Accounting vs Market Information: What Matters More for Stock Return Predictability?

Abstract

We employ machine learning techniques to determine what matters more for stock return predictability: market data or accounting information. Market data holds an advantage—it consistently generates more accurate forecasts and higher portfolio returns. This superiority is remarkably noticeably robust, holding across firm size categories, recent periods, and international markets. In contrast, accounting information adds relatively little value. When incorporated into the forecasting models, it only modestly improves their performance. Finally, the effectiveness of market information originates from volatile periods and hard-to-value assets. Conversely, when valuation uncertainty is low, accounting-based models perform almost as well as their market-based counterparts.

Keywords: stock market, return predictability, the cross-section of stock returns, machine learning, accounting data, market data.

JEL Codes: C52, G10, G12, G15

1. Introduction

Technical and fundamental analysts are archetypal adversaries. Technicians argue that historical market data is the key to successful trading. Conversely, proponents of fundamental methods often disregard technical tools, viewing them as akin to modern alchemy. Instead, they emphasize the importance of accounting and financial metrics, necessary to determine a company's intrinsic value.

These two opposing philosophies fuel a polarized debate about market informational efficiency, which can take several forms. The null hypothesis of the weak form of market efficiency suggests that technical analysis cannot effectively exploit predictable return patterns (Fama, 1970). A broader semi-strong form implies that fundamental analysis is redundant as well. Yet, despite decades of empirical studies, a clear consensus remains elusive. The 2013 Nobel Prize, awarded jointly to Eugene Fama, a proponent of the efficient market hypothesis, and Robert Shiller, known for his skepticism of market efficiency, is perhaps the most vivid sign that the controversy is far from resolved. Some authors strongly support the technical signals (e.g., Han et al., 2013; Avramov et al., 2018; Brogaard & Zareei, 2023). Others remain critical (Sullivan et al., 1999; Bajgrowicz & Scaillet, 2012). In the context of accounting data, Bertram and Grunblatt (2018, 2021) and Yan and Zhang (2017) advocate for fundamental analysis, while Linnainmaa and Roberts (2018) argue that most accounting anomalies likely result from data snooping. This discourse prompts crucial questions: What is the true utility of accounting and market data? Which type of information offers superior return predictability? And do they subsume or complement each other?

In this study, we comprehensively revisit the cross-sectional predictability of stock returns by market and accounting information. To this end, we examine five decades of data from the U.S. market. We compute a comprehensive set of 131 prominent return predictors from the asset pricing literature, including 44 that rely solely on market data and 87 that also use accounting information. Next, we apply machine learning techniques to extract maximum information from the signal space. We take advantage of their unique ability to handle large amounts of data and variables, select the most important predictors, and reveal non-linearities and interactions among them. Specifically, we combine several popular algorithms, including ordinary and penalized regression, dimension reduction techniques, tree methods, and neural networks. Finally, we feed the models with different types of variables to capture the predictability of returns through different types of information.

Our results reveal a visible disparity between the usefulness of the two categories of information. As seen in Figure 1, market variables have a noticeable edge.. The out-of-sample predictive R^2 coefficient, providing an intuitive snapshot of cross-sectional return predictability, is nearly three times higher for market characteristics than for signals derived from accounting data. Put simply, market variables appear more informative about future returns than their accounting counterparts. Moreover, the combination of market and accounting data offers only a modest improvement; it is primarily the market data that matters.

[Insert Figure 1 here]

A variable importance analysis allows us to identify the critical characteristics driving the returns. The predictability by market variables is primarily influenced by indicators of liquidity and frictions, such as the number of zero trading days, short-term reversal, idiosyncratic risk, and market value. On the other hand, within the accounting category, measures of investment, valuation, and profitability take the lead. Notably, the superiority of market characteristics persists across all firm size classes. Even for the largest firms—typically assumed to be informationally efficient and correctly priced—the aggregate importance of market variables overshadows accounting signals.

The effectiveness of market information translates into higher profits from corresponding trading strategies. To illustrate this, we run decile portfolio sorts based on forecasts using different types of information. A long-short value-weighted portfolio that buys (sells) the stocks with the highest (lowest) expected return, as determined by accounting information, yields a monthly six-factor alpha of 0.85% per month. On the other hand, an equivalent strategy based solely on market data earns 1.20% per month. Moreover, combining the two types of information adds limited value, increasing alpha by only 14 basis points to 1.34%.

The preeminence of market characteristics seem to be robust. Similar return patterns persist after excluding low-priced stocks. Furthermore, our findings do not hinge on any particular machine learning model used to extract forecasts from characteristics; they hold across a range of linear and nonlinear algorithms. The results are also consistent in different subperiods. While overall return predictability decreases over time, strategies based on market data continue to outperform their accounting counterparts in the last two decades—even for the largest stocks. However, a closer look at different size categories reveals a noteworthy regularity: the superiority of market data is most pronounced for small and micro stocks. On the other hand, for large companies, strategies based on accounting data perform almost as well as their market-based counterparts—although a difference is still discernible.

Notably, the documented effect is not limited to the United States. Examining ten major developed markets, we observe similar—or even stronger—regularities. The alphas on portfolios based on market data are, on average, more than twice as high as those based on accounting strategies. Combining the two classes of information typically results in minimal, if any, improvement over using market data alone.

Having established the basic properties of return predictability by market and accounting data, we examine the incremental contribution of each type of variables. Do both contain unique information about future returns? To formally investigate the relationship between the two types of information, we conduct spanning regressions and bivariate sorting. Our results show that both market and accounting variables contain independent and incremental information about future returns. Put simply, accounting signals can predict returns even after controlling for market data and vice versa. However, the adjusted returns of market-based strategies consistently outperform those of accounting portfolios in this analysis. Moreover, for the largest stocks, which represent 90% of total market capitalization, the adjusted returns of accounting strategies are remarkably low and on the verge of standard statistical significance thresholds. Put another way, once market signals are considered, accounting information hardly matters for most of the investable equity universe.

Finally, our findings also shed light on the sources of market data superiority. Han et al. (2016) show that the relative effectiveness of fundamental and technical analysis may depend on the uncertainty of stock information. When stock information is highly uncertain, or when the noise-to-signal ratio is very high, fundamental signals, such as financial ratios or valuation ratios, are likely to be inaccurate. As a result, technical signals may dominate, proving more accurate and profitable. In line with this, we observe that market data-based strategies thrive mainly when information is uncertain. In particular, unlike accounting strategies, they extract profitability mainly from hard-to-value stocks. While accounting and market data-based models perform similarly for easy-to-value assets, market-based strategies generate significantly higher returns for assets with high valuation uncertainty. This phenomenon is also consistent with the stronger role of market data for smaller companies, which are typically harder to value. Moreover, a similar phenomenon can be observed in the time series dimension. In stable markets—when fundamental information uncertainty is low—there is no striking difference between the predictability of returns from market and accounting information. In volatile periods, however, the models based on market data clearly take the lead.

Our study relates to two strands of the academic literature. First, we extend the long-standing debate on the weak and semi-strong forms of market efficiency dating back to Fama (1970). More specifically, we connect to the studies of cross-sectional return predictability using market and accounting data. On the market data side, there is a wealth of evidence advocating the utility of technical analysis (e.g., Brock et al., 1992; Lo et al., 2000; Zhu & Zhou, 2009; Menkhoff, 2010; Han et al., 2013; Neely et al., 2014; Jiang et al., 2020; Avramov et al., 2021; Hung & Lai, 2022; Brogaard & Zareei, 2023), and Park and Irwin (2007) and Han et al. (2022) provide excellent surveys. On the other hand, some authors remain skeptical and point to problems with data snooping and transaction costs (e.g., Allen & Karjalainen, 1999; Sullivan et al., 1999; Bajgrowicz & Scaillet, 2012; Yamamoto, 2012). Similarly, the results of fundamental analysis using accounting data are most often supportive-especially when machine learning is involved (Bartram & Grinblatt, 2018; 2021; Hanauer et al., 2022; Yan & Zheng, 2017)—although, for example, Linnainmaa and Roberts (2018) argue that most accounting predictors are likely to be false discoveries.

A related line of research evaluates the value of fundamental analysis from the perspective of analyst recommendations. Again, the results are inconclusive, with Stickel (1992), Womack (1996), and, more recently, Azevedo and Müller (2020) supporting the usefulness of fundamental analysis, while Jaffe (1999) and Metrick (1999) find no benefit from investment newsletters. Very few studies compare and contrast the use of accounting and market data together. A notable exception is Avramov et al. (2018)—a study perhaps most closely related to ours—which examines fundamental and technical forecasts from the television show “Talking Numbers.” They find that while technicians can effectively predict stock returns, fundamental analysts do not add value.

