of our forecasts, we consider two key metrics: Mean Squared Error (MSE) and the percentage of improvement compared to a naive forecast based on the assumption of complete dividend smoothing, where future DPS is assumed to be equal to the lagged period of realized DPS (Yagan, 2015). Based on these two metrics, forecasts derived from earnings multiplied by an estimated payout ratio yield the highest MSE and fail to surpass the benchmark based on lagged DPS at any forecast horizon. This result address two important limitations in the related literature. First, it suggests that actual dividend payouts by firms may not consistently follow historical payout ratios, which aligns with the reduced significance of the target payout ratio, as indicated by survey results (Brav et al., 2005). Second, we employ the common practice of substituting negative EPS and abnormal DPS (e.g., for firms with a history of negative EPS but maintaining a high level of dividend payment) in line with existing literature (Li et al., 2013; Bilinski and Bradshaw, 2022). However, our results indicate that this approach may further introduce errors.

By contrast, analyst forecasts exhibit a lower MSE when compared to forecasts derived from earnings. However, as noted, analyst forecasts demonstrate increasing MSE as the forecast horizon extends. For instance, the MSE increases from 0.008 for one quarter ahead to 0.400 for three years ahead. More importantly, at the three-year-ahead forecast horizon, analyst forecasts fail to surpass the performance of the naive benchmark, resulting in a 39.2% higher MSE compared to the benchmark.

Our ML forecasts (RF and GB) consistently yield the lowest MSE values, outperforming both the lagged DPS benchmark and the analyst forecasts across all forecasting horizons. As analyst forecasts serve as critical input predictors in generating ML forecasts, the accuracy of ML forecasts also tends to decrease as the forecasting horizon increases. However, because our input predictor sets include other essential variables such as EPS, lagged DPS, and stock price, our forecasts can maintain a robust and superior performance compared to those relying on limited information, such as analyst forecasts.

Among the three ML methods, XGB appears to perform the least effectively, producing a higher MSE of 0.113 at the one-year-ahead forecast horizon, even exceeding that of the analyst forecast. Hence, we create a composite ML forecast, denoted as $Comp$, by averaging the predictions from RF and GB. Our findings indicate that this composite approach consistently delivers relatively more accurate forecasts compared to relying solely on either of the two individual ML methods.

Overall, our findings indicate that ML forecasts, which incorporate analyst forecasts alongside other publicly available information, consistently outperform all other forecasting methods across a range of forecasting horizons. While analyst forecasts demonstrate strong performance in the short term, their informational value diminishes as the forecast horizon extends. Additionally, the product of earnings forecasts and payout ratios does not prove to be an effective method for dividend forecasting.

Having established the superior performance of ML forecasts over alternative methods, it is crucial to examine the key contributors to the construction of dividend forecasts. Figure (6) presents the feature importance analysis for one, two, and three-year-ahead DPS using the RF forecast method. [8]

Impurity importance is a measure defined by the extent of the decrease in the sum of squared errors when a specific feature is used to split within each tree individually and then averaged across the entire forecasting model. The analyst forecast of dividends, as well as the realized dividends from the past year, emerge as the most critical features, collectively accounting for over 60% of the impurity importance. Interestingly, when the forecast horizon extended up to three-years-ahead, the significance of past DPS surpasses that of the analyst forecast. Additionally, other predictors such as analyst forecasts of earnings and realized earnings are found to contain valuable information for predicting future dividends. Other fundamental variables related to the firm, such as stock price, operational EPS, debt levels, and tax payments, also contribute significantly to the forecasting outcomes. These results reinforce the rationale of using realized dividends or expected earnings as indicators for inferring future dividend payouts. However, they also emphasize the importance of incorporating a broader range of features to enhance the accuracy and comprehensiveness of dividend

[8] The feature importance results obtained from GB and XGB exhibit similar patterns.

forecasts.

In addition, we employ Partial Dependence Plots (PDPs) to provide further insights into the marginal contributions of key features to the output forecast. PDPs help us gain a deeper understanding of the relationship between the response variable and specific features of interest while taking into account the average effect across the distribution of other variables in the dataset. We select four key variables based on the previously determined feature importance: analyst forecasts of DPS and EPS, past realized DPS, and stock price. As illustrated in Figure (7), the PDPs reveal significant nonlinearity in the relationship between future DPS and these four key features. This empirical evidence, highlighting non-linear patterns in the relationships between features and the outcome variable, may underpin one of the primary reasons for the enhanced forecasting accuracy achieved through the utilization of ML and provides further support for the application of nonparametric tree-based models in forecasting dividends.

4. Does Firm Complexity Affect Forecasting Accuracy?

Our results presented above indicate that ML–based dividend forecasts outperform traditional approaches, such as analyst forecasts. In the theoretical framework described in Section 2, we propose an economic mechanism to explain this outcome. The core concept is that ML models have the capacity to leverage more comprehensive information for forecasting compared to analysts. In this section, we empirically test this hypothesis.

Prior research demonstrates that analysts often incur larger EPS forecast errors when they produce forecasts for more complex firms, such as conglomerates or multi-segment companies, possibly due to a lack of industry-specific expertise (Barinov et al., 2022). Lehavy et al. (2011) and Loughran and McDonald (2020) also find that various measures of analyst forecast imprecision are inversely related to firm complexity. This discussion implies that firm complexity serves as an ideal setting to test our main hypothesis regarding the superior ability of ML to process complex information. To be precise, if ML does indeed possess the capacity to effectively utilize more comprehensive information compared to analysts, then the accuracy of machine-generated forecasts should be less affected by the complexity of firms' information environments when compared to forecasts produced by analysts.

This is the main hypothesis that we examine in this section.

Firm complexity is commonly measured using proxies such as firm size and the number of segments. However, Hoitash and Hoitash (2018) introduce a novel metric called account reported complexity (ARC), which is based on the count of monetary items disclosed in eXtensible Business Reporting Language (XBRL) within company filings. ARC is derived from disclosures that encompass a wide array of items, including investments, R&D expenses, and acquisitions and it effectively captures the overall complexity of a firm's structure. We employ this measure in our study because it more directly captures economically significant information disclosed in company filings.

We collect annual ARC data from the Hoitash & Hoitash website.[9] The ARC dataset covers the period starting from 2009 and includes multiple measures derived from public data available on the SEC's website. Subsequently, we calculate annual squared forecast errors for three-year-ahead DPS forecasts for each firm in our sample by averaging the available monthly squared forecast errors. We choose to assess the accuracy of three-year-ahead DPS forecasts because our earlier findings suggest that analysts face challenges in providing accurate forecasts at longer horizons. This may be attributed to the requirement of analyzing a more extensive dataset when forecasting dividends over longer time horizons; indeed, this is the concept we aim to test. To link our sample with the ARC dataset, we use firms' fiscal year and CIK code, which the ARC data uses as identifiers. This matching process results in a dataset comprising 24,234 firm-year observations spanning from 2009 to 2021.

We conduct two tests. In the first univariate test, we sort the firms by ARC and split them into five portfolios, indicating rising complexity from the first to the fifth group. As shown in Table (3), Panel A, we observe a clear, monotonically increasing trend in analyst forecast errors along with increasing firm complexity. Notably, there is a significant difference of 0.088 in the time-series Mean Squared Error (MSE) for analyst forecasts between the most complex and simplest groups, as indicated in Column C-S of the table. This observation aligns with our expectation, suggesting that analysts face challenges in producing dividend forecasts for firms with more complex information environments.

[9] https://www.xbrlresearch.com/firm-complexity/

We do not observe the same pattern when we examine the ML forecasts. First, we find that ML forecast errors consistently remain lower for each complexity group and exhibit smoother changes across different groups. Furthermore, for ML forecasts, the difference in MSE between the most complex and simplest firms is not statistically significant and it is also about 63% lower than the difference observed for analyst forecasts.

As a more comprehensive test, we construct annual cross-sectional regressions of the two squared errors on ARC for each year, similar to the approach used by Fama and MacBeth (1973) in the Fama-MacBeth regression. In Table (**3**), Panel B, Column (A) presents the regression results on ARC without control variables, while Column (B) includes controls. In addition to firm fundamentals, we incorporate two dummy variables, $Conglo$ and $GeoMulti$. Following Barinov et al. (2022), $Conglo$ takes a value of one if the firm is a conglomerate with multiple business segments in the same year, and $GeoMulti$ is one if the firm generates sales from multiple geographic segments.

Our findings reveal that, regardless of whether controls are included or not, ML forecast errors do not exhibit a significant slope with respect to ARC or other complexity measures. In contrast, analyst forecast errors show a significant positive association with firm complexity, and this relationship remains consistent after controlling for other variables such as firm size and book-to-market ratio. The estimated slope suggests that, holding other variables constant, a hundred-unit increase in monetary items disclosed in XBRL is correlated to a 0.035 increase in analyst forecast squared errors of DPS. As for the other two dummy measures, neither of them exhibits significance in forecasting errors.

In summary, the findings in this section suggest that the advantages of employing ML for dividend forecasts stem from the ML algorithms' ability to effectively analyze more comprehensive information. Therefore, the benefit of using ML in dividend forecasting is particularly pronounced for firms with more complex information environments.

5. Application: Estimating Expected Returns Based on Forecasts of DPS

Our main findings discussed above demonstrate that ML forecasts of DPS are associated with higher forecasting accuracy relative to analysts' forecasts reported in I/B/E/S, which is largely attributed to the ability of ML tools to process more information at a faster speed. This has direct implications in asset pricing and more precisely the estimation of expected returns through stock evaluation methods, such as Dividend Discount Models. Extant studies (Kane et al., 1984; Ely and Mande, 1996) propose that investors leverage both earnings and dividends information, influencing stock prices through their interplay. Information derived from DPS forecasts remains a vital component in asset pricing models, maintaining its relevance even when considering earnings forecasts. We therefore expect, and test, that improved precision in DPS predictions can enhance the overall accuracy of stock price estimations. Moreover, exploring the efficacy of ML in generating more accurate DPS forecasts offers insights into market efficiency. For example, if ML consistently produces more accurate dividend forecasts compared to what is already reflected in security prices, this will suggest a potential inefficiency in the market's assimilation of all available dividend-related information. This would therefore imply opportunities for (further) refining how the market processes and reacts to dividend-related information.

Estimating expected returns is a challenging task, a main difficulty of which arising from the forward-looking nature of expected returns, while many finance models such as the CAPM heavily rely on historical data. To address this challenge, a strand of literature proposes an approach to estimate forward-looking expected returns based on forecasts, commonly referred to as the implied cost of capital (ICC) in the literature. This approach has been shown to measure expected returns effectively (Lee et al., 2009; Hou et al., 2012; Li et al., 2013).

Dividend forecasts are a crucial input in forecast-based expected return models. Given that our ML models can generate dividend forecasts with reduced errors, this may also allow us to obtain a more accurate expected return. Therefore, in this section, we incorporate our ML–generated dividend forecasts into the computation of forecast-based expected returns. First, we revisit the rationale behind forecast-based expected returns as a suitable proxy for expected returns. We then

compare our expected return measure against the traditional method of computing ICC, which relies on forecasts of EPS and the payout ratio. Our findings demonstrate that expected returns calculated using our DPS forecasts outperform the alternatives as a proxy for expected returns.

5.1. Expected return imputed from Free Cash Flow discount models

Conceptually, ICC is defined as the value of the internal return, denoted as r_{ICC}, that solves the infinite Gordon dividend discount model:

$$P_t = \sum_{k=1}^{\infty} \frac{E_t[DPS_{t+k}]}{(1+r_{ICC})^k}, \tag{2}$$

which indicates that the present value of equity price P_t equals the sum of all the discounted future expected dividend payouts $E_t[DPS_{t+k}]$. Combining this with the log-linear approximation demonstrated in Campbell and Shiller (1988), Pástor et al. (2008) shows that under the assumption that both dividend growth and expected returns follow a stationary AR(1) process, r_{ICC} is perfectly correlated with the expected return μ_t:

$$r_{ICC} = \frac{\kappa}{1-\lambda} + (\mu_t - \frac{\kappa}{1-\lambda})\frac{1-\rho}{1-\rho\lambda}, \tag{3}$$

where κ and λ are constants determined by the AR(1) process of μ, and $\rho = 1/(1+exp(\bar{d}-p))$. Therefore, ICC is considered a natural proxy for expected return, as it encapsulates information about future returns without relying on the noisy realized returns (Gebhardt et al., 2001; Easton and Monahan, 2005).

In practice, r_{ICC} can be empirically determined by solving the transformed finite Gordon model, which decomposes the original Eq. (2) into a sum of discounted expected free cash flow terms $EFCF_{t+k}$ (i.e., firm payouts) and a discounted terminal value term TV_{t+T}.

$$P_t = \sum_{k=1}^{T} \frac{E_t[EFCF_{t+k}]}{(1+r_{ICC})^k} + \frac{TV_{t+T}}{(1+r_{ICC})^T}. \tag{4}$$

We initially employ the finite Gordon model to calculate two ICCs (denoted as r_e) as the benchmark expected returns for comparison with our forecasted dividend-based expected return. The first ICC is generated using the analyst forecasts of earnings times the payout ratio as $EFCF$ (Pástor et al., 2008; Lee et al., 2009; Li et al., 2013). This computation is based on three main assumptions

related to the growth process of earnings and the payout ratio, as well as the imputation of abnormal values, as discussed in Table (4). For the calculation of the second ICC, we replace the analyst forecasts of EPS with random forest forecasts using the methods outlined in van Binsbergen et al. (2022). The substitution of $EFCF$ is driven by the aim of mitigating the subjective bias that may arise from analyst forecasts of earnings, which has been identified as a key factor distorting the effectiveness of ICC as a proxy for expected return (Lee et al., 2020). We follow the approach introduced in Pástor et al. (2008) to obtain the two baselines r_e, and the detailed computation process is contained in Online Appendix C.

For the calculation of our dividend-based ICC (denoted as r_d), we utilize the composite ML forecasts of one, two, and three-year-ahead dividends as a proxy for $EFCF$. In this case, we assume that dividend grows perpetually at the average GDP growth rate. Therefore, the computation of our ICC relies solely on available information without the need for imputation of long-term cash flow expectations.

The calculation details of r_d are provided in Table 4. We first require the input of one-year-ahead and two-year-ahead ML forecasts of DPS to be non-missing and non-zero. To address the issue of missing and zero forecasts, which represent approximately 7% of the total sample, we employ an imputation technique. Specifically, we replace these missing or zero forecasts with a value equivalent to four times the one and two-quarter-ahead forecasts, assuming that these firms will make annual dividend payments consisting of four identical quarterly dividends. For the firms without effective forecasts of three-year-ahead DPS, which comprises 14.5% of the total sample, we substitute them with the two-year-ahead DPS forecasts times the last gross growth rate ($(E_t[DPS_{t+2}])^2/E_t[DPS_{t+1}]$), assuming they maintain the same growth rate from the second to the third year. This approach is in line with Pástor et al. (2008). We calculate a rolling average of annual nominal GDP growth rates as the perpetual growth rate of DPS.

We then compute r_d using the constant dividend growth model with the three-year DPS forecasts. To avoid inference from extreme values, for each month, we drop the top and bottom 0.5% of the ICC values consistent with Pástor et al. (2008). The resulting ICC values are constrained within the range of 0 to 0.3, following Hommel et al. (2023), and approximately 3.9% of observations are

removed due to this trimming process.[10] This restriction is applied to ensure that the ICC values are meaningful and do not include negative values, as a negative ICC is not interpretable as a discount rate, and excessively high ICC values may indicate unreliable high reinvestment rates. In total, we obtain 474,255 DPS-based ICC values for 6,567 unique companies spanning from January 2005 to December 2021.

5.2. Testing our expected return: In-sample analysis

In the first test, we investigate the predictability of future realized returns in the aggregate market, as represented by the S&P 500 index, using the aggregate expected return measures. To construct the aggregate expected return, we consider all available constituents of the S&P 500 index for each month spanning from January 2005 to December 2021. We employ both equally-weighted and value-weighted aggregation approaches. After aligning our sample with the S&P 500 companies listed in the CRSP index constitution data file, our sample, on average, comprises 469 firms each month. This sample coverage represents approximately 92% of the total market capitalization.

Prior to examining the relationship between the aggregate expected return and future realized returns, we present a time-series of the S&P 500 value-weighted expected return derived from ML forecasts of DPS in Figure (8), from January 2005 to December 2021. Of particular note is the peak of our expected return, which occurs at 9.9% in April 2009, coinciding with the near end of the 2008 financial crisis. During the two NBER recession periods that fall within our study's time frame, the expected return measure demonstrates a significant increase, consistent with the counter-cyclical behaviour of expected returns as suggested by existing studies (e.g., Li et al. (2012), Li et al. (2013)). Following the first recession, the expected return gradually decreases until the outbreak of the COVID-19 pandemic in February 2020.

Next, we run the following univariate predictive regression framework for different expected return proxies as suggested by Fama and French (1988):

$$\sum_{h=1}^{H} \frac{r_{t+h}}{H} = \beta_0 + \beta_1 \times M_t + \epsilon_{t+h}, \tag{5}$$

[10] Other methods of handling extreme values, such as winsorization at the 1% level, do not significantly affect the subsequent application results.

where r_{t+h} represents the continuously compounded monthly realized returns, either equal-weighted or value-weighted, in excess of the compounded three-month T-bill rate. We present the results for three different monthly prediction horizons, namely $H = 1, 6, 12$, in the interest of brevity, although the results hold at longer horizons as well. M_t donates the vector of expected return proxies. Our primary expected return proxy relies on ML–generated dividend forecasts, and we assess its performance in comparison to expected returns computed using earnings forecasts and various widely-used valuation ratios, consistent with the methodology detailed in Welch and Goyal (2008). Specifically, we include dividend-price ratio (dp), earnings-price ratio (ep), payout ratio (py), default spread ($Default$), and term spread ($Term$). We compute the aggregate market ratios dp, ep, and py using the aggregate dividend and earnings data obtained from Michael R. Roberts' website.[11] $Default$ is determined as the difference between BAA and AAA-rated corporate bond yields, while the $Term$ is calculated as the difference between the 10-year Treasury yield and the three-month bill rate. We source all macroeconomic data from the Federal Reserve Economic Data (FRED) database.

Table (5) presents summary statistics for the relevant variables. The two benchmark expected return measures are calculated using an EPS-based model, as listed in Table (5). These calculations involve ML techniques and analyst forecasts of EPS. Overall, the DPS-based expected return has a lower mean and standard deviation compared to both the equally-weighted and value-weighted EPS-based expected returns. Specifically, the DPS-based equally-weighted expected return has a mean and standard deviation of approximately 8.60% and 0.32%, respectively, which is relatively lower than the EPS-based equally-weighted expected return, with a mean and standard deviation of around 9.5% and 1.12%.

All expected return measures exhibit persistence, with a first-order autocorrelation exceeding 0.90, consistent with the expected characteristics of a robust expected return proxy, as suggested by Lee et al. (2010). This autocorrelation remains significant even at 12 lags. The ADF statistics and corresponding p-values indicate that we can reject the null hypothesis of a unit root for r_d, $r_{e,ML}$, $Default$, D/P, and py at least at 10% level. However, we fail to reject the null hypothesis for analyst forecast of EPS-based expected return measures, $Term$, and the E/P ratio.

[11] http://www.econ.yale.edu/~shiller/data.htm

Table (6) tabulates the results of univariate predictive regressions using Eq.(5) to forecast aggregate realized market returns. The left panel presents the outcomes of regressing equally-weighted returns ($ewret_{t+h}$) on predictors, while the right panel uses value-weighted returns ($vwret_{t+h}$) as the dependent variable. We initially report the Newey and West (1987) t-statistic with lag numbers equal to the forecasting horizon length.

In general, our findings indicate that short-term return predictions are relatively modest. However, even in the short term, the ML–forecasted DPS-based expected return (r_d) consistently outperforms other proxies. Specifically, for the one-month-ahead forecasting of $ewret_{t+1}$ ($vwret_{t+1}$), r_d demonstrates a significant coefficient of 3.178 (2.887), with corresponding t-statistics of 2.071 (2.391) and the highest R-squared of 4.04% (4.18%). Moreover, expected return imputed from ML forecast of EPS, $r_{e,ML}$, also demonstrates a significant coefficient of 0.533 (0.441), with corresponding t-statistics of 2.161 (2.480) and a lower R-squared. In contrast, the expected return measure based on analyst forecast of EPS, $r_{e,AL}$, do not predict one-month-ahead returns effectively. Similarly, the three valuation ratios and two business cycle indicators also fail to forecast one-month returns.

For longer horizons, such as 6 and 12 months ahead, we generally observe higher R-squared and t-statistics for all predictors that is mechanistically caused by the using of overlapping observations and the persistence of the predictors (Stambaugh, 1999). We apply an robust instrumentation approach recently proposed by Kostakis et al. (2015) to reduce this mechanistic bias and find the coefficients of r_d remain constantly statistically significant. Nevertheless, at these extended forecast horizons, r_d continues to outperform other expected return measures and other predictors. In addition, $r_{e,ML}$ also remain robust predict power for different cases. They consistently exhibit the highest t-statistics and R-squared values. This outcome underscores the clear advantage of using expected returns calculated through ML–generated dividend forecasts over other benchmark expected return measures in predicting future realized returns.

5.3. Testing our expected return: Out-of-sample analysis

Having established the superiority of our ML forecasted DPS-based expected return through in-sample regression, the next step is to assess its performance in out-of-sample predictions. This

evaluation focuses on the model's ability to make predictions based solely on historical information available at each forecasting time point. This test holds significance because previous studies, such as Welch and Goyal (2008), show that while expected return predictors work well in in-sample, they tend to perform poorly in out-of-sample. In this test, we use a forecast period from January 2010 to December 2021, with a 60-month estimation window, as well as an forecast period from January 2015 to December 2021, employing a 120-month estimation window.

The out-of-sample R^2 using a length of m estimation window is expressed as (Campbell and Thompson, 2008):

$$R^2_{OS} = 1 - \frac{\sum_{k=1}^{T-m}(r_{m+k} - \hat{r}_{m+k})^2}{\sum_{k=1}^{T-m}(r_{m+k} - \bar{r}_{m+k})^2}. \tag{6}$$

We employ both rolling and expanding window methods to fit the model and predict future returns, a technique also employed by Huang and Kilic (2019). In the rolling window method, the first out-of-sample forecast $\hat{r}_{m+1}$ is obtained using parameters estimated from estimation using the initial m observations. Subsequently, the second forecast is obtained with the estimation using the next m observations starting from the second time point, and so on. Concurrently, the historical average $\bar{r}_{m+k}$ is computed as the average excess return over the most recent m periods. In the expanding window method, we modify the approach by using all available past information for estimating both $\hat{r}_{m+k}$ and $\bar{r}_{m+k}$. As an evaluation criterion, a positive R^2_{OS} indicates that the predictor results in a lower Mean Squared Error (MSE) compared to predictions based on a simple historical average.

We also conduct a test to examine the null hypothesis, where $R^2_{OS} < 0$, as opposed to the alternative hypothesis, where $R^2_{OS} > 0$. This test employs the adjusted Mean Squared Prediction Error (MSPE) statistic introduced by Clark and West (2007):

$$f_{t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - [(r_{t+1} - \hat{r}_{t+1})^2 - (\bar{r}_{t+1} - \hat{r}_{t+1})^2]. \tag{7}$$

f_{t+1} is subsequently regressed on a constant term, and the test relies on the p-value obtained from a one-sided t-statistic.

Table (7) presents the R^2_{OS} values and their corresponding p-values for a one-sided test applied to one-month-ahead equal and value-weighted returns. When employing a 60-month rolling window and 120-month rolling window for out-of-sample forecasts of equal-weighted returns, r_d delivers the

highest positive R^2_{OS} values, reaching 1.26% and 4.66%, respectively. Importantly, all of these R^2_{OS} values are statistically significant, with a minimum significance level of 5%, as determined by the adjusted-MSPE statistic's p-values. In the case of forecasting value-weighted returns, r_d continues to produce positive and statistically significant R^2_{OS} values when using a rolling window approach. Transitioning to the expanding window approach, r_d gains higher positive and statistically significant R^2_{OS}.

On the other hand, both r_e variables fail to produce positive and statistically significant R^2_{OS} values in all instances. Interestingly, in out-of-sample forecasting, the poorly in-sample performed $r_{e,AL}$ occasionally outperforms $r_{e,ML}$, suggesting that a more accurate measure of EPS expectations may not necessarily enhance the ability to capture the out-of-sample dynamics in forecasting realized returns. Among the other predictors, it is observed that only $Default$ occasionally produce positive and significant R_{OS} values at a 10% significance level. However, none of them consistently achieve positive and statistically significant R^2_{OS} values across all cases.

One limitation of the out-of-sample tests, as highlighted by Welch and Goyal (2008) and Huang and Kilic (2019), is its sensitivity to the choice of estimation method and forecast period, a point that our results also confirm. To address this limitation and ensure the consistency of superior out-of-sample performance, we examine the differences between the cumulative squared forecast errors obtained using the historical average return and those obtained using the tested expected return proxy. In this test, a positive slope indicates a sustained advantage in out-of-sample forecasting compared to the historical average.

Figure (9) illustrates the findings. Our expected return measure maintains a consistently positive slope in the forecast period spanning from January 2010 to December 2021, except for the short period of COVID outbreak. In contrast, the expected return imputed from analyst forecasts of earnings displays a negative difference in cumulative squared error from 2013 to 2020, while the expected return imputed from ML forecasts of earnings exhibits an increasing error in recent years. Among the other expected return proxies, only D/P weakly surpasses the historical average return during the tested period, with a difference that remains consistently close to zero. Consequently, we confirm that the ML–forecasted DPS-based expected return demonstrates relatively robust out-of-sample

return predictability.

5.4. Testing our expected return: Variance of measurement errors

Following our analysis presented in the previous section, we demonstrate the superiority of our DPS-based expected return by showcasing its strong track record and superior predictability of future realized returns via predictive regression tests. In this section, we further evaluate its performance using a novel metric, namely the variance of measurement errors. Lee et al. (2020) introduce a parsimonious framework based on treatment effect studies, suggesting that minimizing measurement error variance (MEV) is necessary and sufficient for unbiased expected return proxies. The authors argue that relying solely on general evaluations derived from predictive regression, such as mean squared error and the significance of slope coefficients, is inadequate and can potentially lead to misleading conclusions. While our paper does not specifically assign a treatment effect, we employ this MEV-based evaluation as a valuable complementary assessment for our expected return proxy, derived from ML–based dividend forecasts.

Lee et al. (2020) define that if there is lack of correlation between the news of treatment effect and expected return, the time-series MEV for firm i takes the form of Equation (8):

$$MEV_i = var_i(\hat{er}_{i,t}) - 2cov_i(r_{i,t+1}, \hat{er}_{i,t}) + var_i(er_{i,t}), \tag{8}$$

where $var_i(\hat{er}_{i,t})$ donates the time-series variance of the expected return measure, $cov_i(r_{i,t+1}, \hat{er}_{i,t})$ is the time-series covariance between the tested expected return measure and future realized return, $var_i(er_{i,t})$ is the variance of the true unobservable expected return. Expected return proxy with lower MEV is considered more informative about time-series variations in expected return and leads to less biased estimates of treatment effects.[12]

The rationale for constructing MEV_i is straightforward. It posits that the variance of the measurement error of any expected return proxy rises with the proxy's volatility while declining with the correlation between the expected return proxy and future realized returns. When the expected return proxy closely aligns with future realized returns, the time-series variation in the expected return proxy is more likely to capture the fluctuations in the firm's true expected return.

[12]For a detailed derivation establishing the link between MEV and treatment effect, see Lee et al. (2020).

Offering an alternative mathematical perspective, the construction of MEV_i implies that if the volatility of the expected return proxy fails to translate into a higher covariance between the expected return proxy and the true future return (i.e., when the correlation is low or even negative), the measurement error for this proxy is higher. Also, note that MEV offers advantages in evaluating the performance of long-term ERP by design. It circumvents the use of augmented standard errors to address the mechanistic bias and avoids the inflated estimation of R-squared, both of which are consequences of ERP's persistence.

The third term, $var_i(er_{i,t})$, remains constant across different expected return measures. Therefore, our focus can be directed towards comparing the first two terms of different expected return measures and utilizing the scaled measurement error variance for assessment:

$$SMEV_i = var_i(\hat{er}_{i,t}) - 2cov_i(r_{i,t+1}, \hat{er}_{i,t}) = \sigma_i^2(\hat{er}_{i,t}) - 2corr_i(r_{i,t+1}, \hat{er}_{i,t})\sigma_i(\hat{er}_{i,t})\sigma_i(\hat{r}_{i,t}). \quad (9)$$

Table (8) presents the time-series variance of different expected return proxies, denoted as $Var(\hat{er}_t)$ [13], along with the covariance between expected return proxy and realized returns at various future horizons, $Cov(r_{t+h}, \hat{er}_t)$. Additionally, it displays the calculated scaled measurement error variance, $SMEV_i$, using Eq. (9). It is worth noting that for a constant expected return, equivalent to predicting future returns based on a mean-reversion process, the SMEV will be zero. As a result, an expected return measure with a more negative SMEV is considered a superior expected return proxy.

In the left panel, when using the tested variable to proxy the equal-weighted return, we observe that r_d and D/P exhibit negative SMEV values across all horizons. Particularly, for one-month-ahead $ewret$, r_d displays the most negative SMEV value, amounting to -0.0054. The ML forecast of EPS-based expected return, $Default$, and D/P can also obtain negative SMEV values. In contrast, analyst forecasts of EPS-based expected return measures exhibit positive SMEV values across all horizons, indicating the poor performance of these measures in the time-series dimension. When to proxy value-weighted returns, r_d again consistently exhibits the most negative SMEV. Among other predictors, only D/P and $Default$ manage to attain a negative SMEV, particularly when proxying

[13] Note that the variances of expected return proxies at different horizons are not exactly the same due to decreasing sample sizes with longer horizons. However, the magnitudes of the variances do not vary significantly.

returns at 6 and 12 months ahead.[14] Notably, although $r_{e,ML}$ obtains robust significant predictability of value-weighted return using the regression method as shown in Table **6**, it performs poorly when evaluated using the measurement error method.

In summary, the expected return proxy calculated from ML–based dividend forecasts stands out as the top-performing expected return measure, exhibiting the lowest time-series SMEV. This result remains robust across various horizons and applies to both equal-weighted and value-weighted returns. On the other hand, several other expected return proxies, including r_e, E/P, and *payout*, tend to display higher variances. However, this high variance does not translate into a stronger covariance with realized future returns. This result can be attributed to various factors. For instance, in the case of r_e, the volatility may stem from the noise in analyst forecasts of earnings and the challenges associated with accurately estimating the plowback rate (payout ratio). Our ML–based DPS forecast offers an advantage by directly bypassing the need to estimate the plowback rate, thereby reducing errors in the expected return proxy.

5.5. Testing our expected return: Portfolio analysis

All the tests discussed above are conducted at the aggregate market level. Existing literature also underscores the importance of conducting similar tests at portfolio level (Li et al., 2013; Easton and Monahan, 2005; Guay et al., 2011; Lee et al., 2020). In this test, we initially rank all the stocks traded on the NYSE, NYSE American, and Nasdaq based on their size and book-to-market ratio each June from 2005 to 2021, as is Li et al. (2013). We then create three size categories (small, medium, and large) and three value categories (low, medium, and high) using the NYSE breakpoints for market capitalization and book-to-market ratio. Specifically, a stock is assigned to the small-size portfolio (S) if its firm capitalization is equal to or lower than the bottom 30% breakpoint, to the medium-size portfolio (M) if it falls within the middle 40%, and to the large size portfolio (L) if it falls within the top 30%. The assignment for the high (H), medium (M), and low (L) value portfolios follows the same methodology.

We then align these categorized samples with our DPS and EPS forecast samples, ensuring that

[14] *payout* displays a much larger variance and SMEV compared to the other predictors. This discrepancy arises from the substantially higher magnitude of *payout* originally, as evident in the summary statistics presented in Table (**5**).

we include only firms with three valid expected return estimations for r_d, $r_{e,AL}$, and $r_{e,ML}$. This process results in an average of approximately 200 to 600 firms for each size or value portfolio that contain valid values of all of the three expected returns. To conduct the predictive regression test, we calculate the portfolio expected returns and realized returns by applying value-weighting to the individual firm-level data.

Panel A (B) of Table (9) presents regression results and measurement error tests for portfolios categorized by value (size). In the majority of cases, we find that r_d exhibits statistically significant predictive power across all six portfolios and forecasting horizons, with the exception of the one-month ahead returns for the medium size, medium value, and six-month ahead small size portfolio return. In terms of predictive regression results, both r_e measures ($r_{e,AL}$ and $r_{e,ML}$) also exhibit an overall strong predictive power, even stronger than r_d in predicting portfolios grouped by size. However, when evaluate the expected return performance based on SMEV, we find that only r_d can consistently provide negative values for different portfolios at different horizons, indicate a better performance for ERP than a constant. For both $r_{e,Al}$ and $r_{e,ML}$, they only outperform a constant in certain cases. This result is consistent with the previous aggregate market return test shown in Table (8).