Second, our findings add to the body of research on machine learning applications for predicting the cross-section of equity returns. A rapidly growing body of evidence has demonstrated their effectiveness in extracting information from a variety of sources, not only in stocks (e.g., Gu et al., 2020; Leippold et al., 2022; Cakici et al., 2023; Hanauer & Kalsbach, 2023), but also in other classes such as bonds, options, or commodities (Bali et al., 2020; Bianchi et al., 2021; Bali et al., 2023; Rad et al., 2023). A number of studies have pointed out the sensitivity of the efficiency of machine learning models to various factors, such as the forecast horizon (Blitz et al., 2023), economic restrictions (Avramov et al., 2023), or country characteristics (Cakici et al., 2023). In turn, we focus on the importance of the choice of the underlying set of predictors.

Third, our findings are related to the literature on the role of information uncertainty in the effectiveness of technical analysis. Previous studies have argued that technical signals are particularly useful in periods and market segments where stocks are difficult to value (e.g., Jiang et al., 2005; Zhang, 2006; Zhu & Zhou, 2009; Han et al., 2013, 2016; Detzel et al., 2021). This account is supported by several theoretical models that link the emergence of market trends to information uncertainty (e.g., Brown & Jennings, 1989; Cespa & Vives, 2012; Han et al., 2016). Our results agree with this line of reasoning, showing that return predictability through market data prevails especially for hard-to-value stocks and in volatile market periods.

The remainder of the article proceeds as follows. Section 2 discusses the data and methods. Section 3 reports the baseline results. Section 4 presents further insights and robustness checks. Section 5 scrutinizes the link between return predictability and information uncertainty. Finally, Section 6 concludes our study.

2. Data and Methods

In this section, we summarize our data and methods. We begin with the presentation of the samples of stocks and their characteristics. Next, we outline the prediction models and the evaluation methods of their forecasts.

2.1. Equity Universe

Our main sample covers the U.S. stock market, comprising firms listed on the NYSE, AMEX, and Nasdaq. The study period runs from January 1972 to December 2022 and is dictated by data availability—using earlier data results in a measurable decline in coverage by different stock characteristics. The market data comes from CRSP, and the accounting data is obtained from Compustat. We collect all the market and accounting data using the publicly available code from Jensen et al. (2023).[1] In particular, it also reproduces the authors' data cleaning and preparation procedures. Accordingly, we assume that accounting data is available four months after the fiscal period end. We perform this essential operation to prevent future information, which is not available at the time of the forecasting, from leaking into the prediction models. Moreover, we do not impose any restrictions on firm size, as we control for it in separate tests. Finally, as is common in asset pricing literature, our tests focus on monthly stock returns.

Table 1 presents an overview of our stock sample, with an additional focus on its size structure. As in Cakici and Zaremba (2022), we classify the companies into three categories: big, small, and micro. The *big firms* are the largest companies, which account for 90% of the total market capitalization. The S*mall firms* represent the subsequent 7% of the market capitalization. Finally, the *Micro firms* are the smallest stocks, accounting for the remaining 3% of the market. Our sample contains, on average, 5,549 firms per month with a mean market capitalization of $3.31 billion. Notably, as many as 3,456 of these stocks are microcaps, whose sheer number represents 62% of the sample. Nonetheless, their actual market value is relatively low, amounting to $158 million on average. On the other hand, the big stocks, representing 90% of the total market capitalization, account for only 18% of the companies, translating into less than 1,000 firms per month.

[Insert Table 1 here]

[1] Https://github.com/bkelly-lab/ReplicationCrisis.

2.2. Market and Accounting Characteristics

The essential input to all our prediction models are stock characteristics. To avoid any arbitrariness in their selection, we rely on the representative set of most prominent predictors from Jensen et al. (2023). Noteworthy, our dataset is limited to data based on accounting and market information that can be easily gathered from widely available databases and thus easily implemented in trading strategies. Consequently, we do not account for signals derived from other sources of information, such as insider trading, analyst, customer/supplier, institutional holdings, industry and peer-based signals.

Unlike most studies on machine learning applications to the cross-section of stock returns, we impose an additional restriction on data availability. Initial seminal papers, such as Gu et al. (2020), typically filled the gaps in coverage with simple algorithms, such as replacing missing characteristics with a cross-sectional median for each month. Yet, this practice can materially affect the training process for periods and characteristics with substantial amounts of missing data. The models could be effectively trained largely on artificial and invariant features, which may lead to potentially misleading inferences about their informativeness. Furthermore, many proposed advanced imputation algorithms (e.g., Bryzgalova et al., 2022; Freyberger et al., 2022; Cahan et al., 2023) pose only a partial remedy, as they may not fully account for the unique information content of a given feature that is not captured by other characteristics. To mitigate these challenges, we require that each characteristic be available for at least 30% of the securities in the sample in a given month. If this threshold is not met, the characteristic is dropped from the sample entirely for that month. After filtering out the data-poor months, we follow Gu et al. (2020) and replace the missing characteristics with a cross-sectional median for a given month. Increasing the bar for data coverage in our sample—as described above—affects mainly characteristics derived from long-run historical data, such as seasonality variables, which are computed using a multi-year period. Their availability is limited, particularly in the early years of the study period. Consequently, the number of available characteristics in our sample decreased from 153 available from Jensen et al. (2023) to 131 with sufficient coverage over the entire study period.

To assess the relative information content of market and accounting information, we split the 131 characteristics into 87 accounting (ACCT) and 44 markets (MKT) variables. When classifying the characteristics as ACCT or MKT, we use relatively broad definitions. Specifically, we assign a characteristic to the ACCT group if they require the use of any accounting information, even if market data are utilized as well. Hence, ACCT includes, for example, measures of profitability, accruals, or leverage. On the other hand, the MKT category encompasses all other variables, which could be calculated using solely market data, by which we understand current and past prices, volume, and market capitalization. This class includes indicators of momentum, liquidity, and volatility, among others. It also encompasses variables such as a bid-ask spread or Amihud ratio, which do not belong to a technician's common toolset, yet they rely solely on market data.

Overall, our criteria can be seen as more forgiving of accounting characteristics than market characteristics. If a variable depends on both market and accounting data, we place it in the accounting (ACCT) category. Hence, this group encompasses also valuation measures, which are ratios of accounting data to market capitalization. If we were to reclassify these characteristics as MKT or create a new “mixed” group, this would strengthen the MKT category, which we already see as dominant over ACCT in subsequent tests.

Noteworthily, the accounting features are almost twice as numerous as the market ones. The complete list of characteristics, along with their classification, is available in Appendix A. Importantly, all of them are calculated using the formulas from Jensen et al. (2023).

2.3. Machine Learning Models

To extract aggregate information from accounting and market variables, we resort to machine learning methods. Machine learning models allow us to handle large quantities of characteristics simultaneously, automatically selecting the most valuable predictors of future returns. Furthermore, they cope well with overfitting, which may arise when dealing with dozens of potentially correlated predictors. Last, more complex models can effectively capture non-linearities and interactions in the data, non-detectable by classical methods.

The machine learning literature offers a long (and rapidly growing) list of prediction models. To minimize potential selection bias associated with the choice of individual algorithms, our main inferences rely on the forecast combination model. This approach, with solid roots in statistics (Bates & Granger, 1969; Clemen, 1989; Timmermann, 2006), assumes that multiple models are combined to reduce their variance. As a result, it allows not only to reduce the forecast error (Petropoulos et al., 2022) but also to reduce the risk of model-specific outcomes. To compute a prediction from the combination model, we follow the simple and robust approach of Bali et al. (2023) and equally weight predictions from multiple individual algorithms. Specifically, we use seven models commonly used in cross-sectional studies of asset returns (e.g., Gu et al., 2020; Leippold et al., 2022; Bali et al., 2020, 2023).

Linear Regression with Ordinary Least Squares (OLS): OLS regression involves fitting a predictive model using all available features as inputs. This simple model does not require a regularization, hyperparameter tuning, or validation. However, it is prone to overfitting, especially in high-dimensional scenarios (Gu et al., 2020).

Partial Least Squares (PLS) Dimensionality Reduction: PLS serves as an effective technique for dimensionality reduction, taking into account the relationship between covariates and security returns. The three-pass PLS regression proposed by Kelly and Pruitt (2013, 2015) emphasizes characteristics that are highly correlated with stock payoffs. It aggregates individual predictors into composite factors to maximize the correlation with future returns. The tuning parameter in this regression is the number of components.

Penalized Linear Regression (LASSO, ENET): Penalized linear regression addresses overfitting by introducing a penalty term on the slope coefficients. We use two popular regularization approaches: LASSO (Least Absolute Shrinkage and Selection Operator) and ENET (Elastic Net), as introduced by Zou and Hastie (2005). LASSO penalizes the model based on the absolute values of its coefficients, while ENET combines LASSO and ridge regression components, focusing on the squared coefficients. In our case, both components have equal weight. ENET is more effective in dealing with covariate correlations than LASSO (Zou & Hastie, 2005; Diebold & Shin, 2019). The penalty term serves as a tuning parameter in both models.