5.6. ***Horse race*** **between various expected return proxies**

In summary, based on the "horse race" between various expected return measures, we find that the ML –forecasted DPS– based expected return proxy shows strong in-sample predictive power for both equal- and value-weighted aggregate market returns across various forecasting horizons, outperforming other proxies. The in-sample measurement error variance analysis and the out-of-sample forecasting results further support this result. For robustness, we also examine the predictability of the aggregate market return using the CRSP index. This approach enables us to assess the performance of our expected return proxy without being constrained by the S&P 500 universe. Online Appendix D shows that our expected return measure remains the top performer in both in-sample and out-of-sample tests when we use the CRSP index.

These results suggest that using ML forecasts of dividends as an input in forward-looking ex-

pected return measure estimations is a useful approach that can potentially yield more accurate expected returns. We argue that this approach addresses two important limitations of the conventional earnings-based methods. Specifically, in the earnings-based methods, calculating dividends from earnings forecasts involves multiplying earnings forecasts by the payout ratio. However, payout ratios are often estimated based on past realized values, making them weak forward-looking proxies that potentially introduce bias and noise. Additionally, some studies (Pástor et al., 2008; Li et al., 2013; Hou et al., 2012) replace negative earnings forecasts and select constant payout ratios arbitrarily, further introducing bias. By contrast, our ML–based dividend forecast, which leverages a wider range of information, avoids these issues from imputation and reduces noise in the data.

6. Conclusion

Accurately forecasting dividends is essential for making informed decisions regarding capital allocation, risk management, and asset valuation. Extant literature relies on three prevalent methods for dividend forecasting: (i) historical dividends, (ii) the combination of earnings forecasts and payout ratios, and (iii) analysts' predictions. While these methods provide valuable perspective power on future dividends, the first two rely on assumptions that may not hold in many situations. Moreover, analysts often produce biased forecasts for companies with complex information structures.

In this study, we adopt an approach that integrates both the information generated from these measures and other relevant data not included in these measures into a single forecast generation process. We achieve this by employing a non-parametric tree-based ML approach to forecast dividends. We demonstrate that our ML algorithm produces more accurate forecast that outperforms alternative methods. Furthermore, we highlight that ML–generated dividend forecasts remain unaffected in more complex information structures, while analysts produce less accurate dividend forecasts for firms with complex structures. We next use our ML–generated dividend forecasts as input in the Dividend Discount Model to compute a forward-looking expected return, often referred to as the Implied Cost of Capital. We show that this new expected return proxy contains more predictive power about future realized returns compared to traditional alternatives. This superiority holds true both in-sample and out-of-sample.

Our findings have several practical implications, suggesting that the benefits of ML are more pronounced in complex information environments. As big data continues to grow,[15] investors should increasingly rely on ML for higher precision in forecasting economic measures. On the academic front, our study underscores the importance of exploring ML applications for expected return estimations, providing richer insights into the risk-return relationship within the asset pricing literature.

[15] McKinsey Global Institute: Big data: The next frontier for innovation, competition, and productivity.

References

Allen, F., Bernardo, A. E., Welch, I., 2000. A theory of dividends based on tax clienteles. The journal of finance 55, 2499–2536.

Allen, F., Michaely, R., 2003. Payout policy. Handbook of the Economics of Finance 1, 337–429.

Avramov, D., Cheng, S., Metzker, L., 2022. Machine learning versus economic restrictions: Evidence from stock return predictability. Management Science 69, 2587–2619.

Baker, M., Wurgler, J., 2004. Appearing and disappearing dividends: The link to catering incentives. Journal of Financial Economics 73, 271–288.

Baldauf, M., Mollner, J., 2020. High-frequency trading and market performance. Journal of Finance 75, 1495–1526.

Barinov, A., Park, S. S., Yıldızhan, Ç., 2022. Firm complexity and post-earnings announcement drift. Review of Accounting Studies pp. 1–53.

Bartov, E., Givoly, D., Hayn, C., 2002. The rewards to meeting or beating earnings expectations. Journal of accounting and economics 33, 173–204.

Begenau, J., Farboodi, M., Veldkamp, L., 2018. Big data in finance and the growth of large firms. Journal of Monetary Economics 97, 71–87.

Benamar, H., Foucault, T., Vega, C., 2021. Demand for information, uncertainty, and the response of us treasury securities to news. The Review of Financial Studies 34, 3403–3455.

Berg, T., Fuster, A., Puri, M., 2022. Fintech lending. Annual Review of Financial Economics 14, 187–207.

Bilinski, P., Bradshaw, M. T., 2022. Analyst dividend forecasts and their usefulness to investors. The Accounting Review 97, 75–104.

Blankespoor, E., deHaan, E., Zhu, C., 2018. Capital market effects of media synthesis and dissemination: Evidence from robo-journalism. Review of Accounting Studies 23, 1–36.

Bradshaw, M. T., Lock, B., Wang, X., Zhou, D., 2020. Soft information in the financial press and analyst revisions. Accounting Review 96, 107–132.

Brav, A., Graham, J. R., Harvey, C. R., Michaely, R., 2005. Payout policy in the 21st century$. Journal of Financial Economics p. 45.

Brown, L. D., Call, A. C., Clement, M. B., Sharp, N. Y., 2015. Inside the "black box" of sell-side financial analysts. Journal of Accounting Research 53, 1–47.

Bybee, L., Kelly, B. T., Manela, A., Xiu, D., 2020. The structure of economic news. Tech. rep., National Bureau of Economic Research.

Campbell, J. Y., Shiller, R. J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. The Review of Financial Studies 1, 195–228.

Campbell, J. Y., Thompson, S. B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? The Review of Financial Studies 21, 1509–1531.

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., Mullainathan, S., 2016. Productivity and selection of human capital with machine learning. American Economic Review 106, 124–127.

Chen, H., Cohen, L., Gurun, U., Lou, D., Malloy, C., 2020. Iq from ip: Simplifying search in portfolio choice. Journal of Financial Economics 138, 118 – 137.

Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of econometrics 138, 291–311.

De La O, R., Myers, S., 2021. Subjective Cash Flow and Discount Rate Expectations. The Journal of Finance 76, 1339–1387.

de Silva, T., Thesmar, D., 2021. Noise in expectations: Evidence from analyst forecasts. Working Paper 28963, National Bureau of Economic Research.

Dessaint, O., Foucault, T., Frésard, L., 2023. Does alternative data improve financial forecasting? the horizon effect. Journal of Finance Forthcoming.

Dugast, J., Foucault, T., 2018. Data abundance and asset price informativeness. Journal of Financial economics 130, 367–391.

Dugast, J., Foucault, T., 2023. Equilibrium data mining and data abundance. Journal of Finance Forthcoming.

Easton, P. D., Monahan, S. J., 2005. An evaluation of accounting-based measures of expected returns. The Accounting Review 80, 501–538.

Ely, K. M., Mande, V., 1996. The interdependent use of earnings and dividends in financial analysts' earnings forecasts. Contemporary Accounting Research 13, 435–456.

Fama, E. F., French, K. R., 1988. Dividend yields and expected stock returns. Journal of financial economics 22, 3–25.

Fama, E. F., French, K. R., 1998. Taxes, financing decisions, and firm value. The journal of Finance 53, 819–843.

Fama, E. F., French, K. R., 2001. Disappearing dividends: changing firm characteristics or lower propensity to pay? Journal of Financial Economics 60, 3–43.

Fama, E. F., MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of political economy 81, 607–636.

Farboodi, M., Matray, A., Veldkamp, L., Venkateswaran, V., 2022. Where has all the data gone? Review of Financial Studies 35, 3101–3138.

Farboodi, M., Mihet, R., Philippon, T., Veldkamp, L., 2019. Big data and firm dynamics. AEA Papers and Proceedings 109, 38–42.

Farboodi, M., Veldkamp, L., 2020. Long-run growth of financial data technology. The American Economic Review 110, pp. 2485–2523.

Farre-Mensa, J., Michaely, R., Schmalz, M., 2014. Payout policy. Annu. Rev. Financ. Econ. 6, 75–134.

Floyd, E., Li, N., Skinner, D. J., 2015. Payout policy through the financial crisis: The growth of repurchases and the resilience of dividends. Journal of Financial Economics 118, 299–316.

Francis, J., Philbrick, D., 1993. Analysts' decisions as products of a multi-task environment. Journal of Accounting Research 31, 216–230.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., Walther, A., 2022. Predictably unequal? the effects of machine learning on credit markets. The Journal of Finance 77, 5–47.

Gebhardt, W. R., Lee, C. M., Swaminathan, B., 2001. Toward an implied cost of capital. Journal of accounting research 39, 135–176.

Goldstein, I., Yang, L., 2015. Information diversity and complementarities in trading and information acquisition. Journal of Finance 70, 1723–1765.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. The Review of Financial Studies 33, 2223–2273.

Guay, W., Kothari, S., Shu, S., 2011. Properties of implied cost of capital using analysts' forecasts. Australian Journal of Management 36, 125–149.

Hoitash, R., Hoitash, U., 2018. Measuring accounting reporting complexity with xbrl. The Accounting Review 93, 259–287.

Hoitash, R., Hoitash, U., Yezegel, A., 2021. Can sell-side analysts' experience, expertise and qualifications help mitigate the adverse effects of accounting reporting complexity? Review of Quantitative Finance and Accounting 57, 859–897.

Hommel, N., Landier, A., Thesmar, D., 2023. Corporate valuation: An empirical comparison of discounting methods. Working Paper 30898, National Bureau of Economic Research.

Hou, K., van Dijk, M. A., Zhang, Y., 2012. The implied cost of capital: A new approach. Journal of Accounting and Economics 53, 504–526.

Huang, D., Kilic, M., 2019. Gold, platinum, and expected stock returns. Journal of Financial Economics 132, 50–75.

Huberman, G., Regev, T., 2001. Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. The Journal of Finance 56, 387–396.

Hulbert, M., 2020. Use AI for picking stocks? Not so fast. Wall Street Journal .

James, G., Witten, D., Hastie, T., Tibshirani, R., et al., 2013. An introduction to statistical learning, vol. 112. Springer.

Kane, A., Lee, Y. K., Marcus, A., 1984. Earnings and dividend announcements: is there a corroboration effect? The Journal of Finance 39, 1091–1099.

Kaplan, Z., Pérez-Cavazos, G., 2022. Investment as the opportunity cost of dividend signaling. The Accounting Review 97, 279–308.

Kelly, B. T., Malamud, S., Zhou, K., 2022. The virtue of complexity in return prediction. Working Paper 30217, National Bureau of Economic Research.

Kim, O., Verrecchia, R. E., 1994. Market liquidity and volume around earnings announcements. Journal of Accounting and Economics 17, 41–67.

Kostakis, A., Magdalinos, T., Stamatogiannis, M. P., 2015. Robust econometric inference for stock return predictability. The Review of Financial Studies 28, 1506–1553.

Leary, M. T., Michaely, R., 2011. Determinants of Dividend Smoothing: Empirical Evidence. Review of Financial Studies 24, 3197–3249.

Lee, C., Ng, D., Swaminathan, B., 2009. Testing international asset pricing models using implied costs of capital. The Journal of Financial and Quantitative Analysis 44, 307–335.

Lee, C., So, E., Wang, C., 2010. Evaluating implied cost of capital estimates. SSRN eLibrary 6, 51.

Lee, C., Watts, E. M., 2021. Tick size tolls: Can a trading slowdown improve earnings news discovery? Accounting Review 96, 373–401.

Lee, C. M. C., So, E. C., Wang, C. C. Y., 2020. Evaluating Firm-Level Expected-Return Proxies: Implications for Estimating Treatment Effects. The Review of Financial Studies 34, 1907–1951.

Lehavy, R., Li, F., Merkley, K., 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. The Accounting Review 86, 1087–1115.

Li, Y., Ng, D. T., Swaminathan, B., 2012. Predicting time-varying value premium using the implied cost of capital: Implications for countercyclical risk, mispricing and style investing. Available at SSRN 2082108.

Li, Y., Ng, D. T., Swaminathan, B., 2013. Predicting market returns using aggregate implied cost of capital. Journal of Financial Economics 110, 419–436.

Lintner, J., 1956. Distribution of incomes of corporations among dividends, retained earnings, and taxes. The American economic review 46, 97–113.

Loughran, T., McDonald, B., 2020. Measuring firm complexity. Journal of Financial and Quantitative Analysis pp. 1–55.

Michaely, R., Moin, A., 2022. Disappearing and reappearing dividends. Journal of Financial Economics 143, 207–226.

Miller, M. H., Rock, K., 1985. Dividend policy under asymmetric information. The Journal of Finance 40, 1031–1051.

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica: Journal of the Econometric Society pp. 703–708.

Pástor, L., Sinha, M., Swaminathan, B., 2008. Estimating the intertemporal risk–return tradeoff using the implied cost of capital. The Journal of Finance 63, 2859–2897.

Pedersen, L. H., 2022. Game on: Social networks and markets. Journal of Financial Economics 146, 1097–1119.

So, E. C., 2013. A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts? Journal of Financial Economics 108, 615–640.

Stambaugh, R. F., 1999. Predictive regressions. Journal of financial economics 54, 375–421.

van Binsbergen, J. H., Han, X., Lopez-Lira, A., 2022. Man versus Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases. The Review of Financial Studies 36, 2361–2396.

Van Kervel, V., Menkveld, A. J., 2019. High-frequency trading around large institutional orders. Journal of Finance 74, 1091–1137.

Verrecchia, R. E., 1982. Information acquisition in a noisy rational expectations economy. Econometrica pp. 1415–1430.

Welch, I., Goyal, A., 2008. A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. Review of Financial Studies 21, 1455–1508.

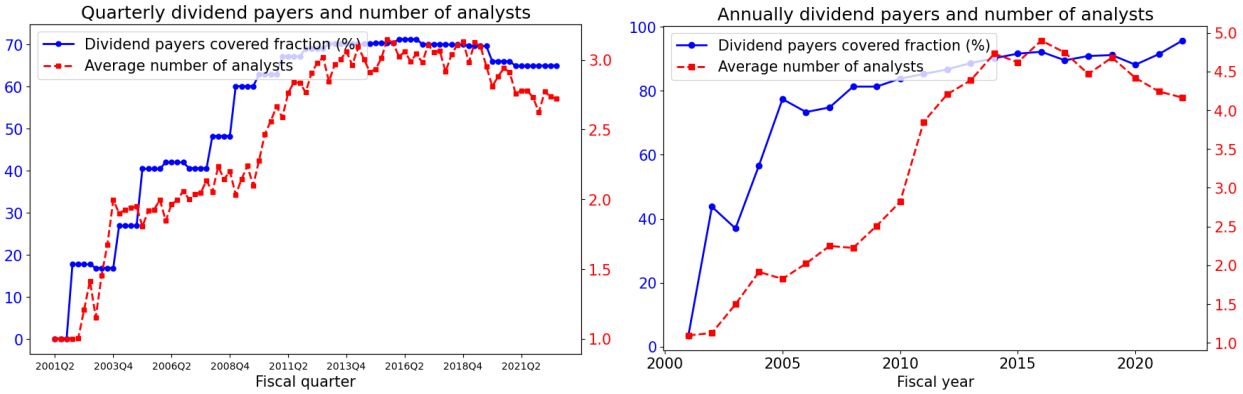
Weller, B. M., 2018. Does algorithmic trading reduce information acquisition? Review of Financial Studies 31, 2184–2226.

Yagan, D., 2015. Capital Tax Reform and the Real Economy: The Effects of the 2003 Dividend Tax Cut. American Economic Review 105, 3531–3563.

Zwiebel, J., 1996. Dynamic capital structure under managerial entrenchment. The American Economic Review 86, 1197–1215.

Figure 1: Analyst forecast of DPS coverage

This figure displays the proportion of quarterly (a) and annually (b) dividend payers covered by analyst forecasts for the US companies from 2001 to 2022. The blue line with circular markers represents the dividend payers covered fraction, while the red line with square markers represents the average number of analysts report for each firm-level forecast. The data of actual and analyst forecast of fiscal annual and fiscal quarter dividend is from I/B/E/S.



(a) Quarterly analyst forecast coverage

(b) Annual analyst forecast coverage

Figure 2: Dividend payout variation over time

This figure display the percentage of fraction of firms that change dividend payment compared to the previous level (left blue line), and the median of absolute deviation from target payment (right red line) from 1970 to 2021. The deviation is computed as suggested by Leary and Michaely (2011): $dev_{it} = TPR_i * E_{i,t} - D_{i,t-1}$. TPR_i stands for target payout ratio, estimated as the median of 10-year payout ratio. $E_{i,t}$ and $D_{i,t-1}$ represent earnings and dividend, respectively. The data are from Compustat annual fundamentals. Financial firms (SIC codes 6000-6999) and non-dividend-paying firms are excluded.

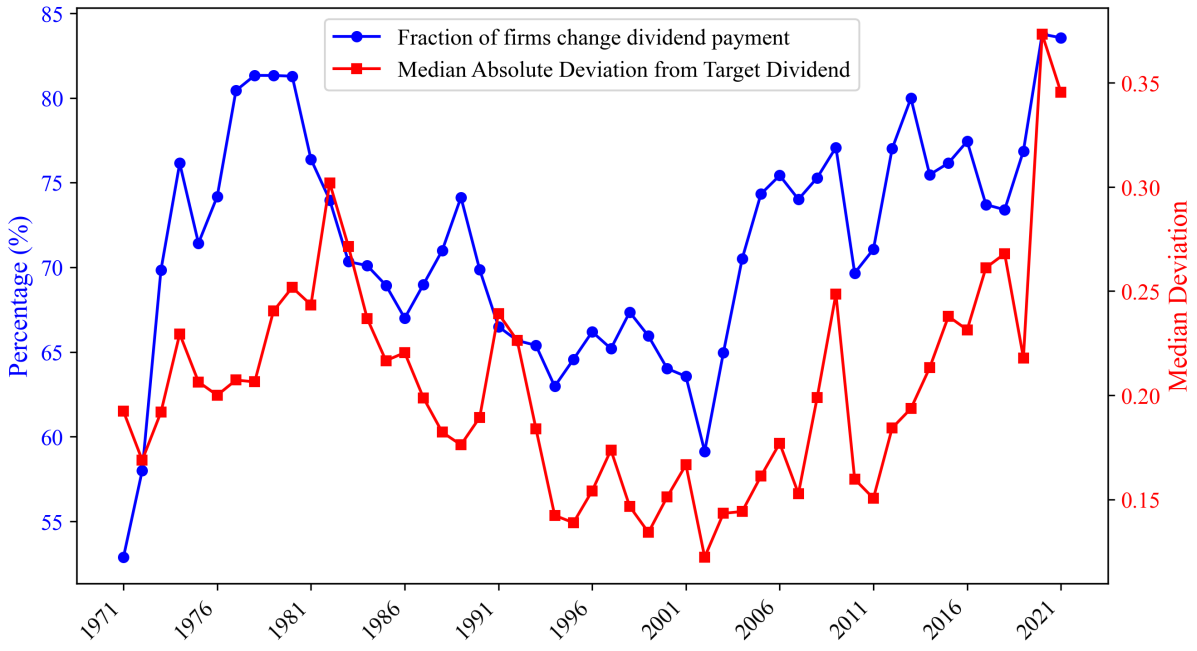


Figure 3: Estimation results of two-stage partial adjustment model

This figure plots the median of estimated SOA (circle mark line), dividend change's sensitivity to earnings (square mark line), and R-square (red dash line) from 1970 to 2021. The estimation using 10-year-rolling window at firm-level and using the two-stage method proposed by Leary and Michaely (2011): $\Delta D_{i,t} = \alpha + \beta dev_{i,t-1} + \varepsilon_{i,t}$, where $dev_{i,t-1} = TPR_i * E_{i,t} - D_{i,t-1}$. SOA is the estimated coefficient $\hat{\beta}$, and sensitivity to earnings is reflected by $\hat{\beta} * TPR_i * E_{i,t}$. TPR_i is estimated at first-stage using 10-year-rolling medians.The data are from Compustat annual fundamentals. Financial firms (SIC codes 6000-6999) and non-dividend-paying firms are excluded.

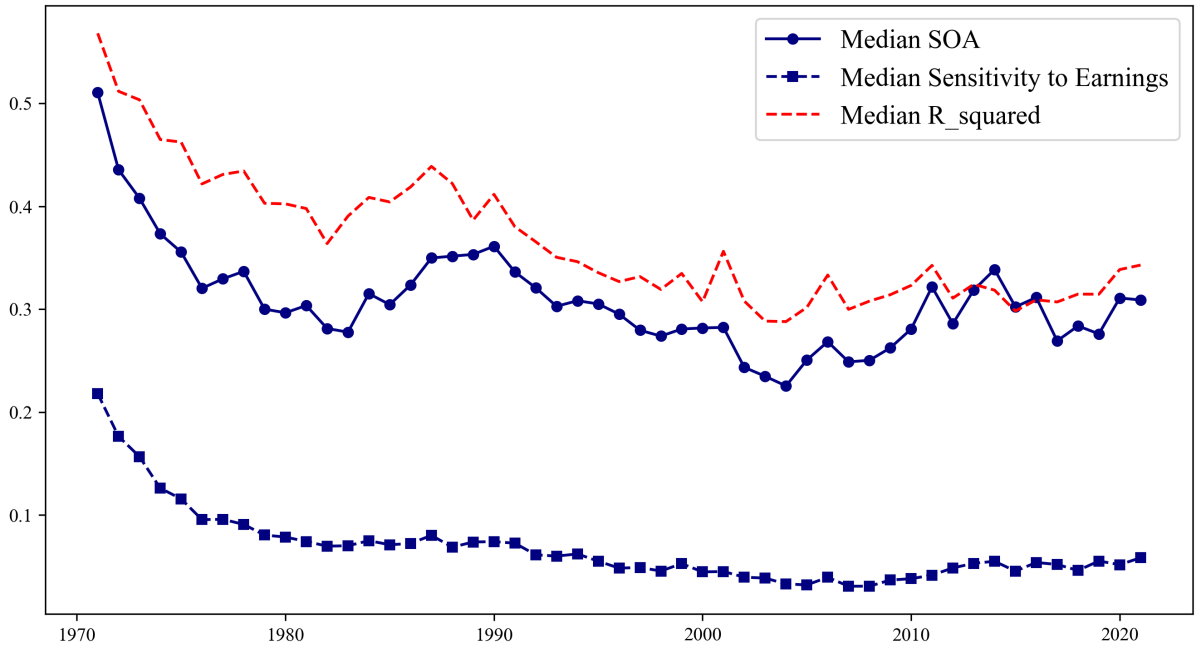


Figure 4: Flowchart of three tree-based regression models

This flowchart illustrates the methodologies employed by three advanced regression algorithms: Random Forest (RF) Regression, Gradient Boosting (GB) Regression, and Extreme Gradient Boosting (XGB) Regression. It provides a visual representation of the key procedural steps involved in the model-building process, leading to the final prediction output.

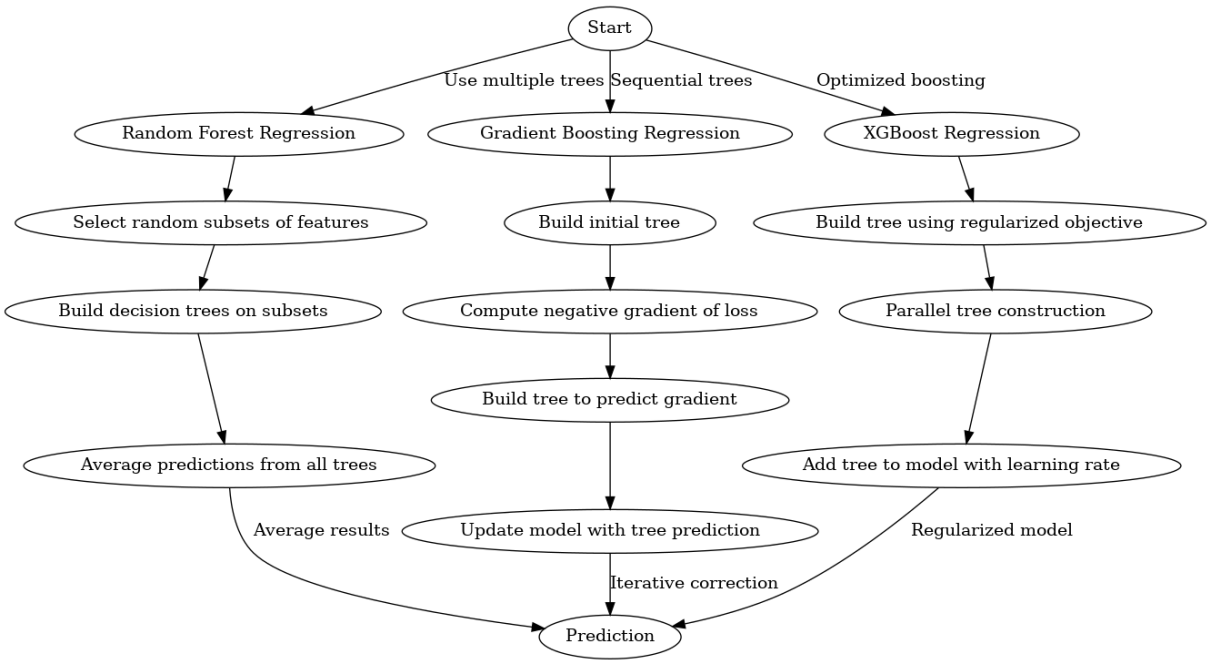


Figure 5: Example of single regression decision tree

The figure illustrates an example of a regression tree extracted from the Random Forest Regression model for one-year-ahead DPS forecasts. Each node in the tree contains the splitting condition, splitting criteria measured by Mean Squared Error (MSE), the number of observations, and the final output forecast value. The darker-colored nodes indicate areas of the tree with low impurity, signifying that the samples within these nodes have similar target values, resulting in more homogeneous predictions for these samples. AL_1y_DPS and AL_1y_EPS donate the analyst forecast of DPS and EPS, and L1Y_DPS represents the lagged actual DPS payout.

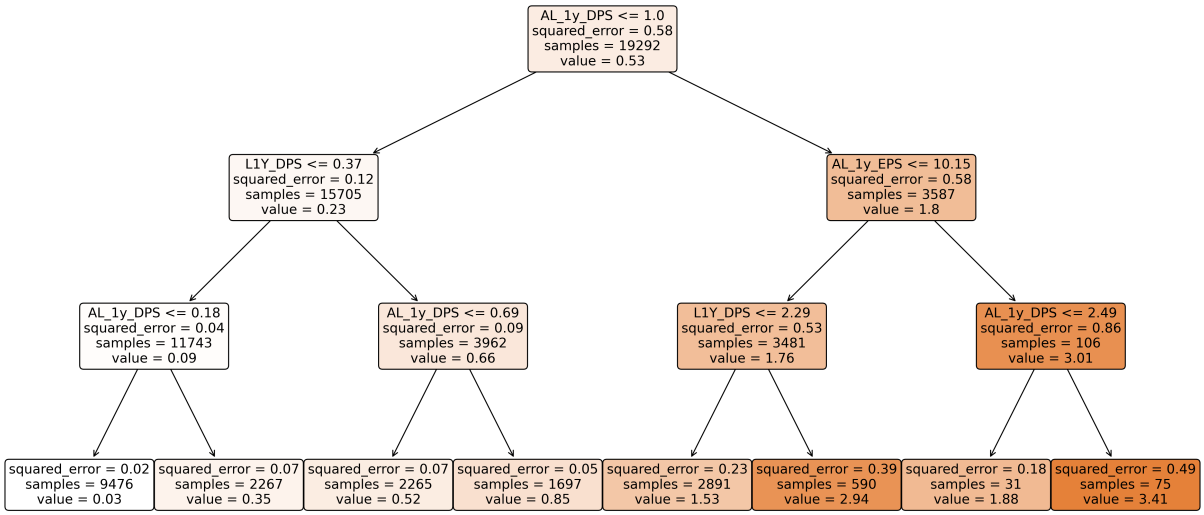


Figure 6: Importance from Random Forest Forecast

This figure showcases the top 10 most influential features in one, two, and three-years ahead forecasts(from left to right), as generated by Random Forest Forecast. The significance of each feature is measured by the reduction in impurity as the feature is incorporated into the forecasting model. All importance scores have been normalized to ensure a cumulative sum of one. The prefix 'AL' implies analyst forecasts, 'L1Y' denotes variables realized during the last fiscal year.

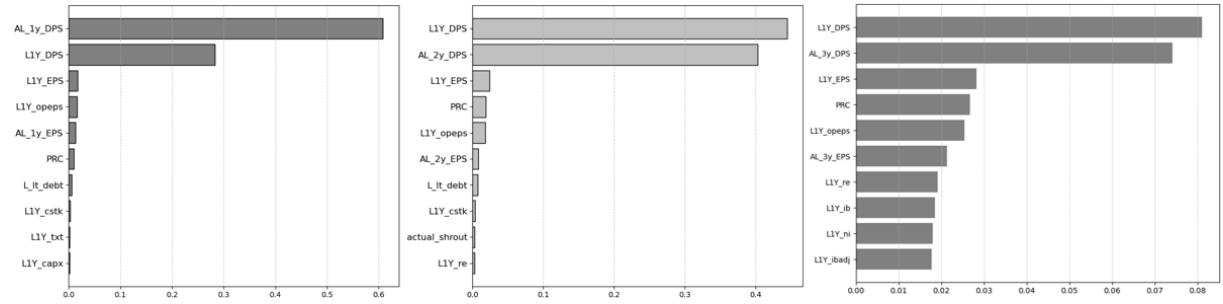


Figure 7: Nonlinearity between future actual DPS and key features

This figure presents Partial Dependence Plots (PDPs) derived from a two-year-ahead DPS forecast using RF regression. The forecasting model utilizes the standardized predictors described earlier. The time period covered in the analysis spans from 2003 to 2022.

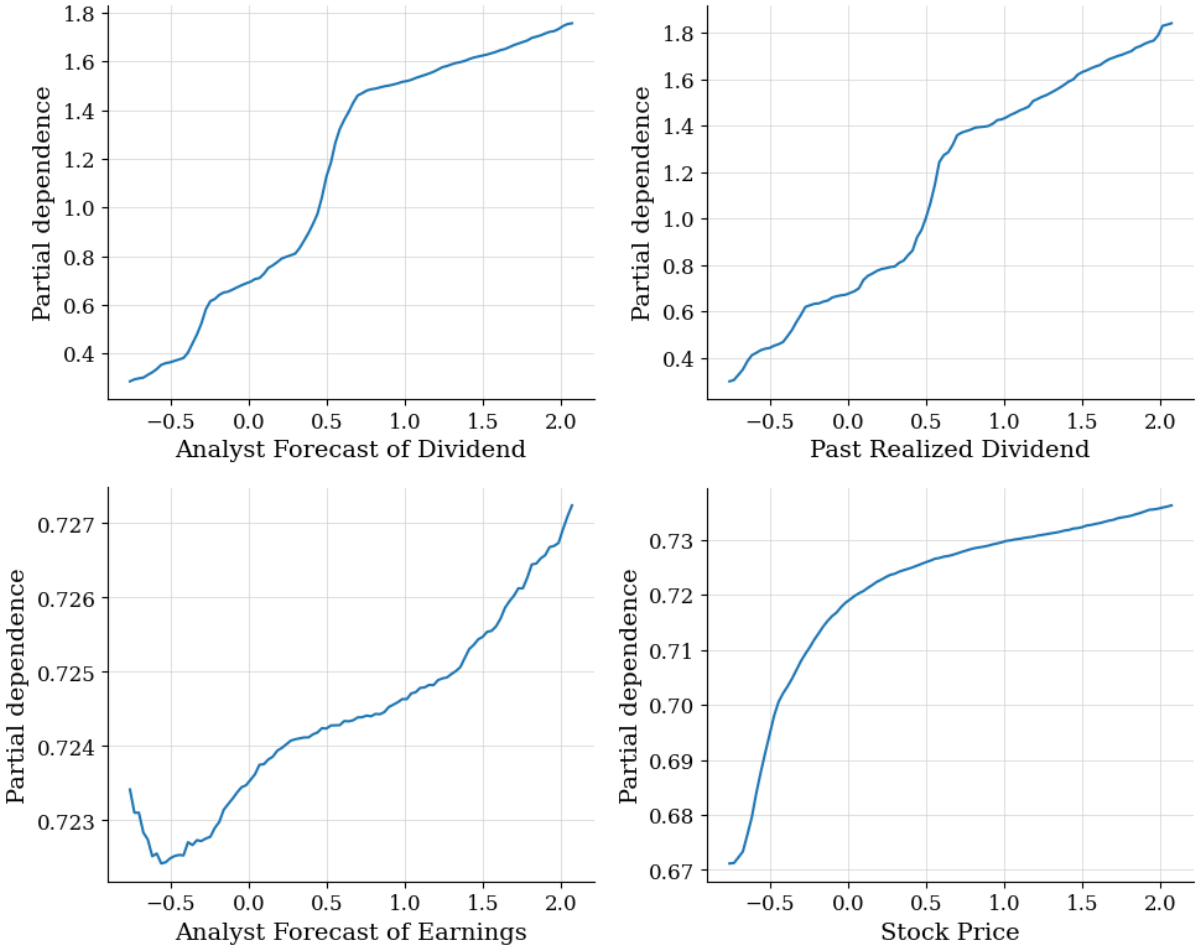


Figure 8: Time-series of the S&P 500 aggregate ICC

The figure plots the aggregate ICC computed as the value-weighted average of ICC values for all included S&P 500 constituents. The data covers the period from January 2005 to December 2021, with ICC expressed in an annualized format. These ICC values are derived from our machine learning forecasts of DPS, encompassing predictions up to three years ahead. The shaded regions on the graph correspond to NBER recession periods, which include the 2008 financial crisis and the 2020 COVID-19 outbreak.