Tree Models (RF, GBRT): Tree models, including Random Forests (RF) and Gradient Boosted Regression Trees (GBRT), offer flexibility in capturing interactions and non-linearities in returns. They classify observations (stock characteristics) into subcategories or "leaves." Trees are constructed iteratively, and their structure is determined by splitting variables and decision nodes. Overfitting of individual trees is mitigated by strong regularization. RF uses bootstrap aggregation or “bagging” to average multiple trees based on bootstrapped subsamples of the original data. GBRT, on the other hand, fits subsequent trees based on residuals from previous trees and aggregates predictions by successively multiplying numerous trees with a learning rate (0.1). We use the least squares boosting method (Breiman, 2001; Hastie et al., 2008) to fit the GBRT model, typically using between 100 and 200 learning cycles.

Feed-Forward Neural Network (FFNN): Feed-forward neural networks consist of an "input layer" representing stock characteristics, a variable number of "hidden layers" containing activation functions to transform the characteristics, and an "output layer" that converts the results from the hidden layers into return predictions. The neural network allows for modeling non-linearities and interactions, with flexibility increasing with the number of layers. Following Bali et al. (2022), we use a single hidden layer FFNN to take advantage of shallow learning (Gu et al., 2020). This hidden layer consists of eight neurons using rectified linear units as the activation function. We optimize the model using the Adam algorithm of Kingma and Ba (2014).

For each of these individual models, we estimate their parameters, tune hyperparameters, and evaluate their prediction performance using typical methods from machine learning literature. We assume a fixed (rolling) in-sample period spanning 15 years. At each re-estimation, we randomly split into a training sample, encompassing 70% of the observations, and a validation sample, comprising the remaining 30%. The training set estimates the model's parameters subject to pre-specified hyperparameters. The validation set, in turn, allows us to fine-tune the model hyperparameters to minimize the objective loss function. Last, we test the model's accuracy with a testing (holdout) dataset encompassing the subsequent twelve months. Notably, the testing period is an unseen portion of the data and never enters the training or validation sets. Notably, this holdout set allows for an unbiased model assessment of how the model is expected to perform on new, unseen data, offering an out-of-sample evaluation of its forecasting accuracy.

Given our study period of January 1972 to December 2022, the first in-sample period ends in December 1986, and the subsequent testing period encompasses data from January to December 1987. Given the computational intensity of machine learning models, we follow Gu et al. (2020) and Leippold et al. (2022), among others, and re-estimate them annually. We then continue this procedure until we reach the end of the sample. The total test period is thus January 1987 to December 2022, or 432 months.

2.4. Performance Evaluation

Our principal measure of prediction performance is the out-of-sample R^2 coefficient:

$$R^2_{OOS} = 1 - \frac{\sum_{(i,t)\in T_3} (r_{i,t+1} - \hat{r}_{i,t+1})^2}{\sum_{(i,t)\in T_3} {r_{i,t+1}}^2}, \tag{1}$$

with $\hat{r}_{i,t+1}$ and $r_{i,t+1}$ indicating the predicted and realized monthly stock i returns in month t+1, and T_3 represents the testing sample. R^2_{OOS} is calculated using the full sample of all monthly return observations pooled across stocks and time. High R^2_{OOS} values indicate that a model’s forecasts capture well the variation in stock returns.

While R^2_{OOS} is probably the most popular measure of prediction accuracy, yet, it might prove problematic to interpret or even irrelevant in certain contexts. Particularly, it may fail to capture the viewpoint of a quantitative portfolio manager who employs sorting to construct a portfolio. What matters in such cases is a model's ability to rank assets in line with their actual ex-post realized returns. Essentially, can it distinguish between winners and losers? R^2_{OOS} might obscure the link between predicted and realized returns, with the cross-sectional correlation getting overshadowed by return variances (Coqueret, 2022). As a result, investors may still realize measurable economic gains—even if R^2_{OOS} is negative (Kelly et al., 2022).

In particular, it may not reflect the perspective of a quantitative portfolio manager who employs sorting to build a portfolio. What matters to them is how effectively a model ranks assets according to their actual ex-post returns. In other words, can it distinguish between winners and losers? Nevertheless, the link between predicted and realized returns may be blurred for R^2_{OOS}, as the cross-sectional correlation gets lost in the return variances (Coqueret, 2022). Finally, even if R^2_{OOS} is negative, investors may still be able to make measurable economic gains (Kelly et al., 2022).

To address this issue, we complement R^2_{OOS} with simple cross-sectional correlation coefficients. Specifically, we compute time-series averages of monthly Parson product-moment ($\bar{\rho}_P$) and Spearman rank-based ($\bar{\rho}_S$) correlation coefficients between predicted and realized returns. The resulting numbers provide an intuitive snapshot of the relationship between model predictions and realized payoffs.

3. Baseline Findings

We begin the discussion of results with an overview prediction performance of different groups of variables. Next, we scrutinize which characteristics are particularly important. Finally, we investigate portfolio strategies based on different categories of information.

3.1. Prediction Performance

Figure 2, Panel A, focuses on the R^2_{OOS} coefficient. A quick glimpse at the results reveals that the market variables seem substantially more critical than their accounting counterparts. More specifically, for the total sample, the R^2_{OOS} equals 0.51%, a figure qualitatively similar to earlier findings in seminal studies of the U.S. market, such as Gu et al. (2020). Significantly, removing the accounting variables from the set of predictors does not substantially impair the forecasting accuracy. The R^2_{OOS} value for the MKT variables amounts to 0.47%, closely aligning with the full sample. On the other hand, limiting the model inputs to the accounting variables results in a visible decline in return predictability. Specifically, R^2_{OOS} drops to only 0.16%. To sum up, the information content of market variables appears incomparably richer than the accounting variables, manifesting in more than three times higher R^2_{OOS} coefficient. Furthermore, considering the market and accounting variables jointly leads to only marginal improvement versus the return predictability versus the market variables alone.

[Insert Figure 2 here]

The described pattern in return predictability remains consistent across different firm size categories: the market characteristics beat accounting characteristics everywhere. However, the precise magnitude of outperformance is uneven. The difference in market and accounting variables is most striking for microcaps, where the MKT R^2_{OOS} more than triples the ACCT-based predictions. For the small and big stocks, the spread is lower. On the other hand, in these two cases, there is no meaningful difference between R^2_{OOS} for models based on all market variables and those relying on market data only. In other words, the benefits of using incremental information from accounting variables appear negligible.

Panels B and C of Figure 2 focus on average cross-sectional correlation coefficients—measures that more closely align with investment practice. Notably, the results remain consistent in this setting. For all measures and all size categories, the market variables prove more efficient in sorting ranking stocks in line with ex-post returns than the accounting characteristic. The average correlation coefficients for the market features consistently noticeably surpass those of accounting data. Furthermore, using both accounting and market data leads to only marginal improvement relative to the market data.

To sum up, our initial evidence suggests that market variables encapsulate most of the valuable information about future stock returns. The information content of accounting data is either relatively low or already discounted in the market variables. Consequently, incorporating accounting data in the prediction model leads to only a modest improvement in prediction accuracy.

3.2. Which Characteristics Matter?

Having established the essential prediction performance, we now look at the relative importance of different characteristics. To capture the contribution of specific features to overall return predictability, we calculate variable importance (VI) as in Kelly et al. (2019). Specifically, for each market or accounting characteristic, we compute the decline in R^2_{OOS} caused by setting a given variable to zero while keeping all else fixed. In this way, we are able to capture the critical factors that influence the cross-sectional variation of stock returns while taking into account the comprehensive set of 131 signals within the system. For presentation purposes, following Gu et al. (2020), we rescale VI so that their sum equals 1.

We begin with a bird-eye view of different categories of variables. To this end, we compute VI for the prediction model based on all 131 characteristics in our sample. We consider the most general case, examining the total sample of all stocks over the entire study period. To capture the importance of the entire groups of accounting and market variables, we closely follow the approach of Bali et al. (2022) to estimate aggregate VI (rescaled previously to sum to 1) within the categories representing similar economic intuitions—as specified in Table 1. We aim to determine which groups matter the most for the cross-section of stock returns. Figure 3 displays the results of this exercise.

[Insert Figure 3 here]

The market characteristics prevail across all market segments. In the full market sample, their aggregate VI amounts to 58% versus 42% for the accounting variables. Admittedly, the values for distinct size categories may differ. For example, the aggregate VI of market variables for micro firms is 61% and for small firms, it is only 51%. Nonetheless, in all cases, the accounting features matter less. In particular, the market data turns out to be particularly important for the biggest firms in the market, where their aggregate VI equals 68%. This casts doubt on the widespread view that large and liquid stocks are remarkably efficient, undermining the profitability of technical analysis. Even in this segment, market variables are more informative than accounting ones—despite being outnumbered by a factor of two.