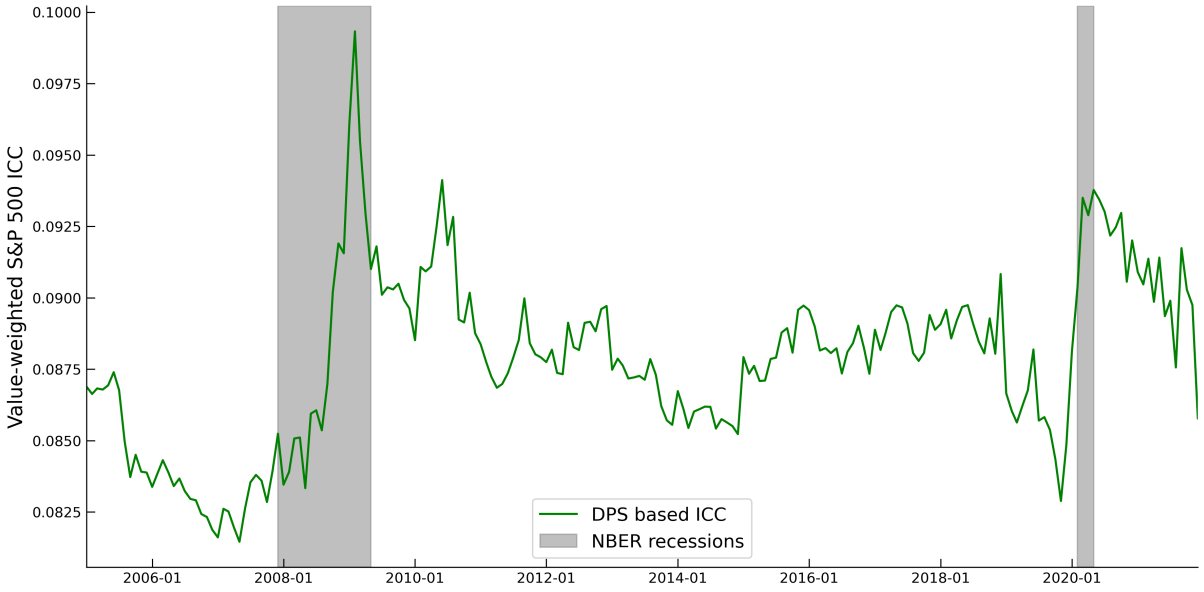


Figure 9: Difference of cumulative squared error between historical average and ERP

The figure displays the difference of cumulative forecast squared error of value-weighted S&P 500 return between the benchmark forecast using the historical average return and forecasts using the tested ERP, including: EPS and DPS-based ICC, default spread ($Default$), term spread ($Term$), dividend-price ratio (D/P), earnings-price ratio (E/P), and log payout ratio (py). The compare time period spans from January 2010 to December 2021. The out-of-sample forecasts are obtained with a 60-month rolling window method.

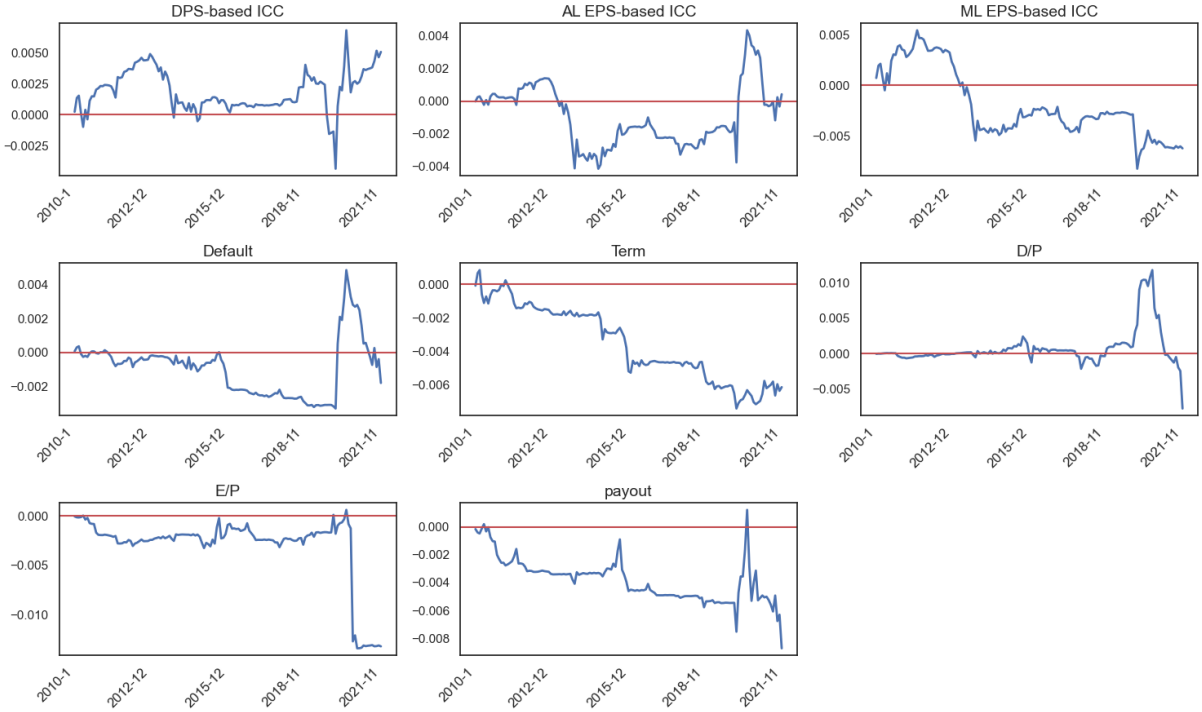


Table 1: Variables Utilized in Constructing Machine Learning Forecasts for DPS

The table presents an overview of the three information sets utilized in constructing the machine learning forecast. Analyst forecasts, realized DPS, and EPS data are obtained from I/B/E/S. Quarterly and annual firm fundamentals data are obtained from the WRDS CRSP/Compustat Merged database. Monthly stock price, return, and trading volume data are obtained from CRSP, while macroeconomic indicators are sourced from FRED.

Predictors	**Category**	**Description**
$X_{i,t}^{AL}$	Analyst forecasts	1. Analyst forecast of DPS and EPS. 2. Analyst forecast of long-run EPS growth rate averaged by industries.
$X_{i,t}^{Fin}$	Firm fundamentals	1. Realized last annually and quarterly DPS and EPS. 2. 58 quarterly financial fundamentals and calculated financial ratios. 3. Monthly stock price, return, and number of shares.
X_t^{Macro}	Macroeconomic indicators	GDP Growth, industrial production index (IPT) growth, consumption expenditure growth, and unemployment rate.

Table 2: Summary Statistics and Accuracy Evaluation of DPS Forecasts

The table presents summary statistics for actual DPS values and various DPS forecast results in the first panel, followed by accuracy evaluation in the second panel. The forecast period spans from January 2004 to December 2022. *Actual* represents the actual DPS values, while *AF* and $EPS * payout$ denote the analyst's DPS forecast and the analyst's forecast of EPS multiplied by the most recent year's payout ratio, respectively. *RF*, *GB*, and *XGB* correspond to forecasts generated using three tree-based machine learning regression methods: Random Forest, Gradient Boost Trees, and XGBoost Trees. *Comp* indicates the composite machine learning forecast, computed as the average of *RF* and *GB*. In the second panel, MSE represents the Mean Squared Error, calculated as the mean squared difference between actual and forecasted DPS values. % imp signifies the percentage reduction in MSE achieved by the forecasting methods compared to the naive forecast using the lagged one-period actual DPS. A negative sign ('-') indicates a reduction exceeding 100%.

Panel A: Summary statistics

	one-quarter		two-quarter		three-quarter		one-year		two-year		three-year	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Actual	0.204	0.248	0.210	0.254	0.216	0.260	0.678	0.945	0.730	1.014	0.815	1.111
AF	0.204	0.244	0.208	0.245	0.214	0.251	0.670	0.926	0.721	0.974	0.823	1.089
$EPS * payout$	0.175	0.248	0.187	0.263	0.197	0.279	0.613	0.955	0.693	1.024	0.839	1.249
RF	0.203	0.228	0.209	0.228	0.215	0.232	0.672	0.863	0.724	0.877	0.808	0.939
GB	0.203	0.226	0.209	0.230	0.215	0.234	0.674	0.858	0.725	0.883	0.810	0.946
XGB	0.202	0.230	0.209	0.232	0.215	0.236	0.675	0.879	0.726	0.897	0.814	0.968
Comp	0.203	0.227	0.209	0.228	0.215	0.233	0.673	0.859	0.725	0.878	0.809	0.939
Obs.	385847		372989		352092		534255		475311		316083	
N. of firms	5173		5095		4994		7213		6520		5047	

Panel B: Accuracy evaluation

	one-quarter		two-quarter		three-quarter		one-year		two-year		three-year	
	MSE	% imp	MSE	% imp	MSE	% imp	MSE	% imp	MSE	% imp	MSE	% imp
DPS_{t-1}	0.015		0.016		0.016		0.217		0.251		0.287	
AF	0.008	47.92	0.010	34.81	0.013	18.39	0.110	49.18	0.239	4.50	0.400	-39.20
$EPS * payout$	0.029	-89.53	0.038	-	0.044	-	0.387	-78.11	0.445	-77.65	0.769	-
RF	0.007	55.01	0.009	42.55	0.011	33.04	0.105	53.68	0.211	15.64	0.271	5.83
GB	0.007	52.53	0.009	40.42	0.011	31.04	0.107	50.80	0.214	14.64	0.277	3.44
XGB	0.007	52.06	0.010	40.24	0.011	30.26	0.113	48.03	0.225	10.33	0.287	0.05
Comp	0.007	54.81	0.009	42.45	0.011	33.01	0.106	51.38	0.210	16.30	0.268	6.58

Table 3: Firm Complexity and Forecast Errors

Panel A presents the time-series averages of squared errors for analyst forecasts and machine learning forecasts within five portfolios sorted by firm complexity. Firm complexity is quantified by ARC/100, representing one hundred units of monetary items disclosed. The Mean Squared Error (MSE) for each group is calculated as the value-weighted average of all individuals' three-year-ahead DPS forecast errors within the respective year. Panel B presents time-series average results for annual cross-sectional regression tests that examine the relationship between forecast errors and complexity. In Column (A), the regression includes ARC as the sole explanatory variable, while Column (B) introduces additional controls, encompassing firm fundamentals such as the logarithms of size and book-to-market ratio, return on assets (ROA), debt-equity ratio (d/e), dividend payout ratio (payout), number of employees (emp), and return volatility (ret_vol). Furthermore, two other firm complexity measures, $Conglo$ and $GeoMulti$, are included. $Conglo$ is a dummy variable that takes a value of one if a firm has multiple business segments in the same year, while $GeoMulti$ is one if a firm generates sales from multiple geographic segments in the same year. Ind_FE indicates the including of industry fixed effects. The t-statistics (in parentheses) are computed using the Fama-Macbeth method for the average slopes. The data spans from 2009 to 2021. *, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

Panel A: portfolio sort						
	1	2	3	4	5	C-S
AL_MSE	0.202	0.219	0.239	0.246	0.290	0.088*
t-stat						1.920
ML_MSE	0.143	0.133	0.153	0.171	0.175	0.032
t-stat						1.204

Panel B: Regression test				
	Analyst forecast		Machine learning forecast	
	(A)	(B)	(A)	(B)
ARC	0.039***	0.035***	0.004	-0.003
	(4.865)	(3.772)	(0.754)	(-0.507)
$Conglo$		-0.028		-0.024
		(-1.140)		(-1.323)
$GeoMulti$		-0.057**		-0.022
		(-2.095)		(-1.162)
$Size$		0.039***		0.033***
		(4.140)		(4.769)
B/M		-0.004		0.000
		(-0.293)		(-0.030)
ROA		0.324***		0.206***
		(10.29)		(7.921)
D/E		0.001***		0.000**
		(3.665)		(1.718)
$Payout$		0.004***		0.002
		(11.28)		(6.109)
EMP		0.000		0.000
		(-0.594)		(-0.166)
Ret_{vol}		0.364***		0.121
		(2.434)		(1.218)
$Intercept$	0.311***	-0.243*	0.280***	-0.177*
	(9.475)	(-1.899)	(11.005)	(-1.931)
Ind_FE	Y	Y	Y	Y
R^2	3.618%	5.217%	2.798%	4.342%

Table 4: ICC calculation approaches

The table presents a comparison between the ICC imputed from the conventional EPS-based method (denoted as r_e) and the DPS-based ICC obtained from a simplified Gordon growth model (denoted as r_d).

ICC	Assumption	Computation model
r_e	1. The growth of earnings after year 2 mean-reverts exponentially to GDP growth rate g. 2. The plowback rate follows the iterated process of $b_{t+k} = b_{t+k-1} - (b_1 - b)/T$. 3. In steady-state plowback rate b times the return on new investment equals the steady-state earnings growth rate.	$P_t = \sum_{k=1}^{T=15} \frac{E_t[EPS_{t+k}] \times (1 - b_{t+k})}{(1 + r_{e,t})^k} + \frac{E_t[EPS_{t+T}]}{r_{e,t}(1 + r_{e,t})^T}$
r_d	Dividend after year 2 will grow perpetually at the average GDP growth rate.	$P_t = \sum_{k=1}^{T=2} \frac{E_t[DPS_{t+k}]}{(1 + r_{d,t})^k} + \frac{E_t[DPS_{t+3}]}{(r_{d,t} - g)(1 + r_{d,t})^T}$

Table 5: Summary statistics of S&P 500 return and predictors

The table presents summary statistics of all variables entering our in-sample prediction tests covering the period from January 2005 to December 2021. The predictors include EPS (both $r_{e,AL}$ and $r_{e,ML}$), DPS-based ICC (r_d), calculated using both equal and value-weighted methods, default spread ($Default$), term spread ($Term$), dividend-price ratio (D/P), earnings-price ratio (E/P), and the natural logarithm of the payout ratio (py). All variables, except the payout, are expressed as annualized percentages. The table also includes columns for AR(1) and AR(12), indicating the autocorrelation coefficients at one and twelve lags, respectively. ADF represents the augmented Dickey-Fuller test statistic, and the corresponding p-value is provided.

Variable	Mean	Std. Dev.	AR(1)	AR(12)	ADF	p-Val
ew_{ret}	9.367	60.309	0.095	0.028	-12.867	0.000
ew_{r_d}	8.595	0.318	0.925	0.262	-3.325	0.014
$ew_{r_{e,AL}}$	9.511	1.126	0.961	0.558	-2.028	0.274
$ew_{r_{e,ML}}$	9.916	1.685	0.934	0.208	-2.619	0.089
vw_{ret}	9.014	50.852	0.078	0.036	-13.111	0.000
vw_{r_d}	8.643	0.300	0.925	0.317	-3.266	0.016
$vw_{r_{e,AL}}$	9.387	1.145	0.078	0.036	-1.950	0.309
$vw_{r_{e,ML}}$	9.784	1.697	0.929	0.172	-2.708	0.073
$Default$	12.648	5.459	0.953	0.081	-3.743	0.004
$Term$	14.656	10.270	0.985	0.682	-1.727	0.417
D/P	1.974	0.334	0.950	0.124	-3.305	0.015
E/P	4.758	1.295	0.982	0.315	-2.462	0.125
py	-0.362	0.210	0.980	0.083	-4.113	0.001

Table 6: Predicting future realized returns by using expected returns: univariate analysis

The table presents the results of predictive regressions for equal (left panel) and value-weighted (right panel) S&P 500 returns, covering the period from January 2005 to December 2021. The analysis is conducted at forecast horizons of 1, 6, and 12 months. We estimate the following regression:

$$\sum_{h=1}^{H}(r_{t+h})/H = \beta_0 + \beta_1 \times M_t + \epsilon_{t+h}$$

where r_{t+h} represents the continuously compounded monthly returns, either equal-weighted or value-weighted, in excess of the compounded three-month T-bill rate. M_t donates the vector of predictors, including EPS ($r_{e,AL}, r_{e,ML}$) and DPS-based ICC (r_d), computed using equal or value-weighted methods, default spread ($Default$), term spread ($Term$), dividend-price ratio (D/P), earnings-price ratio (E/P), and log payout ratio ($payout$). The NW-t column represent the Newey and West (1987) t-statistic, calculated with the number of lags equal to the forecast horizon length. The IVX-p column reports the p-value computed regarding to a Wald test of significance with the robust instrumentation approach developed by Kostakis et al. (2015).

		Equal-weighted				**Value-weighted**			
Variable	Horizon	β_1	NW-t	IVX-p	R-sq (%)	β_1	NW-t	IVX-p	R-sq (%)
r_d	1	3.178	2.071	0.011	4.04	2.887	2.391	0.006	4.18
	6	3.520	4.898	0.002	23.77	3.076	4.387	0.001	23.94
	12	3.215	5.287	0.002	41.11	2.749	3.741	0.001	37.55
$r_{e,AL}$	1	0.290	0.549	0.618	0.42	-0.012	-0.032	0.686	0.00
	6	0.405	1.181	0.370	4.02	0.020	0.066	0.768	0.01
	12	0.494	3.692	0.228	12.36	0.131	0.927	0.909	1.21
$r_{e,ML}$	1	0.533	2.161	0.017	3.21	0.441	2.480	0.014	3.12
	6	0.588	4.219	0.004	19.05	0.429	3.549	0.009	15.07
	12	0.478	3.754	0.009	26.49	0.322	2.761	0.029	16.77
$Default$	1	0.029	0.018	0.723	0.00	-0.381	-0.317	0.238	0.17
	6	1.021	1.359	0.379	4.13	0.421	0.648	0.992	1.08
	12	1.162	4.030	0.245	11.07	0.605	2.425	0.705	4.34
$Term$	1	0.283	0.706	0.649	0.23	0.101	0.304	0.939	0.04
	6	0.273	0.987	0.656	1.06	0.099	0.469	0.934	0.22
	12	0.339	1.549	0.525	3.46	0.164	0.924	0.894	1.18
D/P	1	1.106	0.526	0.601	0.53	0.389	0.245	0.778	0.09
	6	1.876	1.860	0.270	6.75	1.006	1.179	0.740	2.98
	12	2.102	5.419	0.208	16.21	1.286	3.843	0.543	8.77
E/P	1	-0.101	-0.250	0.836	0.07	-0.003	-0.011	0.786	0.00
	6	-0.281	-0.999	0.355	2.55	-0.162	-0.756	0.629	1.30
	12	-0.257	-1.379	0.381	4.34	-0.141	-1.039	0.671	1.90
py	1	0.022	0.776	0.331	0.85	0.011	0.494	0.831	0.28
	6	0.031	2.713	0.117	7.94	0.019	2.046	0.373	4.57
	12	0.025	3.983	0.168	11.11	0.015	3.050	0.473	5.58

Table 7: Out-of-sample predictability test

The table presents out-of-sample predictive test results for one-month-ahead S&P 500 returns, covering the period from January 2005 to December 2021. The forecast period spans from January 2010 (2015) for a 60 (120) months estimation window and ends at December 2021. Panel (A) and Panel (B) display results for equal-weighted and value-weighted returns, respectively. The left and right panels indicate an estimation strategy with a rolling window and expanding window, respectively. The tested expected return proxies include EPS ($r_{e,AL}, r_{e,ML}$) and DPS-based ICC (r_d), default spread ($Default$), term spread ($Term$), dividend-price ratio (D/P), earnings-price ratio (E/P), and log payout ratio ($payout$). The first column reports out-of-sample R-squared value R^2_{OS} calculated as $R^2_{OS} = 1 - \frac{\sum_{k=1}^{T-m}(r_{m+k} - \hat{r}_{m+k})^2}{\sum_{k=1}^{T-m}(r_{m+k} - \bar{r}_{m+k})^2}$. The p-val (second column) stands for the p-value from the one-side test using the adjusted Mean Squared Prediction Error statistic regress on a constant, calculated as $f_{t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - [(r_{t+1} - \hat{r}_{t+1})^2 - (\bar{r}_{t+1} - \hat{r}_{t+1})^2]$. We only report the p-val for the predictor with a positive R^2_{OS}. *, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

	Rolling window				Expanding window			
	60 months		120 months		60 months		120 months	
	$R^2_{OS}(\%)$	p-val	$R^2_{OS}(\%)$	p-val	$R^2_{OS}(\%)$	p-val	$R^2_{OS}(\%)$	p-val
Panel A: Equal-weighted								
r_d	1.26**	0.048	4.66**	0.017	4.79**	0.013	6.19**	0.020
$r_{e,AL}$	-0.14		1.36	0.121	0.41	0.233	0.38	0.261
$r_{e,ML}$	-3.59		-0.48		0.24	0.105	0.02	0.198
$Default$	-0.88		1.52*	0.091	0.04	0.410	-0.13	
$Term$	-2.86		-2.07		-0.34		-0.83	
D/P	-3.46		-6.17		0.04	0.340	-0.44	
E/P	-7.13		-2.33		-1.44		-0.14	
py	-4.34		-0.41		-0.98	0.224	0.72	0.163
Panel B: Value-weighted								
r_d	1.77**	0.038	3.91**	0.036	5.39***	0.007	5.66**	0.025
$r_{e,AL}$	-0.90		-0.70		-0.42		-0.52	
$r_{e,ML}$	-2.89		0.49	0.113	-0.45		0.56	0.140
$Default$	-0.75		1.61*	0.062	0.16	0.289	-0.30	
$Term$	-2.67		-1.97		-0.68		-0.89	
D/P	-3.23		-6.37		-0.01		-0.33	
E/P	-4.92		-1.87		-0.64		-0.86	
py	-3.55		-0.38		-0.58		0.33	0.223

Table 8: Evaluation of expected return proxy by variance of measurement errors

The table presents the measurement error tests for one, six, and twelve-month-ahead S&P 500 returns, covering the period from January 2005 to December 2021. The left and right panels display results for equal-weighted and value-weighted returns, respectively. The tested expected return proxies include EPS ($r_{e,AL}, r_{e,ML}$) and DPS-based ICC (r_d), default spread ($Default$), term spread ($Term$), dividend-price ratio (D/P), earnings-price ratio (E/P), and log payout ratio ($payout$). The first and second column, $Var(\hat{er}_t)$ and $Cov(r_{t+h}, \hat{er}_t)$,indicates time-series variance and covariance between ERP and realized return, respectively. The third column represents the calculated scaled measurement error variance $SMEV_i$ using: $SMEV_i = var_i(\hat{er}_{i,t}) - 2cov_i(r_{i,t+h}, \hat{er}_{i,t})$. All values are reported after multiplied by 100, as suggested by Lee et al. (2020).

		Equal-weighted			**Value-weighted**		
ERP	Horizon	$Var(\hat{er}_t)$	$Cov(r_{t+h}, \hat{er}_t)$	$SMEV_i$	$Var(\hat{er}_t)$	$Cov(r_{t+h}, \hat{er}_t)$	$SMEV_i$
r_d	1	0.0010	0.0032	-0.0054	0.0009	0.0026	-0.0043
	6	0.0010	0.0036	-0.0061	0.0009	0.0028	-0.0047
	12	0.0010	0.0033	-0.0056	0.0009	0.0026	-0.0042
$r_{e,AL}$	1	0.0127	0.0037	0.0053	0.0132	-0.0002	0.0135
	6	0.0130	0.0053	0.0025	0.0134	0.0003	0.0129
	12	0.0131	0.0065	0.0002	0.0132	0.0017	0.0098
$r_{e,ML}$	1	0.0285	0.0152	-0.0019	0.0289	0.0127	0.0034
	6	0.0292	0.0172	-0.0051	0.0297	0.0127	0.0042
	12	0.0300	0.0143	0.0013	0.0304	0.0098	0.0108
$Default$	1	0.0021	0.0001	0.0020	0.0021	-0.0010	0.0042
	6	0.0021	0.0021	-0.0022	0.0021	0.0007	0.0007
	12	0.0021	0.0025	-0.0028	0.0021	0.0012	-0.0002
$Term$	1	0.0074	0.0021	0.0032	0.0074	0.0007	0.0060
	6	0.0075	0.0021	0.0034	0.0075	0.0007	0.0062
	12	0.0078	0.0026	0.0025	0.0078	0.0012	0.0053
D/P	1	0.0011	0.0012	-0.0013	0.0011	0.0003	0.0006
	6	0.0010	0.0019	-0.0028	0.0010	0.0009	-0.0008
	12	0.0009	0.0020	-0.0030	0.0009	0.0012	-0.0014
E/P	1	0.0168	-0.0017	0.0202	0.0168	0.0004	0.0161
	6	0.0171	-0.0048	0.0267	0.0171	-0.0024	0.0219
	12	0.0169	-0.0043	0.0255	0.0169	-0.0020	0.0209
py	1	4.4096	0.0970	4.2157	4.4096	0.0375	4.3346
	6	4.4878	0.1371	4.2136	4.4878	0.0779	4.3320
	12	4.6172	0.1149	4.3874	4.6172	0.0631	4.4911

Table 9: Univariate predictive regression for value and size portfolios

The table reports results from (i) the univariate predictive regression test and (ii) time-series mean Scaled Measurement Errors Variance (SMEV) for three *value* sorted portfolios (Panel A): low (L), medium (M), and high (H), and three *size* sorted portfolios (Panel B): small (S), medium (M), and big (B). The period analysed spans from January 2005 to December 2021 and portfolios are formed in June of each year. The estimate the following regression:

$$\sum_{h=1}^{H}(r_{t+h})/H = \beta_0 + \beta_1 \times r_t + \epsilon_{t+h}$$

We focus on comparing the performance of three ICCs (r_d, $r_{e,AL}$, and $r_{e,ML}$), computed by value-weighting the individual constituents by capitalization. Columns 't-stat' represents Newey and West (1987) t-statistic with the number of lags equal to the forecast length; '$SMEV$' represents the calculated scaled measurement error variance $SMEV_i$ using: $SMEV_i = var_i(\hat{er}_{i,t}) - 2cov_i(r_{i,t+h}, \hat{er}_{i,t})$. The values are reported after multiplying 100 as suggested by Lee et al. (2020). *, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

		r_d			$r_{e,AL}$			$r_{e,ML}$		
	Horizon	β_1	t-stat	$SMEV$	β_1	t-stat	$SMEV$	β_1	t-stat	$SMEV$
Panel A: Value sorting										
	1	2.29***	2.97	-0.005	0.12	0.29	0.008	0.56***	3.02	-0.003
L	6	2.31***	3.13	-0.005	0.09	0.30	0.009	0.49***	4.00	0.000
	12	2.32***	3.01	-0.004	0.15	0.85	0.008	0.34***	3.08	0.010
	1	3.05**	2.04	-0.005	0.14	0.42	0.018	0.18	1.24	0.050
M	6	3.73***	5.87	-0.006	0.26	1.08	0.012	0.31***	3.97	0.030
	12	3.34***	5.03	-0.006	0.32***	2.86	0.009	0.28***	4.67	0.034
	1	1.64	1.33	-0.004	0.28	1.88*	0.032	0.05	0.29	0.174
H	6	1.91***	2.96	-0.005	0.38***	3.23	0.015	0.20**	2.49	0.100
	12	1.88***	3.65	-0.005	0.40***	3.79	0.013	0.25***	6.50	0.086
Panel B: Size sorting										
	1	2.53*	1.77	-0.004	0.37**	2.42	0.013	0.20*	1.95	0.118
S	6	1.74	1.36	-0.003	0.57***	2.76	-0.004	0.27***	3.00	0.067
	12	1.79**	1.97	-0.003	0.66***	4.04	-0.008	0.35***	6.91	0.033
	1	1.89	1.61	-0.003	0.35	1.25	0.006	0.29*	1.87	0.033
M	6	1.80*	1.89	-0.003	0.55**	2.23	-0.001	0.42***	4.56	0.010
	12	2.03**	2.41	-0.003	0.65***	3.98	-0.004	0.41***	7.23	0.010
	1	3.12***	3.05	-0.005	0.15	0.38	0.011	0.39**	2.22	0.009
B	6	3.21***	4.23	-0.005	0.18	0.69	0.010	0.42***	4.62	0.006
	12	2.95***	3.95	-0.005	0.24**	2.42	0.008	0.32***	3.49	0.015

Online Appendix for

"Dividend Forecasting in the Age of Machine Learning"

Appendix A. Gaining of forecast accuracy from more complex information structure: evidence from a MSE decomposition framework

In this appendix section, we present supplementary proofs for the benefits of more information bringing to the forecast accuracy via an extended MSE decomposition approach suggested by de Silva and Thesmar (2021). As a result, we derive and discuss the feasibility of the sufficient conditions for the forecast formed using more complex information structure to outperform the ones with limited information. This outcome is essential for the necessity of incorporating more information to construct dividend forecast instead of simply relying on parsimonious models.

Assume a two-period-economy that contains the current time, denoted by t, and the future or forecasted period that is denoted by $t + h$. At the current time, there are two information sources relevant to forecast the firm i's dividend: X_i, which is public information available to everyone, and P_i, the analyst's private information that is unobservable to the public. In general, the private information can originate either from analyst's personal source of information (e.g., private communication with management (Brown et al., 2015)), or generated by her own skills used in understanding the firm's payout policy (So, 2013). Since the dividend payout for each fiscal period is fundamentally determined by the firm's management broad, the input of private information is important for analysts to generate accurate forecasts. To simplify notations, we eliminate the time subscripts and state the first lemma 1:

Lemma 1 *For either linear or nonlinear information structure embedded in X_i and P_i, the true payout for firm i follows the true DGP without loss of generality*

$$y_i = x_i + p_i + \varepsilon_i. \tag{1}$$

where:

- $x_i = \mathbb{E}(y_i|X_i)$ *is the conditional forecast based on public available* X_i.
- $p_i = \mathbb{E}(y_i|X_i, P_i) - \mathbb{E}(y_i|X_i)$ *is the separate component forecast generated by private information set* P_i.
- ε_i *is the zero mean i.i.d. error term.*

We adopt the framework proposed by de Silva and Thesmar (2021), where the three parts are assumed to be orthogonal to each other. p_i is defined differentially to capture the variations in the forecast when additional private-information is introduced. Essentially, Eq. (3) signifies that the forecast of y_i relies on information from three distinct components: publicly available data x_i (e.g., earnings and historical dividends, as previously mentioned), private information p_i exclusive to analysts, and an *i.i.d.* error term ε_i.

For investors, they can generate dividend forecasts through two approaches: (a) based on past dividends, expected earnings, and the payout ratio according to the Lintner's model and, (b) by relying entirely on consensus analyst forecasts. With respect to the first approach, we do not specify the method practitioners use to arrive at their expected earnings. They may employ a firm characteristic model outlined by So (2013) or directly utilize analyst earnings forecasts. However, it is worth noting that we assume the entire set of this information to be publicly available and reflected in X_i.

Based on the second approach, which relies exclusively on the consensus analyst forecast, it is presumed to incorporate all accessible public and private information to produce the consensus div-

idend forecast. Nevertheless, the biases and inherent unpredictability in individual forecasts may not necessarily be fully mitigated when combined. This leads us to Lemma 2, which combines the structure of these two forecasting methods.

Lemma 2 *Dividend forecasts that are constructed solely rely on Lintner's model F_i^L or the analyst consensus* [1] *of dividend can be respectively decomposed as:*

$$F_i^L = x_i + \eta_i, \tag{2}$$

$$F_i^{AL} = x_i + p_i + b_i, \tag{3}$$

where:

- *η_i is the desperation generated by only utilizing a subset of public information in Lintner's forecast.*
- *$b_i = \mathbb{E}(F_i^{AL} - y_i | X_i, P_i)$ is the component distorted by analyst subjective bias.*

For the first Lintner's forecast, denoted as F_i^L, we introduce a random noise term η_i to account for potential distortions arising from utilizing a limited information set—namely, historical dividends and expected earnings. In the case of the analyst forecast, denoted as F_i^{AL}, we include a bias term b_i to characterize the component responsible for deviations between the analyst forecast and the true values, capturing the analyst forecast error. This structural framework implies that, despite analysts having access to both public and private information, inherent biases may cause their expectations to deviate from the true values. In practice, such biases can stem from various factors, including analysts' subjective reactions to external shocks, a tendency to maintain consistent forecast levels (as discussed by So (2013) and Hilary and Hsu (2013)), and the complexity of firms' information environments (Brown et al., 2015; Hoitash et al., 2021).

[1] To simplify, we examine the forecasting structure at an aggregate level, focusing on a representative investor and the analyst consensus, rather than delving into individual forecasts.

As argued by de Silva and Thesmar (2021), the framework's flexibility has a notable strength: it does not adhere to the concept of full-information rational expectations but rather accommodates deviations arising from both private information and biases. We can then formulate a forecast approach that incorporates both publicly available information and analyst forecasts following their announcements. At the end of each forecast period, two dividend forecasts become available: one is based on all historical public information X_t, while the other relies on the consensus analyst forecast. The combined forecast is derived by assigning weights to each component, as outlined in Lemma 3.