We now zoom in on the importance of individual variables. To this end, we re-estimate the prediction models based on accounting or market variables and scrutinize the most important features. Figure 4 reveals the findings.

[Insert Figure 4 here]

The top market features (Figure 4, Panel A) are dominated by variables reflecting market liquidity (e.g., *zero_trades_21d*, *zero_trades_126d*, *ami_126d*) and firm size (*market_equity*)—the last one matters particularly for the big firms. The micro and small firms, in turn, are more reliant on short-lived signals derived from the last month of daily data, such as *ret_1_0*, *max5_21d*, or *ivol_capm_21d*). The variables derived from longer horizons, such as *prc_highprc_252d*, still matter but score lower in the importance ranking.

For the accounting variables (Figure 4, Panel B), the key determinants differ partly across the firm size segments. For small companies, a crucial role is played by valuation ratios, such as *at_me*, *be_me*, or *div12m_me*. For small and big firms, what matters more is profitability (*gp_at*) or financial standing (*f_score*). Furthermore, investment patterns (*lti_gr1a*, *sti_gr1a*) prove relevant across all firm categories. Lastly, the variable importance across the accounting characteristics proves more democratic, spreading the contribution across a larger number of variables. On the other hand, for the market variables, the predictability concentrates on a handful of key characteristics.

3.3. Portfolio Strategies

Our analyses thus far indicate that market variables are more informative for future returns than accounting predictors. Does this pattern also translate into superior portfolio performance? Or, in other words, do strategies based on market data outperform those based on accounting characteristics?

To explore this question, we run univariate portfolio sorts on predicted returns based on different categories of data. Specifically, each month, we rank all the stocks on their forecasts from the combination model and group them into deciles. Next, we form value-weighted portfolios. Lastly, we also construct long-short portfolios, which buy the decile of stock with the highest expected returns and short those with the lowest. The performance of this portfolio serves as an intuitive acid test for the efficiency of different models from an investor perspective. We evaluate portfolio returns with the six-factor model of Fama and French (2018). The model accounts for the most common asset pricing factors: market, size, value, profitability, investment, and momentum:

$$R_t = \alpha + \beta_{MKT} MKT_t + \beta_{SMB} SMB_t + \beta_{HML} HML_t + \beta_{WML} WML_t + \beta_{RMW} RMW_t + \beta_{CMA} CMA_t + \varepsilon_t, \tag{2}$$

where R_t is the month t excess return on the evaluated portfolio and β_{MKT}, β_{SMB}, β_{HML}, β_{WML}, β_{RMW}, and β_{CMA} are estimated measures of factor exposure to risk factors: market risk factor (MKT), small minus big (SMB), high minus low (HML), winners minus losers (WML), robust minus weak (RMW), and conservative minus aggressive (CMA), all calculated as in Fama and French (2018). α represents the alpha, measuring the monthly abnormal return, and ε_t denotes the error term.

Table 2 reports the performance of one-way sorted portfolios, with Panel A focusing on value-weighted strategies. The strategies based on market variables (Panel B.2) generate higher returns than those based on accounting data (Panel C.1). MKT strategies earn, on average, 1.52% per month, somewhat outperforming their ACCT counterparts, which yield 1.25% per month. The difference cannot be captured by the exposures to the six Fama and French (2018) factors, and the alphas on MKT strategies remain noticeably higher as well, reaching 1.20% relative to 0.85% for the ACCT data. In a nutshell, relying on simple price, volume, and capitalization information, broadly available and easy to understand, proves more beneficial than resorting to more sophisticated accounting information.

[Insert Table 2 here]

Interestingly, the strategies based on all possible characteristics—both market and accounting alike—beat those based on just one category. However, the performance improvement is relatively modest. The forecasts based on all variables yield a value-weighted portfolio return of 1.71% with a six-factor model alpha of 1.34%. The synergies may originate both from selecting the best predictors from the two worlds—which contain some unique and independent information—as well as from capturing interactions between MKT and ACCT variables.

A closer look at the alphas on individual decile portfolios casts further light on the sources of the long-short strategy abnormal returns. For market variables, the alpha originates mainly from the short leg, where it amounts to -0.76%, compared with 0.43% on the top decile. This pattern is frequently linked with mispricing (e.g., Stambaugh et al., 2012, 2015), as the barriers to short selling constitute one of the forms of limits to arbitrage. On the other hand, the strategies based on accounting data fail to display such an asymmetry. In fact, the top decile produces even marginally higher absolute alphas (0.48%) than the bottom decile (-0.37%).

4. Further Insights and Robustness Checks

Having understood the overall effect of market and accounting variables, we now proceed with further analysis in robustness checks. First, we examine the sensitivity of our results to firm size, the exclusion of low-priced stocks, changes in forecasting models, subperiod analysis, and trading costs. We then test whether both types of variables contain independent information about future returns. Finally, we turn to international market evidence.

4.1. The Role of Firm Size

Is the outperformance of MKT variables consistent across different firm sizes? Or does it originate only from a particular segment, such as small stocks? The low-capitalization companies are commonly regarded as a reservoir of market inefficiencies. In consequence, numerous market predictors, such as those associated with momentum or liquidity (Hong et al., 2000; Hou et al., 2020), thrive in this market segment but disappoint in the bigger firms. More importantly, small stocks are also typically harder to value, which may matter for the relative efficiency of technical and fundamental signals (Han et al., 2016). To shed light on results consistency across firm size segments, we reproduce the portfolio sorts within the categories of micro, small, and big companies. We follow the classification based on aggregate market capitalization, as outlined in Table 1. Notably, for brevity, in all tests henceforth, we focus on value-weighted portfolios, which align more with a practical investor perspective.

Table 3 presents the performance of decile portfolios formed on ALL, MKT, and ACCT information within different size categories. Overall, our baseline conclusions hold across all market subsets. In each size segment, the strategies based on market variables outperform their accounting counterparts. However, the sheer magnitude of this outperformance, as well as the portfolio returns, may vary. The biggest discrepancies can be spotted for microcaps. In this case, the ALL, MKT, and ACCT strategies earn 3.94%, 3.44%, and 2.47% per month, respectively. In other words, the MKT variables beat ACCT by almost 100 basis points. On the other hand, in big caps, these differences—

while still present—are noticeably more timid. The MKT and ACCT strategies generate average monthly returns of 0.89% and 0.58%, respectively. Moreover, at the alpha level, the abnormal returns on the MKT and ACCT portfolios are 0.62% and 0.58%, respectively, indicating that the superiority of accounting data shrinks to only 3 basis points.

[Insert Table 3 here]

In summary, while the superior performance of market-based predictors is not just a statistical artifact arising from some illiquid microcaps with meaningless economic significance, the effect is visibly stronger among the smaller firms. In other words, among the largest firms in the market, accounting strategies perform almost as well as those using market data. This pattern may be related to the uncertainty of fundamental information, as the smaller firms are typically harder to value. We explore this issue further in Section 5 of this article.

4.2. The Effect of Low-Priced Stocks

As in most seminal asset pricing studies (e.g., Gu et al., 2020; Leippold et al., 2021), our equity universe comprises all stocks in the market, irrespective of their share price. However, low stock prices may imply substantial microstructure effects, such as bid-ask bounce. This, in turn, could artificially inflate the returns on market-based strategies, generating elevated returns that cannot be harvested in practice. The described phenomenon may affect particularly the microcap segment, typically less liquid than market blue chips. To alleviate such concerns, we reproduce our analysis with the low-priced stocks excluded from the sample. Precisely, each month t, we discard the stock with a share price below \$1 at the end of month t-1. Next, we reproduce our key tests.

Figure 5 illustrates the R^2_{OOS} values for different sets of variables in the sample with the low-priced stocks excluded. In a nutshell, all essential patterns hold. In particular, the return predictability is always stronger for the models based on MKT than for those relying on ACCT data. The differences are most apparent for microcaps but hold consistently for small and big caps alike. In other words, even in the biggest and most liquid stocks—when the microstructure effects associated with low prices are reduced—the market data still proves more informative about future returns than accounting information.

[Insert Figure 5 here]

Table 4 translates these patterns of return predictability into investment strategies based on portfolio sorts. Again, market variables dominate, demonstrating their superiority over accounting data. For example, the long-short ACCT strategy generates a six-factor alpha of 0.91% with a Sharpe ratio of 0.78. In comparison, the abnormal returns of the MKT strategy are almost 60% higher. Specifically, they generate a monthly alpha of 1.43% with a Sharpe ratio of 0.87. In other words, the higher R^2_{OOS} coefficient can be forged into more profitable investment strategies. Finally, the portfolios based on the predictions using all possible variables (ALL) outperformed both the MKT and ACCT sets. However, the magnitude of the outperformance over the MKT is not striking. For example, the monthly alpha is 1.49%, only six basis points higher than the MKT,

[Insert Table 4 here]

4.3. Individual Forecasting Models

The combination model, which we use in most of our calculations, helps minimize selection bias because it robustly averages several popular algorithms. But do our results hold for individual forecasting models? To answer this question, we replicate our portfolio tests for the predictions of the seven individual models used to construct the COMB forecast: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. For brevity, we focus only on the performance of the long-short portfolios, using the ALL, MKT, and ACCT sets of variables to determine the forecasts. Table 5 reports the results of this exercise in terms of mean returns and alphas, and Figure 6 supplements these with Sharpe ratios.

[Insert Table 5 here]

[Insert Figure 6 here]

A quick review of model performance allows two important observations to be made. First, the pattern of predictability associated with different categories of information is not limited to any particular model. For example, for most models, portfolio mean raw and abnormal returns are higher for MKT-based predictions than for those based on ACCT. Specifically, monthly alphas for MKT strategies range from 0.70% (PLS) to 2.09% (FFNN), while ACCT portfolios earn abnormal returns ranging from 0.15% (RF) to 0.91% (FFNN). In addition, the models that extract information from all variables often outperform both MKT and ACCT.

Second, while the outperformance of ACCT over MKT is widespread, it varies across models. In particular, the more complex strategies—those that account for nonlinearities and interactions—show the most striking differences between MKT and ACCT strategies. For example, consider FFNN. The MKT strategy's alpha of 2.09% is more than double that of ACCT, which is 0.91%. On the other hand, the analogous figures for the Elastic Net are 0.92% and 0.78%, respectively, so the difference shrinks to a minuscule 16 basis points. This suggests a greater potential for nonlinearities and interactions among the market-based variables that tree models and neural networks can effectively capture. An analogous pattern can easily be seen in the Sharpe ratios in Figure 5: the differences in performance between the MKT and ACCT models for RF, GBRT, and FFNN are obvious, while the effect is less clear for the simpler linear models.

4.4. Performance Through Time

Does the predictability of superior returns based on market variables persist over time? A substantial body of academic evidence suggests that the profitability of stock anomalies tends to decline over time, either due to investor learning or changes in market liquidity and efficiency (Schwert, 2003; Chordia et al., 2011; McLean & Pontiff, 2016; Zaremba et al., 2020). However, this pattern does not apply universally across the world (Jacobs, 2016; Jacobs & Müller, 2020) and may also vary depending on the type of predictor. On the one hand, the accessibility of accounting information has increased significantly in recent decades. On the other hand, technological advances have increased the ability of investors to quickly process large amounts of market data, allowing for the rapid deployment of complex technical signals.

Figure 7 shows the cumulative returns of long-short strategies based on different categories of information. A quick glance reveals a clear decline in the predictability of returns: over the past two decades, profits have been visibly lower than before. The effect is evident across all firm size classes, although it is not uniform. While the returns of large- and small-cap strategies have virtually stagnated, microcaps have continued to generate positive returns, albeit significantly lower than before. Notably, across all size segments, the decline in profitability appears to be more pronounced for strategies based on accounting information than for those based on market data. Again, the effect is particularly pronounced for microcaps, where cumulative returns have continued to grow even in recent years.

[Insert Figure 7 here]

Table 6 formalizes the discussion further by reporting mean returns and alphas in two equal subperiods: 1987 to 2004 and 2005 to 2022. Again, the decline in profitability is evident. While in the years 1972 to 2004 all strategies generated positive and significant profits in all company segments, in the last years - 2005 to 2022 - they survived mainly in microcaps. However, regardless of statistical significance, the market data continues to show higher returns in all three different size segments. For example, for microcaps, the average strategy returns for MKT and ACCT data are 1.94% and 1.20%, respectively, from 2005 to 2022. For large caps, the MKT and ACCT returns are 0.33% and 0.01%, respectively. These return differences remain visible after adjusting for returns using Fama and French's (2018) six-factor model.

[Insert Table 6]

In summary, the return predictability of market variables is consistently superior to accounting variables over time. This is true even in recent decades, when various types of data are widely available and investors can exploit various technical signals faster and more effectively than ever before.

4.5. Transaction Costs

Novy-Marx and Velikov (2016) demonstrate that the transaction costs of equity anomalies may depend strongly on the type of trading signal. Many strategies based on market data, such as short-term reversal, idiosyncratic volatility, or reversal, imply substantial portfolio turnover and, thus, trading costs. On the other hand, those relying on profitability or valuation ratios require less intensive trading. Do the strategies based on aggregate accounting and market information also share similar patterns?

To answer this question, we calculate the average monthly portfolio turnover of our portfolio strategies. As in Novy-Marx and Velikov (2016) and Barroso and Detzel (2021), among others, we compute monthly turnover as the average share of the portfolio being replaced each month:

$$PT_t = \frac{1}{2}\sum_{i=1}^{n} \left| w_{i,t-1} \times (1 + r_{i,t}) - w_{i,t} \right|, \tag{3}$$

where $w_{i,t-1}$ and $w_{i,t}$ are the weights of stock i in two consecutive months, and $r_{i,t}$ is the month t return on stock i. We calculate a one-sided (rather than two-sided) turnover measure to avoid double-counting of buys and sells.

Next, by juxtaposing the trading intensity with average returns, we estimate the breakeven transaction costs. Specifically, the breakeven cost rate is defined as the level of one-way trading costs (in %) at which the average net return declines to zero. Since many anomaly premia can be effectively harvested via both long-short and long-only approaches (Blitz et al., 2020), we conduct these calculations for both a strategy tracking the top decile stocks, as well as to long-short portfolio that simultaneously buys the top stocks and sell those with the lowest expected returns. Table 7 presents the results of this experiment.

[Insert Table 7 here]

First, consider the long-short portfolios in Panel A. While the MKT strategies tend to have higher average returns, they are also associated with higher portfolio turnover. For all firms, the average monthly turnover for MKT and ACCT is 188% and 174%, respectively. Moreover, the differences are even more pronounced for certain size segments, such as microcap portfolios, where turnover is 170% and 127% for MKT and ACCT, respectively. As a result, the break-even transaction costs for these two information categories are comparable and typically differ by only a few basis points. They also depend on the size of the company and are typically higher for small and micro stocks, where average earnings are more robust. Thus, for microcaps, the break-even costs of MKT and ACCT strategies are 100 and 97 basis points, respectively, while for large and small companies they are 26 and 21 basis points, respectively.

The long-only strategies require significantly less trading (Panel B) and are therefore associated with lower portfolio turnover. This, in turn, leads to higher breakeven costs —even though the strategies do not extract profits from the short positions. However, the relationship between market and accounting data remains similar, with the former showing simultaneously higher gross profits and portfolio turnover than the latter. As a result, the breakeven costs remain pretty similar. For example, for micro caps, the breakeven rates are slightly higher for ACCT strategies (126 versus 114 bps), while for large caps, MKT strategies prevail (62 versus 54 bps).

To conclude, the strategies based on market information generate both higher pre-cost returns and portfolio turnover. In consequence, the breakeven costs of both MKT and ACCT strategies are comparable—regardless of the firm size segment considered.

4.6. Independent Information in Accounting and Market Variables

The evidence thus far suggests that signals derived from market data are more informative about future returns than accounting variables. They predict future returns more accurately and turn into more profitable investment strategies. However, do both sets of data contain independent and incremental information about future returns? Or, perhaps the market variables already discount all accounting information, rendering it redundant from an investor perspective? To answer these questions, we conduct two types of analysis: mean-variance spanning tests and bivariate portfolio sorts.

4.6.1. Spanning Tests

We begin with mean-variance spanning tests. In this exercise, we regress the returns on a strategy formed using one information set on its counterpart based on another type of information:

$$R_{X,t} = \alpha_N + \beta_Y R_{Y,t} + \varepsilon_t, \tag{4}$$

$$R_{X,t} = \alpha_B + \beta_Y R_{Y,t} + \beta_{MKT} MKT_t + \beta_{SMB} SMB_t + \beta_{HML} HML_t + \beta_{WML} WML_t + \beta_{RMW} RMW_t + \beta_{CMA} CMA_t + \varepsilon_t. \tag{5}$$

$R_{X,t}$, and $R_{Y,t}$ in Eq. (4) and (5) denote monthly portfolio returns on strategies based on different types of information, and the coefficient β_Y measures the mutual exposure.

The $R_{X,t}$, and $R_{Y,t}$ strategy returns are calculated as long-short portfolios buying (selling) the decile of stocks with the highest (lowest) expected returns, identically as in Section 3.3. Compared to (4), Eq. (5) additionally controls for the exposure to Fama and French's (2018) six factors, and α_N and α_N denote alphas from the "narrow" (without control factors) and "broad" (with control factors) regression specification, respectively.