Lemma 3 *Assume the consensus forecast of dividend and all public information is available, the combined forecast (donated as F_i^{Com}) can be regarded to be formed by choosing a weight β where $0 \leq \beta \leq 1$.*

$$F_i^{Com} = (1-\beta)x_i + \beta(x_i + p_i + b_i) = x_i + \beta(p_i + b_i). \tag{4}$$

Under Bayesian expectations, the optimal weight is determined by the relative precision (inverse of variance) for each component of the forecast and can be calculated when the distribution of each forecast is observable (Chen and Jiang, 2006; So, 2013). However, our goal is to describe a more general framework for optimal forecasting that goes beyond the constraints of the Bayesian approach. In this broader context, we do not fix the value of the weight parameter β. Instead, our aim is to investigate how variations in this weight parameter can impact the overall Mean Squared Error (MSE) of the combined forecast. It is worth noting that when $\beta = 0$, the combined forecast corresponds to a special case of the econometric forecast as discussed in de Silva and Thesmar (2021), which relies exclusively on publicly accessible information X_i.

Based on the Eq. (2), (3), and (4), we can calculate the Mean Squared Error (MSE) for the forecasts of all the firms' future dividends using the three strategies: F^L, F^{AL}, and F^{Com}, as the evaluation of overall forecast accuracy. The cross-sectional MSE is defined as: $MSE = E[(F_i - y_i)^2]$. We assume

that the chosen weight β is neither 0 nor 1, meaning that the combined forecast does not solely rely on either analyst forecasts or general forecasts based on public information.

We then formally introduce Assumption 1 for simplify the computation:

Assumption 1 *All the conditional forecast component* $[x_i, p_i, \varepsilon_i, \eta_i, b_i]$ *are orthogonal to each other, e.g.,* $E(p_i|x_i) = 0$. *The unpredictable residual* ε_i *has a mean of zero.*

Operating under the assumption of orthogonality, the Mean Squared Error (MSE) for each forecast can be calculated as follows:

$$\begin{aligned}
MSE^L &= E[(\eta_i - p_i - \varepsilon_i)^2] \\
&= E(\eta_i^2) + E(p_i^2) - 2E[\eta_i]E[p_i] + var(\varepsilon_i) \\
MSE^{AL} &= E[(b_i - \varepsilon_i)^2] \\
&= E(b_i^2) + var(\varepsilon_i) \\
MSE^{Com} &= E[((\beta - 1)p_i + \beta b_i - \varepsilon_i)^2] \\
&= (\beta - 1)^2 E(p^2) + \beta^2 E(b_i^2) + var(\varepsilon_i) - 2\beta(1 - \beta)E(p_i)E(b_i)
\end{aligned}$$

With a large forecasting sample size, we assume that the private information component's expected value averages out across different companies. Put it simply, we acknowledge that private information can influence dividend forecasts for individual firms, such as when analysts interact with managers to gain insights into their attitudes toward dividend payouts for the next fiscal period. However, when the sample size is fairly large, we assume that the collective effect of this private information on the forecast is distributed symmetrically. This assumption can be formally justified as:

Assumption 2 *With a large forecasting sample size and applying the central limit theorem, we assume that the expected value of the private information component averages out across different companies, such that* $E(p) = E[\mathbb{E}(y_i|X_i, P_i) - \mathbb{E}(y_i|X_i)] = 0$.

With $E(p) = 0$, the difference in MSE can be calculated as follows:

$$\begin{aligned} MSE^{Com} - MSE^{L} &= \beta(\beta-2)E(p_i^2) + \beta^2 E(b_i^2) - E(\eta_i^2) + 2\beta(\beta-1)E(p_i)E(b_i) + 2E(p_i)E(\eta_i) \\ &= \beta(\beta-2)E(p_i^2) + \beta^2 E(b_i^2) - E(\eta_i^2), \qquad \text{(i)} \\ MSE^{Com} - MSE^{AL} &= (1-\beta)^2 E(p_i^2) - (1-\beta^2)E(b_i^2) + 2\beta(\beta-1)E(p_i)E(b_i) \\ &= (1-\beta)^2 E(p_i^2) - (1-\beta^2)E(b_i^2). \qquad \text{(ii)} \end{aligned}$$

To determine the solution for the weight parameter β that ensures the first difference (i) is less than zero, indicating that the combined forecast outperforms Lintner's forecast, we can equivalently seek solutions for the inequality $\beta(\beta-2)E(p_i^2) + \beta^2 E(b_i^2) < 0$. Since the expectation of a squared term, $E(\eta_i^2)$, is non-negative, solutions to this simplified inequality satisfy the original inequality (i). Additionally, considering that β ranges from 0 to 1, the sufficient condition for making (i) negative is:

$$\beta < \frac{2E(p_i^2)}{E(p_i^2) + E(b_i^2)}. \qquad \text{(a)}$$

The sufficient condition to make (ii) negative can also be determined as:

$$\beta > \frac{E(p_i^2) - E(b_i^2)}{E(p_i^2) + E(b_i^2)}. \qquad \text{(b)}$$

Let's consider the feasibility of the above two conditions. In the first condition, we can reasonably assume that for short-term horizon forecasts, such as one-quarter, two-quarters, and one-year-ahead

forecasts, private information is relatively reliable, and analysts do not have significant subjective bias. Assuming that all the second-moment terms in (a) and (b) have comparable magnitudes, $E(p_i^2)$ tends to be higher than $E(b_i^2)$ in short-term forecasts. Therefore, the right-hand side of (a) is higher than 1, which naturally satisfies the first condition since β has been set to be lower than 1.

Similarly, for longer horizon forecasts (two years ahead, or beyond), private information might be sparse, and subjective bias or noise tends to be more significant. In this case, the left-hand side of (b) tends to be negative, naturally satisfying the second condition since $\beta > 0$. Therefore, for short-term forecasts, the combined forecast tends to outperform Lintner's forecast due to the advantages of private information. As the forecast horizon extends, the combined forecast naturally outperforms the analyst forecast as analysts' subjective biases increase.

Overall, we extend the MSE decomposition framework introduced by de Silva and Thesmar (2021) to demonstrate how dividend forecast accuracy benefits from combining information from both public sources and analysts. We also explain why forecasts based solely on Lintner's model or analysts may have a large MSE. Importantly, our findings do not rely on strict assumptions regarding the form of analyst beliefs or the true DGP for dividends, and they do not necessarily suggest that Lintner's or analyst forecasts underperform in all situations. Rather, our goal is to demonstrate how dividend forecasts can potentially benefit from richer information. This perspective aligns with the findings of van Binsbergen et al. (2022), which suggest that incorporating additional features in the optimal estimator reduces errors because irrelevant information is automatically filtered out.

Appendix B. Details on machine learning forecast formation

Appendix B.1. Lists of input variables

Table B1 presents our data sources and the variables used to construct the machine learning forecast of DPS. We obtain quarterly financial ratios from WRDS through the CRSP/Compustat Merged quarterly fundamentals dataset. The calculation of financial ratios follows the method recommended by the Financial Ratios Suite available on WRDS. To ensure data availability at the forecast time points, all variables are based on one-quarter lagged values. We match the I/B/E/S actual file with the summary file containing analysts' consensus forecasts, using Ticker and forecast fiscal period as matching criteria. To facilitate the merging of I/B/E/S with other datasets, we use the link table provided by WRDS, ensuring that the firm's historical CUSIP matches in both I/B/E/S and CRSP datasets.

Table B1: **Variable Definitions and Categories**

Label	Definition	Category
AL_h_DPS	analyst forecast of h-period-ahead DPS	Analysis forecasts
AL_h_EPS	analyst forecast of h-period-ahead EPS	Analysis forecasts
L_h_DPS	past period of actual DPS	Historical actual
L_h_EPS	past period of actual EPS	Historical actual
prc	monthly price	Stock information
ret	monthly total return	Stock information
shrout	monthly shares outstanding	Stock information
de_ratio	Total Debt/Equity	Solvency
debt_assets	Total Debt/Total Assets	Solvency
capital_ratio	Capitalization Ratio	Capitalization
totdebt_invcap	Total Debt/Invested Capital	Capitalization
equity_invcap	Common Equity/Invested Capital	Capitalization
debt_invcap	Long-term Debt/Invested Capital	Capitalization
sale_invcap	Sales/Invested Capital	Efficiency
at_turn	Asset Turnover	Efficiency
sale_equity	Sales/Stockholders Equity	Efficiency
lt_debt	Long-term Debt/Total Liabilities	Financial Soundness
cash_lt	Cash Balance/Total Liabilities	Financial Soundness
npm	Net Profit Margin	Profitability
opmad	Operating Profit Margin After Depreciation	Profitability
aftret_equity	After-tax Return on Total Stockholders' Equity	Profitability
roa	Return on Assets	Profitability
roe	Return on Equity	Profitability
pe_op_basic	Price/Operating Earnings (Basic, Excl. EI)	Valuation
bm	book-to-market ratio	Valuation
ptb	Price/Book	Valuation
ps	Price/Sales	Valuation
retain_eps	retained earnings per share	Fundamentals
txtq	income taxes	Fundamentals
oancfy	Operating Activities - Net Cash Flow	Fundamentals
ivncfy	Investing Activities - Net Cash Flow	Fundamentals
fincfy	Financing Activities - Net Cash Flow	Fundamentals
chechy	Cash and Cash Equivalents Increase (Decrease)	Fundamentals
capxy	Capital Expenditures	Fundamentals
fopoy	Funds from Operations	Fundamentals
ppentq	Property Plant and Equipment	Fundamentals
txditcq	Deferred Taxes and Investment Tax Credit	Fundamentals
nopiq	Non-Operating Income (Expense)	Fundamentals
apq	Account Payable/Creditors	Fundamentals
xoprq	Operating Expense	Fundamentals
mibtq	Noncontrolling Interests	Fundamentals
capsq	Capital Surplus/Share Premium Reserve	Fundamentals

Table B1 *Continued*

Label	Definition	Category
xsgaq	Selling, General and Administrative Expenses	Fundamentals
dpq	Depreciation and Amortization	Fundamentals
size	Capitalization	Fundamentals
niq	Net Income	Fundamentals
ibq	Income Before Extraordinary Items	Fundamentals
cstkeq	Common Stock Equivalents	Fundamentals
ivchy	Increase in Investments	Fundamentals
sivy	Sale of Investments	Fundamentals
dltisy	Long-Term Debt - Issuance	Fundamentals
dltry	Long-Term Debt - Reduction	Fundamentals
opepsy	Earnings Per Share from Operations	Fundamentals
aqcy	Acquisitions	Fundamentals
tstkq	Treasury Stock	Fundamentals
dlcq	Debt in Current Liabilities	Fundamentals
revtq	Revenue	Fundamentals
capsq	Capital Surplus/Share Premium Reserve	Fundamentals
cstkq	Common/Ordinary Stock (Capital)	Fundamentals
rectq	Receivables	Fundamentals
invtq	Inventories	Fundamentals
pstkrq	Preferred/Preference Stock	Fundamentals
atq	total asset	Fundamentals
seqq	Stockholders Equity	Fundamentals
repurch	repurchases scaled by shares: (prstkc-pstkrv)/shares	Fundamentals
gdp	GDP gross	Macro economics
cg	consumption growth	Macro economics
indg	Growth of industrial production	Macro economics
unemp	Unemployment rate	Macro economics

Appendix B.2. Description of three tree-based techniques

We apply three ensemble tree-based regression methods: random forest (RF), gradient-boost (GB) trees, and extreme gradient-boosting (XGB) trees to make more accurate, robust and stable predictions.

In RF, the fundamental idea revolves around bootstrapping samples of the training dataset and constructing a decision tree for each of these samples. Specifically, given a training dataset D of size N, for each of the k trees to be grown, a sample of size N is drawn from D with replacement.

For each decision tree split, a random subset of m predictors (features) is selected from the total p predictors, and the best split on these m is used to split the node. The typical choice for m is $\sqrt{p}$ for regression problems. The final prediction of the RF regressor is an average of the predictions from all the individual decision trees. Mathematically, for a new data point x, the RF regression prediction $\hat{y}$ can be expressed as: $\hat{y}(x) = \frac{1}{k}\sum_{i=1}^{k} h_i(x)$. Where $h_i(x)$ represents the prediction of the i^{th} decision tree for the data point x.

GB focuses on minimizing the residuals left by the previous trees. The algorithm begins by initializing with a simple model, often the mean or another straightforward statistic, and computes the residuals. For each subsequent step, a decision tree is trained to predict the negative gradient (or, essentially, the residuals) of the loss function with respect to the previous cumulative prediction. The output for this tree is then scaled by a factor known as the learning rate and added to the previous predictions. Mathematically, given a loss function $L(y, F(x))$, where y is the true value and $F(x)$ is our predicted value, the negative gradient (or pseudo-residual) r at step m for every data point i can be expressed as: $r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$. A tree is then fit to these residuals, and the resulting predictions are scaled by a factor ν (the learning rate) and added to the previous model: $F_m(x) = F_{m-1}(x) + \nu\sum_{j=1}^{J}\gamma_{jm} I(x \in R_{jm})$ where J is the number of terminal nodes, R_{jm} is the region of the j^{th} terminal node for the m^{th} tree, and γ_{jm} is the value predicted for region R_{jm}.

XGB is a sophisticated optimization of the traditional gradient boosting framework. Distinct from GB, XGB aims to minimize a regularized objective function:

$$Obj(\Theta) = \sum_i l(y_i, \hat{y}_i) + \sum_j \Omega(f_j)$$

where $l(y_i, \hat{y}_i)$ denotes the training loss, capturing the discrepancy between the predicted $\hat{y}_i$ and actual y_i values, while $\Omega(f_j)$ signifies a regularization term, penalizing model complexity to mitigate

overfitting. This regularization integrates both L1 and L2 forms, offering a robust mechanism against potential overfitting. An important feature of XGB is its capability to handle missing data, intuitively imputing values during tree construction. Additionally, unlike traditional gradient boosting, which grows trees depth-first, XGB employs a depth-first strategy, subsequently pruning trees using the "max_depth" parameter, leading to optimized tree architectures. Integrated cross-validation during each boosting iteration facilitates pinpointing the optimal number of boosting rounds.

Appendix B.3. Forecast formation process for DPS

We construct our real-time forecasts for from one to three-quarters-ahead and one to three-years-ahead DPS using a rolling-window approach, split by time. This process involves six steps, starting from the original dataset:

1. The original dataset contains monthly firm-level values of predictor $X_{i,t}$, and includes realized future values of one and two-year-ahead DPS from 2003 to 2022.

2. We replace missing values in all variables, except for the realized DPS and the analyst forecast of DPS, with industry-time medians. Industries are classified according to the Fama-French 49 industry classification.

3. We create the first training dataset using the data from the year 2003. We winsorize all the variables at a 1% level and standardize all the predictors in the training set to have a mean of zero and a unit variance.

4. We set the year 2004 as the gap year for the one-year-ahead forecast and years 2004 and 2005 as the two gap years for the two-year-ahead forecast. We create the first test dataset using the data from the years 2005 (for the one-year-ahead forecast) and 2006 (for the two-year-ahead forecast). Similar to the training data, in the test data, we winsorize all the variables at a 1%

level and standardize all the predictors using the mean and variance obtained from the training dataset

5. We fit the model using the training set data with one of the three tree-based methods. We apply a 5-fold cross-validation to tune the hyperparameters for each algorithm at first and then choose the top 10 most important features to fit the forecast model.

6. We use the fitted model to generate forecasts based on three algorithms on the test set. Moving time forward, we repeat Steps 3-5 until the last test set within the year of 2022.

We implement the RF and GB three regressions, along with their cross-validation program, using the `scikit-learn` package in Python. The XGB regression program is also supported by a dedicated package `xgboost` in Python [2]. For the tree-based algorithms, we apply 5-fold cross-validation to select the optimal combination of hyperparameters from a specified range.

- **RF**. We employ a Grid Search methodology over a predefined parameter space supported by the `GridSearchCV` function. We set the number of trees, `n_estimators`, to be 1000. We choose the maximum depth of the trees, `max_depth`, between 4 and 8, controlling how deep each tree can grow and consequently its complexity. The minimum number of samples required to split an internal node, `min_samples_split`, is chosen from 2 and 8. We consider the proportion of features for the best split, `max_features`, from 0.5 (indicating half of the features) to 1 (using all features). The minimum number of samples required to be at a leaf node, `min_samples_leaf`, has values of 1 and 4.

- **GB**. We use the `GridSearchCV` function to choose parameters for the Gradient-Boost tree regression method. First, we select the number of trees, `n_estimators`, from 500 to 10,000. The maximum depth of the trees, `max_depth`, has possible values of 1 and 3. We consider the

[2] see https://xgboost.readthedocs.io/en/stable/python/python_intro.html for more information

learning_rate, which determines the step size at each iteration while moving towards a minimum of the loss function, with values of 0.001 and 0.1.