We run all the tests for all three information sets: market, accounting, and all. Furthermore, we consider both the total sample of all stocks as well as different size categories: micro, small, and big. Table 8 reports the results of this examination.

[Insert Table 8 here]

The spanning regressions reveal several interesting findings. To begin with, in most specifications, both accounting and market data contain independent information about future stock returns. In other words, regressing them on each other yields positive and significant alphas. Consistent with this, neither market nor accounting variable-based strategies can subsume the returns of all variables together. This confirms that, in most cases, both sets contribute incrementally to return prediction.

Nevertheless, the abnormal returns on market strategies after controlling for their exposure to accounting strategies are typically higher than vice versa. For example, for the entire sample of all stock Eq. (5), alphas on market and accounting strategies amount to 0.97% and 0.62%, respectively. In other words, market strategies remain more profitable than accounting ones. The alpha differences are the biggest for microcaps, where they exceed one percentage point. On the other hand, in the significant firm segment, they are small yet still noticeable.

Notably, in certain specifications, strategies based on market data fully subsume those based on accounting data, while the opposite is not valid. This is, for example, the case of regression (4) (without control factors) applied to the big firm segment, representing 90% of the entire global market capitalization. The accounting alpha shrinks to just 0.29% (t-stat = 1.48) and no longer statistically differs from zero. Admittedly, this phenomenon is not confirmed by the regression (5)—which controls for the factor exposure—applied to the same size segment; the alpha in this specification equals 0.46% (t-stat = 2.39), statistically differing from zero at the conventional levels. Nonetheless, still, the results underline the limited utility and value-added of accounting variables—especially in the small and big cap segments, which represent the vast majority of the investable equity universe.

Lastly, it is worth noting that the strategies based on all variables frequently—though not always—generate significant and positive alphas even after accounting for the exposure to both market and accounting data. It means that both categories of information interact, generating additional premium compared to when utilized on a standalone basis. This observation holds for both microcaps and big stocks but to a lesser extent in small firms.

4.6.2. Bivariate Portfolio Sorts

To corroborate our conclusions from mean-variance spanning tests, we now proceed with bivariate portfolio sorts. In this experiment, we sort all stocks independently into quintiles based on the forecast using only market or accounting information. The intersection of these two distinct sorts into quintiles produces 25 double-sorted portfolios. For each of the MKT and ACCT quintiles, we compute long-short portfolios buying the stocks with the highest (lowest) expected returns based on one category of information within a similar level of expected returns based on the other type of data. Finally, we also calculate average returns across all five quintiles to capture the incremental premium of one type of information after controlling for the other. As in the spanning regressions, we repeat this exercise for all firm size categories.

The results of this analysis, as summarized in Table 9, align with our earlier findings. In most cases, both market and accounting data provide incremental information about future returns. That is, market variable-based strategies generate significant raw and risk-adjusted returns after controlling for the accounting-based forecasts and vice versa. Nonetheless, payoffs on the market-based strategies are typically higher than on their accounting counterparts. For example, in the total sample of all firms, the average Fama-French (2018) alpha on the five MKT strategies is 0.61%, compared to 0.48% for the accounting data.

[Insert Table 9 here]

Similar to all previous results, the magnitude of the MKT outperformance is uneven and depends on the size of the company. The largest differences are observed for microcaps, where the average MKT alpha after controlling for ACCT is 1.94%, while the ACCT alpha after controlling for MKT is 1.30%. The absolute differences are smaller for other segments. However, for large firms, the average payoffs and alphas for many ACCT strategies are low and on the edge of statistical significance—or even insignificant. In particular, the average return across all five quintiles formed on ACCT predictions is only 0.26% (t-stat = 1.65) versus 0.62% return on MKT strategies.

To conclude, the spanning regressions and bivariate sorts alike point to a substantial amount of independent information about future returns embedded in both market and accounting data. In other words, neither market data encapsulates all accounting information nor vice versa. Nevertheless, the information content of market variables is typically richer, resulting in stronger portfolio returns. Furthermore, in certain market segments, such as big caps, it renders the abnormal returns on accounting strategies minuscule and barely significant.

4.7. Evidence from International Markets

Do our findings extend to international markets? Asset pricing literature is highly U.S.-centric, with most studies published in top-tier journals focusing on U.S. stocks (Karolyi, 2016). At the same time, formulating broad generalizations based solely on the U.S. data can be risky, as finance literature has already highlighted cases when the conclusions from the United States fail to hold in the global setting (e.g., Goyal & Wahal, 2015; Jacobs & Muller, 2020; Cakici & Zaremba, 2022; Cakici et al., 2023). At the same time, non-U.S. stocks make up a significant fraction of global equity markets.

To cast light on the information content of market and accounting data internationally, we replicate our baseline portfolio tests in ten major developed markets: Australia, Canada, France, Germany, Hong Kong, Italy, Japan, Singapore, Sweden, and the United Kingdom. Notably, the availability of stock characteristics in international markets tends to be lower than in the United States. Hence, the choice of these ten countries aims to ensure a considerable number of return-predicting variables available throughout the entire research period. As seen in the rightmost section of Table 10, their total number varies between 78 in Italy to 116 in the United Kingdom. Notably, except for Italy and Japan, in all other markets, there are more accounting than market characteristics.

The study period for foreign stocks is from January 1996 to December 2022, with the first ten years serving as an in-sample period and the actual testing period beginning in 2006. All the data come from Compustat and are compiled using the same procedures as in our U.S. tests. Furthermore, each month, we discard 10% of stocks with the lowest stock price to reduce the undue impact of microstructure issues in international markets. Finally, following standard conventions in global studies (e.g., Fama & French, 2012, 2017), we express all returns in U.S. dollars.

With the international data at hand, we reproduce our previous portfolio formation procedures. Specifically, we train the models using market, accounting, or all variables. We employ a ten-year in-sample period and re-estimate the prediction models each year. The models are trained and validated within individual markets. Our forecasts come from a combination model, aggregating the same seven individual models as for the tests. Lastly, the stocks are sorted into deciles based on expected returns to form value-weighted portfolios. For conciseness, we focus solely on the long-short strategies, buying (selling) the 10% stocks with the highest (lowest) predicted returns. We then evaluate them with the model (2) of Fama and French (2018), with its factors derived from local market data.

Figure 8 presents the Sharpe ratios on the spread portfolios in international markets. Overall, the evidence aligns with the superiority of the market variables over their accounting counterparts. In seven out of ten countries, the Sharpe ratios on the strategies based on market data exceed those relying on accounting information, and in some instances—such as Australia and Canada—the difference is striking. Only in three countries where the market-based predictions fail to outperform: France, Italy, and Japan—though in Italy, both categories work on par, generating similar risk-adjusted returns.

[Insert Figure 8 here]

Interestingly, unlike in the United States, not in all countries, the strategies based on all characteristics outperform both market and accounting strategies in terms of their Sharpe ratios. More specifically, only in three countries—Hong Kong, Singapore, and the United Kingdom—the prediction model based on a pooled sample of all features generated the best risk-return profile. The strategy based on market variables only typically works as well as it—or even proves superiors, as in the cases of Australia and Canada.

Table 10 further formalizes the overview of international markets by reporting the mean returns and alphas on different types of strategies. Clearly, the portfolios based on market variables visibly outperform those based on accounting data. More specifically, market strategies typically produce raw and risk-adjusted payoffs more than twice as high as their accounting counterparts. The average alpha on portfolios formed on market data across the ten analyzed markets equals 1.78%, while its counterpart for accounting data is only 0.82%. Looking at individual markets, the market strategies beat the accounting ones in seven out of ten countries. In some of them, such as Australia and Canada, the differences in alphas reach as much as three percentage points. On the other hand, the worst performance belongs to the strategies in Japan, when they earn the six-factor model alpha of 0.51% versus 0.80% for accounting strategies.

[Insert Table 10 here]

Notably, unlike for the United States, Table 10 suggests that the strategies based on all variables typically generate comparable performance as those based on market data only. While both approaches commonly beat the accounting strategies, their levels of returns and alphas are relatively similar. This observation further undermines the value added of using accounting data for global return predictions. Apparently, most of the essential information is already discounted in market variables.

To sum up, while international markets reveal a certain degree of variability across countries, the global evidence generally supports the U.S. findings. The strategies based on market variables not only typically beat those based on accounting data but also perform on par with forecasting models that all market and accounting predictors together.

5. Return Predictability and Information Uncertainty

The effectiveness of fundamental and technical analysis may depend on the uncertainty of stock information (Han et al., 2016). When stock information is highly uncertain, or when the noise-to-signal ratio is very high, fundamental signals are likely to be inaccurate. As a result, investors tend to rely more heavily on market information. In consequence, technical signals are likely to be more profitable when the available information is uncertain.