- **XGB**. The xgboost program in Python supports the GridSearchCV function. Since xgboost allows the calibration of a large number of hyperparameters, we select critical ones following Chen et al. (2024) to avoid overcomplexity. We choose the number of trees, n_estimators, from 100 to 500. The maximum depth of the trees, max_depth, takes values of 4 and 8. Learning_rate is selected from 0.001 and 0.1. min_child_weight, selected from values 1 or 3, affects the sensitivity of the model to specific data patterns. The portion of data used for building each tree, subsample, is either 80% or 100%. Finally, for the data sampling method, sampling_method, we experiment with both uniform and gradient-based approaches.

Appendix B.4. Computation hardware

For the tree-based regression methods, the rolling window prediction with 5-fold cross-validation is highly computationally intensive. To save computation time, we split the entire dataset by time and perform parallel computing using the cloud computing service provided by the Edinburgh Compute and Data Facility (ECDF [3]). Each estimation uses 20 cores, each with 4GB of RAM, resulting in a total of 40 CPUs, and takes around 8 hours.

[3] www.ecdf.ed.ac.uk

Appendix C. The EPS-based expected return calculations

In this section, we describe how we calculate the conventional EPS-based expected return measures, following the approach used by Pástor et al. (2008), Lee et al. (2009), and Li et al. (2013). We denote the EPS-based expected return as r_e. We obtain EPS expectations, $E_t[EPS_{t+k}]$, up to three years ahead from either machine learning forecasts or analyst forecasts of EPS. If the machine learning forecast of EPS is unavailable or negative, we replace it with the available and positive analyst forecast values. To generate machine learning forecasts of EPS, we follow the approach outlined by van Binsbergen et al. (2022). This approach employs a random forest regression model that considers predictors such as analyst forecasts, firm financial ratios from WRDS, and macroeconomic variables. This machine learning forecast serves as a statistical benchmark expectation, helping to isolate substantial subjective bias in analyst forecasts of earnings.

Similar to our DPS forecast, the machine learning forecast of EPS is generated using a rolling window approach and is designed to be out-of-sample. We select hyperparameters for the model based on the values suggested by van Binsbergen et al. (2022): `n_estimators` is set to be 2000, `max_depth` is 7, `min_samples_split` is 1%, and `min_samples_leaf` is 5.

For missing values and negative forecasts of EPS, we use the most recent available realized values, following the approach suggested by Li et al. (2013). For instance, if $E_t[EPS_{t+2}]$ is missing or negative and $E_t[EPS_{t+1}] > 0$, and the most recent realized $EPS_t > 0$, then $E_t[EPS_{t+2}]$ is replaced by $E_t[EPS_{t+1}] \times (E_t[EPS_{t+1}]/EPS_t)$. Similarly, the negative $E_t[EPS_{t+1}]$ is replaced by $EPS_t \times (1 + (E_t[EPS_{t+2}]/EPS_t - 1)^{1/2})$. If $E_t[EPS_{t+3}]$ is missing or negative, we use $E_t[EPS_{t+2}]$ times the analyst forecast of long-term EPS growth rate. Eventually, r_e is obtained by solving the following nonlinear

finite Gordon growth model:

$$P_t = \sum_{k=1}^{T=15} \frac{E_t[EPS_{t+k}] \times (1 - b_{t+k})}{(1 + r_{e,t})^k} + \frac{E_t[EPS_{t+k}]}{r_{e,t}(1 + r_{e,t})^T},$$

where b_{t+k} is the estimated plowback rate. The terminal value term is calculated using the no-growth perpetuity theory for the long run, with the formula $TV_{t+T} = E_t[EPS_{t+T+1}]/r_{ICC}$ (Damodaran, 2012). Plowback rate estimates that fall outside the range [0, 1] are adjusted to 0 and 1 accordingly.

To generate EPS forecast for $t + 4$ to $t + T + 1$, we assume that the earnings growth rate for year $t + 3$, $g_3 = E_t[EPS_{t+3}]/E_t[EPS_{t+2}] - 1$, exponentially reverts to a long-run steady value represented by the long-run nominal gross domestic product (GDP) growth rate. In cases where the growth rate g_3 is negative, we replace it with the year $t + 2$ earnings growth rate, denoted as g_2. Consequently, for $k = 4, ..., T + 1$, we calculate the earnings growth and earnings expectations as follows:

$$\begin{aligned} g_{t+k} &= g_{t+k-1} \times exp[log(g/g_3)/T], \\ E_t[EPS_{t+k}] &= E_t[EPS_{t+k-1}] \times (1 + g_{t+k}) \end{aligned}$$

The initial one-year ahead plowback rate, denoted as b_1, for year $t + 1$ is estimated as one minus the most recent year's total payout ratio, which is available from the Financial Ratios Suite by WRDS. We assume that the plowback rates revert linearly to a steady-state plowback rate, denoted as b, following the formula: $b_{t+k} = b_{t+k-1} - (b_1 - b)/T$. According to the sustainable growth rate formula, in the steady state, the earnings growth rate equals the return on new investment (ROI) multiplied by the plowback rate. We further assume that ROI is the same as the expected return for new investments. Therefore, the plowback rates for years $t + 3$ to $t + T$ can be computed as $b_{t+k} = b_{t+k-1} - (b_1 - b)/T$ for $k = 3, ..., T$.

Appendix D. Additional Predictive Regression Tests for expected return proxies

Appendix D.1. Predictability for CRSP Return

In this subsection, we perform a robustness check on our expected return by assessing its ability to predict the aggregate market return represented by the CRSP index. We use expected return proxies aggregated across the entire available sample, allowing us to evaluate the effectiveness of all the expected return measures without the restriction of the S&P 500 stock universe. The three samples, each containing ICC values ($r_d, r_{e,AL}, r_{e,ML}$) within the range of 0 to 0.3 suggested by Hommel et al. (2023), span from January 2005 to December 2021 and consist of approximately 500,000 firm-month observations. We construct the aggregate ERP by either equal-weighted or value-weighted methods to predict the corresponding CRSP NYSE/NYSE American/NASDAQ/ARCA returns obtained from WRDS. Table D1 summarizes the results of the predictive regression and also the scaled measurement errors variance ($SMEV$).

As TableD1 reports, We find that overall the tested ERP have better predictability to CRSP equal and value weighted returns, showing by more statistically significant coefficients and lower scaled measurement error variances. r_d has a weaker performance in predicting one-month ahead equal weighted return: its coefficient is no longer significant at even 10% level. In contrast, the performance of $r_{e,ML}$ has improved significantly measured by a lower and negative $SMEV$ for most cases. For a more comprehensive test, we continue to do the out-of-sample predictive test using different estimation methods.

Table D2 presents the results of the out-of-sample prediction tests for CRSP returns. Again, the overall performance of all ERP has gained significant improvement, indicating that these models provide a relatively better explanation for the dynamics of CRSP returns compared to those of the S&P 500. We find that r_d consistently maintains the strongest and most significant predictive power

Table D1: **Predictive regression for equal and value-weighted CRSP market return**

This table reports predictive regression results and the scaled measurement error variance in the prediction of CRSP market excess returns from January 2005 to December 2021. The left and right panels display results for equal-weighted and value-weighted returns, respectively. The columns $\beta 1$ display the estimated coefficient. columns NW-t show the Newey and West (1987) t-statistic with the number of lags equal to the forecast length. The IVX-p columns report the p-value computed regarding to a Wald test of significance with the robust instrumentation approach developed by Kostakis et al. (2015). The estimated predictive regression model is: $\sum_{h=1}^{H}(r_{t+h})/H = \beta_0 + \beta_1 \times M_t + \epsilon_{t+h}$.
where r_{t+h} represents h-periods-ahead compounded CRSP excess return. The vector of predictors M_t are defined the same as above tables. The fourth column $SMEV$ represents the scaled measurement error variance calculated by $SMEV_i = var_i(\hat{er}_{i,t}) - 2cov_i(r_{i,t+h}, \hat{er}_{i,t})$.

		Equal-weighted				Value-weighted			
variable	Horizon	β_1	NW-t	IVX-p	$SMEV$	β_1	NW-t	IVX-p	$SMEV$
r_d	1	2.192	1.567	0.13	-0.0046	2.813	2.193	0.02	-0.0042
	6	2.732	3.036	0.04	-0.0059	3.331	4.093	0.00	-0.0052
	12	2.803	4.880	0.04	-0.0057	2.889	3.699	0.00	-0.0046
$r_{e,AL}$	1	0.463	0.919	0.44	0.0008	0.039	0.082	0.87	0.0096
	6	0.733	1.930	0.13	-0.0046	0.090	0.251	1.00	0.0088
	12	0.740	3.370	0.11	-0.0048	0.195	1.186	0.74	0.0065
$r_{e,ML}$	1	0.637	2.343	0.01	-0.0066	0.502	2.429	0.01	-0.0001
	6	0.858	4.130	0.00	-0.0161	0.517	3.726	0.01	-0.0008
	12	0.757	4.593	0.00	-0.0111	0.382	3.066	0.02	0.0060
$Default$	1	0.909	0.574	0.49	-0.0017	-0.286	-0.225	0.43	0.0033
	6	1.735	2.342	0.08	-0.0052	0.534	0.773	0.73	-0.0001
	12	1.637	6.316	0.08	-0.0048	0.703	2.597	0.52	-0.0009
$Term$	1	0.211	0.485	0.73	0.0043	0.068	0.193	0.94	0.0064
	6	0.183	0.547	0.76	0.0048	0.057	0.250	0.91	0.0067
	12	0.252	0.844	0.63	0.0039	0.130	0.705	0.92	0.0057
D/P	1	2.043	0.976	0.19	-0.0034	0.485	0.291	0.98	0.0000
	6	2.735	2.673	0.08	-0.0045	1.139	1.276	0.56	-0.0013
	12	2.685	6.334	0.09	-0.0041	1.405	4.246	0.43	-0.0017
E/P	1	-0.317	-0.761	0.33	0.0275	-0.011	-0.034	0.91	0.0172
	6	-0.453	-1.447	0.13	0.0326	-0.172	-0.742	0.55	0.0230
	12	-0.417	-2.106	0.15	0.0309	-0.161	-1.126	0.56	0.0223
$payout$	1	0.040	1.458	0.05	4.0607	0.012	0.546	0.63	4.3011
	6	0.043	3.407	0.02	4.1045	0.020	2.033	0.28	4.3055
	12	0.034	4.684	0.05	4.3029	0.016	3.413	0.35	4.4654

for CRSP returns in all cases. Interestingly, the EPS-based expected return proxy, derived from the machine learning forecast of EPS, also demonstrates several positive and significant predictive per-

formance with when using a 60-month expanding window estimation and 120-months rolling window estimation. However, its performance is highly dependent on the choice of estimation method. For instance, when predicting the equal-weighted return using a 60-month rolling window estimation, $r_{e,ML}$ exhibits a notably negative out-of-sample R^2 (specifically, -6.7%), which undermines its robustness compared to r_d. Additionally, the business cycle variable default spread keep exhibiting significant and positive R^2_{OS} for several cases.

In summary, our predictive tests, both in-sample and out-of-sample, for CRSP returns affirm the superiority of our expected return measure. While $r_{e,ML}$ demonstrates commendable performance in in-sample prediction and measurement error variance, its performance in out-of-sample scenarios lacks consistent robustness. Furthermore, a notable limitation of using CRSP returns to test the expected return generated from the entire sample is that the results are significantly influenced by the participation of individual stocks and the overall sample size. Given that the composition of the S&P 500 is more controlled and changes less frequently over time, we advocate using the aggregate market return represented by the S&P 500 return, as our primary test benchmark.

Table D2: **Out-of-sample predictive regression test for CRSP market return**

This table presents out-of-sample predictive test results for one-month-ahead CRSP excess returns, covering the period from January 2005 to December 2021. The forecast period spans from January 2010 (2015) for a 60 (120) months estimation window and ends at December 2021. The top and bottom panels display results for equal-weighted and value-weighted returns, respectively. The left and right panels indicate an estimation strategy with a rolling window and expanding window, respectively. Definitions of the tested ERP are the same as in the above tables. The first column reports out-of-sample R square R^2_{OS} calculated by: $R^2_{OS} = 1 - \frac{\sum_{k=1}^{T-m}(r_{m+k} - \hat{r}_{m+k})^2}{\sum_{k=1}^{T-m}(r_{m+k} - \bar{r}_{m+k})^2}$. The second column p-val stands for the p-value from the one-side test results using the adjusted-MSPE statistic regress on a constant, calculated as: $f_{t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - [(r_{t+1} - \hat{r}_{t+1})^2 - (\bar{r}_{t+1} - \hat{r}_{t+1})^2]$. We only report the p-val for the predictor with a positive R^2_{OS}. *, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

Equal-weighted	Rolling window				Expanding window			
	60 months		120 months		60 months		120 months	
	$R^2_{OS}(\%)$	p-val	$R^2_{OS}(\%)$	p-val	$R^2_{OS}(\%)$	p-val	$R^2_{OS}(\%)$	p-val
r_d	0.80**	0.046	4.30	0.040**	2.89**	0.036	4.07**	0.040
$r_{e,AL}$	−5.03		0.32	0.199	0.79	0.154	0.60	0.222
$r_{e,ML}$	−6.70		0.14	0.177	1.88*	0.075	0.86	0.158
$Default$	1.40	0.142	3.45**	0.015	0.50	0.132	0.81	0.106
$Term$	−4.44		−1.62		−0.65		−0.85	
D/P	−2.03		−3.14		0.77	0.135	0.52	0.212
E/P	−7.72		−2.15		−1.75		0.55	0.291
py	−0.65		2.60	0.082	−0.08	0.358	2.38**	0.031
Value-weighted	Rolling window				Expanding window			
	60 months		120 months		60 months		120 months	
	$R^2_{OS}(\%)$	p-val	$R^2_{OS}(\%)$	p-val	$R^2_{OS}(\%)$	p-val	$R^2_{OS}(\%)$	p-val
r_d	0.97*	0.063	2.86*	0.069	4.14**	0.028	4.67*	0.063
$r_{e,AL}$	−0.84		0.26	0.287	−0.33		−0.24	
$r_{e,ML}$	−2.53		0.90*	0.087	1.04*	0.063	1.31	0.107
$Default$	−0.31		2.38**	0.038	−0.14		−0.31	
$Term$	−3.11		−1.61		−0.85		−0.54	
D/P	−2.70		−5.61		0.13	0.332	−0.16	0.467
E/P	−6.31		−1.89		−0.73		−0.49	
py	−2.90		0.10	0.353	−0.36		0.55	0.174

References

Brown, L. D., Call, A. C., Clement, M. B., and Sharp, N. Y. (2015). Inside the "black box" of sell-side financial analysts. *Journal of Accounting Research*, 53(1):1–47.

Chen, Q. and Jiang, W. (2006). Analysts' weighting of private and public information. *The Review of financial studies*, 19(1):319–355.

Chen, Y., Calabrese, R., and Martin-Barragan, B. (2024). Interpretable machine learning for imbalanced credit scoring datasets. *European Journal of Operational Research*, 312(1):357–372.

Damodaran, A. (2012). *Investment valuation: Tools and techniques for determining the value of any asset*, volume 666. John Wiley & Sons.

de Silva, T. and Thesmar, D. (2021). Noise in expectations: Evidence from analyst forecasts. Working Paper 28963, National Bureau of Economic Research.

Hilary, G. and Hsu, C. (2013). Analyst forecast consistency. *the Journal of Finance*, 68(1):271–297.

Hoitash, R., Hoitash, U., and Yezegel, A. (2021). Can sell-side analysts' experience, expertise and qualifications help mitigate the adverse effects of accounting reporting complexity? *Review of Quantitative Finance and Accounting*, 57:859–897.

Hommel, N., Landier, A., and Thesmar, D. (2023). Corporate valuation: An empirical comparison of discounting methods. Working Paper 30898, National Bureau of Economic Research.

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, 28(5):1506–1553.

Lee, C., Ng, D., and Swaminathan, B. (2009). Testing international asset pricing models using implied costs of capital. *The Journal of Financial and Quantitative Analysis*, 44(2):307–335.

Li, Y., Ng, D. T., and Swaminathan, B. (2013). Predicting market returns using aggregate implied cost of capital. *Journal of Financial Economics*, 110(2):419–436.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society*, pages 703–708.

Pástor, L., Sinha, M., and Swaminathan, B. (2008). Estimating the intertemporal risk–return tradeoff using the implied cost of capital. *The Journal of Finance*, 63(6):2859–2897.

So, E. C. (2013). A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts? *Journal of Financial Economics*, 108(3):615–640.

van Binsbergen, J. H., Han, X., and Lopez-Lira, A. (2022). Man versus Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases. *The Review of Financial Studies*, 36(6):2361–2396.