Notably, the described dependence can manifest itself in both cross-sectional and time-series dimensions, i.e., at both the stock and market levels. For example, Detzel et al. (2021) document that technical analysis is profitable for hard-to-value assets, and the link between stock-level information uncertainty holds for various return patterns, such as price momentum and post-earnings announcement drift (Jiang et al., 2005; Zhang, 2006), moving averages (Han et al., 2013), and the trend factor (Han et al., 2016). On the other hand, in the time series setting, the more uncertain the future information, the more volatile the stock price risk. As a result, time-varying market volatility may also affect the predictability of returns from market data. High aggregate volatility suggests that fundamental signals may be imprecise, leading investors to rely more on technical signals. This narrative aligns with the superior profitability of technical and trend factor strategies in volatile periods (Zhu & Zhou, 2009), as well as with various theoretical models. For example, Brown and Jennings (1989) argue that rational investors can benefit from forming expectations based on historical prices and that this advantage increases with asset volatility, and Cespa and Vives (2012) show that the presence of asset payoff uncertainty can give rise to rational trends in asset prices.

Building on the studies above, we run two analyses to explore the link between information uncertainty and return predictability by market and accounting information. First, in the cross-sectional dimension, we scrutinize the information content of market and accounting data in hard-to-value stocks. Second, in the time-series dimension, we investigate the interplay between return predictability and market volatility.

5.1. The Role of Hard-to-Value Stocks

We examine the cross-sectional impact of information uncertainty with the use of two-way independent sorts. To identify the hard-to-value stocks, we calculate a composite measure of valuation uncertainty. To this end, for each stock each month, we compute four proxies for valuation uncertainty employed commonly in finance literature (e.g., Kumar, 2009; Ben David et al., 2019, 2023; Xiong et al., 2020): firm age (*age*), quarterly return on assets (*niq_at*), share turnover (*turnover_126d*), and idiosyncratic risk (*ivol_ff3_21d*). Following earlier studies, we assume that young unprofitable firms with high turnover and idiosyncratic risk are harder to value than old and profitable companies characterized by low turnover and idiosyncratic risk. Finally, we calculate an average cross-sectional rank to obtain an aggregate measure of valuation uncertainty.

With this measure in hand, we conduct double sorts. Specifically, we intersect two independent sorts into quintiles on i) valuation uncertainty and ii) return predictions from the COMB model, supplied with market and accounting information. In doing so, we want to see how the models based on different types of data perform on easy-to-value and hard-to-value stocks. Table 11 reports the results of this exercise, with Panels A and B focusing on market and accounting information, respectively.

[Insert Table 11 here].

A side-by-side comparison of Panels A and B reveals a remarkable difference in the behavior of strategies based on market and accounting information. The profits of long-short portfolios based on market data exhibit a visible link to valuation uncertainty: their magnitude increases substantially for hard-to-value assets. Specifically, the average return of the long-short strategy implemented in the hardest-to-value quintile is 1.39 percentage points higher than its counterpart in the easiest-to-value quintile. The difference is statistically significant at conventional levels and cannot be attributed to exposure to common factors.

On the other hand, accounting strategies reveal no similar pattern. The mean returns on the long-short COMB strategies are qualitatively similar across all quintiles. Moreover, the difference between the hardest and easiest to value groups is only 0.25 percentage points, not significantly different from zero. In other words, we cannot detect any reliable relationship between valuation uncertainty and return predictability. In fact, the mean returns across all quintiles, ranging from 0.59% to 0.94%, are qualitatively similar to the mean returns of market strategies in the easiest-to-value quintile (0.70%).

In summary, Table 11 identifies the source of the superior performance of market-based strategies. The return predictability by market variables derives its superiority mainly from difficult-to-value assets, where it vastly outperforms accounting strategies. Specifically, in the hardest-to-value quintile, the long-short strategy based on market and accounting data generates mean monthly gains of 2.09% and 0.84%, respectively. On the other hand, in the easier-to-value stocks, the differences are minimal and both types of information lead to comparable performance. In short, it is the hard-to-value stocks that matter.

5.2. Time-Series Variation in Aggregate Volatility

Having assessed the cross-sectional variation in return predictability, we now turn to the time-series dimension. Specifically, following Zhu and Zhou (2009), we examine the relationship between time-varying aggregate volatility and the information content of accounting and market data. To this end, we compute the value-weighted average stock volatility (rvol_21d) across all firms for each month. We then divide the full study into two halves by the median of the aggregate volatility and examine the performance of the COMB strategies within the sub-periods. Table 12 summarizes the results. Specifically, we present the mean returns and six-factor model alphas of the long-short portfolios constructed using different data types within the periods of above-median and below-median volatility.

[Insert Table 12 here].

In the low-volatility regimes, both the MKT and ACCT strategies perform comparably well and generate qualitatively similar payoffs. In fact, once we account for common factor exposure, the performance of the ACCT strategies is even slightly stronger, yielding a monthly alpha of 0.69% versus 0.48% for MKT. However, the situation changes significantly when we focus on the high volatility regime. In this case, the strategies based on market information clearly outperform, generating abnormal returns that are more than twice as high as those of the accounting strategies. More precisely, the alpha of the MKT and ACCT portfolios is 2.28% and 0.98%, respectively. In other words, while the ACCT alpha increased by only 31 basis points relative to the low-volatility regime, the performance of the MKT strategy was more than quadrupled. Finally, the results at the low return level are less striking, but the differences are still clear and substantial: the average monthly return of the MKT portfolios is one percentage point higher than its ACCT counterpart.

In conclusion, the evidence from the time-series variation in return predictability also supports the view that market information predictability prevails in periods of high information uncertainty. While in stable periods the magnitude of abnormal returns is comparable for both types of data, in volatile markets it is market data that takes the lead. Simply put, technical signals extract value mainly in periods and segments where valuation is particularly uncertain.

6. Concluding Remarks

In this study, we compare the cross-sectional predictability of stock returns by accounting and market information. To this end, we examine five decades of data from the U.S. market and compute a comprehensive set of 131 prominent return predictors. To capture the relevance of different types of information, we classify them into two groups – market and accounting – based on the data from which they originate. Finally, we employ various machine learning techniques, including simple and penalized regression, dimension reduction techniques, tree models, and neural networks, to extract return predictions from different categories of signals.

Our results reveal remarkable differences in the predictive performance of different data types. Market variables are more informative than accounting variables, yielding more accurate return forecasts. The out-of-sample R^2 coefficient for predictions based on market data alone is about three times higher than those based on accounting characteristics. The superior predictive properties of market variables hold across firm size categories and outperform accounting signals even in the large firm segment. Moreover, considering market and accounting variables together only marginally improves predictive accuracy, suggesting that the contribution of accounting variables is limited.

A closer look at the importance of different predictor categories sheds further light on the contribution of market and accounting variables. When all variables are considered together, the overall importance of market variables is visibly higher than that of accounting predictors. The critical signals are typically related to market frictions, such as the number of zero-return days or short-term reversals. Interestingly, market variables also dominate for large-cap stocks, which are generally considered to be the most efficient and leave little money on the table for technical analysts.

The superior predictive power of market variables translates into higher returns for the corresponding investment strategies. A long-short value-weighted portfolio that buys (sells) the stocks with the highest (lowest) expected returns has a six-factor model alpha of 1.20% when the return forecasts are based on market information. For accounting information, the alpha is 0.85%.

The dominance of market variables is robust to many considerations. It holds for different firm size categories, as well as after excluding low-priced stocks. Furthermore, it survives both in composite forecasts and individual forecast models. Moreover, although the magnitude of return predictability varies over time, it also survives in subperiods. In addition to that, despite unequal portfolio turnover, both market and accounting data allow for comparable breakeven trading costs. Finally, our conclusions are not limited to the U.S. market. After studying ten major international markets, we find that models based on market data overperform in most of them. The effectiveness of accounting forecasts tends to be lower, which leads to poorer performance of foreign investment portfolios.

Given the apparent superiority of market variables, do both market and accounting variables actually contain independent and valuable information about future returns? To shed light on this issue, we run mean-variance spanning tests and bivariate portfolio sorts that compare forecasts based on different types of information. Interestingly, while the performance of market-based strategies is typically stronger after controlling for accounting information than vice versa, they both contain incremental information about future returns. Nevertheless, the marginal contribution of accounting information is sometimes small, and in some instances, such as for large-cap stocks, it is on the verge of statistical significance.

Finally, our results shed light on the origins of market information superiority. The effectiveness of technical signals prevails when fundamental information is uncertain. As a consequence, market-based strategies perform particularly well for hard-to-value stocks and in volatile periods. On the other hand, when information uncertainty is low, book-based models work almost as well as their market data-based counterparts.

Our study provides clear, practical implications. Using accounting data—often less accessible, pricier, and more difficult to analyze—may be of limited value. Investors can often achieve similar, if not better, performance by relying solely on readily available market data and technical analysis tools—especially for those stocks that are most difficult to value.

References

Allen, F., & Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. *Journal of Financial Economics*, *51*(2), 245–271.

Amihud, Y., Hameed, A., Kang, W., & Zhang, H. (2015). The illiquidity premium: International evidence. *Journal of Financial Economics*, 117(2), 350–368.

Avramov, D., Cheng, S., & Metzker, L. (2023). Machine learning vs. economic restrictions: Evidence from stock return predictability. *Management Science*, *69*(5), 2587–2619.

Avramov, D., Kaplanski, G., & Levy, H. (2018). Talking Numbers: Technical versus fundamental investment recommendations. *Journal of Banking & Finance*, *92*, 100-114.

Avramov, D., Kaplanski, G., & Subrahmanyam, A. (2021). Moving average distance as a predictor of equity returns. *Review of Financial Economics*, *39*(2), 127-145.

Azevedo, V., & Müller, S. (2020). Analyst recommendations and mispricing across the globe. *Available at SSRN 3705141*.

Bajgrowicz, P., & Scaillet, O. (2012). Technical trading revisited: False discoveries, persistence tests, and transaction costs. *Journal of Financial Economics*, 106(3), 473-491.

Bali, T. G., Beckmeyer, H., Moerke, M., & Weigert, F. (2023). Option return predictability with machine learning and big data. *Review of Financial Studies*, *36*(9), 3548-3602.

Bali, T. G., Goyal, A., Huang, D., Jiang, F., & Wen, Q. (2020). Predicting corporate bond returns: Merton meets machine learning. *Georgetown McDonough School of Business Research Paper*, (3686164), 20-110.

Barroso, P., & Detzel, A. (2021). Do limits to arbitrage explain the benefits of volatility-managed portfolios? *Journal of Financial Economics*, 140(3), 744-767.

Bartram, S. M., & Grinblatt, M. (2018). Agnostic fundamental analysis works. *Journal of Financial Economics*, 128(1), 125-147.

Bartram, S. M., & Grinblatt, M. (2021). Global market inefficiencies. *Journal of Financial Economics*, 139(1), 234-259.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, *20*(4), 451–468.

Ben-David, I., Birru, J., & Rossi, A. (2019). Industry familiarity and trading: Evidence from the personal portfolios of industry insiders. *Journal of Financial Economics*, 132(1), 49-75.

Ben-David, I., Kim, B., Moussawi, H., & Roulstone, D. (2023). Corporate Transactions in Hard-to-Value Stocks. *Review of Corporate Finance Studies*, 12(3), 539-580.

Bianchi, D., Büchner, M., & Tamoni, A. (2021). Bond risk premiums with machine learning. *Review of Financial Studies*, 34(2), 1046-1089.

Blitz, D., Baltussen, G., & van Vliet, P. (2020). When equity factors drop their shorts. *Financial Analysts Journal*, 76(4), 73-99.

Blitz, D., Hanauer, M. X., Hoogteijling, T., & Howard, C. (2023). T]The term structure of machine learning alpha. *Journal of Financial Data Science,* in press.

Breiman, L. (2001). Random forests. *Machine Learning*, pp. 45, 5–32.

Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, 47(5), 1731-1764.

Brogaard, J., & Zareei, A. (2023). Machine learning and the stock market. *Journal of Financial and Quantitative Analysis*, 58(4), 1431-1472.

Brown, D. P., & Jennings, R. H. (1989). On technical analysis. *Review of Financial Studies*, 2(4), 527-551.

Bryzgalova, S., Lerner, S., Lettau, M., & Pelger, M. (2022). Missing financial data. Available at SSRN 4106794.

Cahan, E., Bai, J., & Ng, S. (2023). Factor-based imputation of missing values and covariances in panel data of large dimensions. *Journal of Econometrics*, 233(1), 113-131.

Cakici, N., & Zaremba, A. (2022). Salience theory and the cross-section of stock returns: International and further evidence. *Journal of Financial Economics*, 146(2), 689-725.

Cakici, N., Fieberg, C., Metko, D., & Zaremba, A. (2023). Machine learning goes global: Cross-sectional return predictability in international stock markets. *Journal of Economic Dynamics and Control*, *155*, 104725.

Cespa, G., & Vives, X. (2012). Dynamic trading and asset prices: Keynes vs. Hayek. *Review of Economic Studies*, 79(2), 539-580.

Chordia, T., Roll, R., & Subrahmanyam, A. (2011). Recent trends in trading activity and market quality. *Journal of Financial Economics*, 101(2), 243-263.

Clemen, R.T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583.

Coqueret, G. (2022). Persistence in factor-based supervised learning models. *Journal of Finance and Data Science*, 8, 12–34.

Detzel, A., Liu, H., Strauss, J., Zhou, G., & Zhu, Y. (2021). Learning and predictability via technical analysis: evidence from Bitcoin and stocks with hard-to-value fundamentals. *Financial Management*, 50(1), 107-137.

Diebold, F. X., & Shin, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives. *International Journal of Forecasting*, *35*(4), 1679-1691.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2), 383–417.

Freyberger, J., Höppner, B., Neuhierl, A., & Weber, M. (2022). Missing data in asset pricing panels (No. w30761). National Bureau of Economic Research.

Goyal, A., & Wahal, S. (2015). Is momentum an echo? *Journal of Financial and Quantitative Analysis*, 50(6), 1237-1267.

Green, J., Hand, J. R., & Zhang, X. F. (2017). The characteristics that provide independent information about average US monthly stock returns. *Review of Financial Studies*, *30*(12), 4389–4436.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223–2273.

Han, Y., Liu, Y., Zhou, G., & Zhu, Y. (2021). Technical analysis in the stock market: A review. Available at SSRN 3850494.

Han, Y., Yang, K., & Zhou, G. (2013). A new anomaly: The cross-sectional profitability of technical analysis. *Journal of Financial and Quantitative* Analysis, 48(5), 1433-1461.

Han, Y., Zhou, G., & Zhu, Y. (2016). A trend factor: Any economic gains from using information over investment horizons? *Journal of Financial Economics*, 122(2), 352-375.

Hanauer, M. X., & Kalsbach, T. (2023). Machine learning and the cross-section of emerging market stock returns. *Emerging Markets Review*, *55*, 101022.

Hanauer, M. X., Kononova, M., & Rapp, M. S. (2022). Boosting agnostic fundamental analysis: Using machine learning to identify mispricing in European stock markets. *Finance Research Letters*, 48, 102856.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning, second edition.* New York: Springer.

Hong, H., Lim, T., & Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance*, 55(1), 265–295.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *Review of financial studies*, *33*(5), 2019-2133.

Hung, C., & Lai, H. N. (2022). Information asymmetry and the profitability of technical analysis. *Journal of Banking & Finance*, *134*, 106347.

Jacobs, H. (2016). Market maturity and mispricing. *Journal of Financial Economics*, 122(2), 270–287.

Jacobs, H., & Müller, S. (2018). ... and nothing else matters? On the Dimensionality and Predictability of International Stock Returns. *On the Dimensionality and Predictability of International Stock Returns (November 20, 2018).*

Jacobs, H., & Müller, S. (2020). Anomalies across the globe: Once public, no longer existent? *Journal of Financial Economics*, 135(1), 213-230.

Jaffe, J. F., & Mahoney, J. M. (1999). The performance of investment newsletters. *Journal of Financial Economics*, *53*(2), 289–307.

Jensen, T. I., Kelly, B., & Pedersen, L. H. (2023). Is there a replication crisis in finance? *Journal of Finance*, 78(5), 2465-2518.

Jiang, G., Lee, C. M., & Zhang, Y. (2005). Information uncertainty and expected returns. *Review of Accounting Studies*, 10, 185-221.

Jiang, J., Kelly, B. T., & Xiu, D. (2020). (Re-) Imag (in) ing price trends. *Chicago Booth Research Paper*, (21–01).

Kelly, B. T., Malamud, S., Zhou, K. (2022). The virtue of complexity in machine learning portfolios. *Journal of Finance*, in press.

Kelly, B., & Pruitt, S. (2013). Market expectations in the cross-section of present values. *Journal of Finance*, 68(5), 1721-1756.

Kelly, B., & Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, 186(2), 294-316.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kumar, A. (2009). Hard-to-value stocks, behavioral biases, and informed trading. *Journal of Financial and Quantitative Analysis*, 44(6), 1375-1401.

Leippold, M., Wang, Q., & Zhou, W. (2022). Machine learning in the Chinese stock market. *Journal of Financial Economics*, *145*(2A), 64-82.

Linnainmaa, J. T., & Roberts, M. R. (2018). The history of the cross-section of stock returns. *Review of Financial Studies*, 31(7), 2606-2649.

Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance*, *55*(4), 1705-1765.

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, *71*(1), 5–32.

Menkhoff, L. (2010). The use of technical analysis by fund managers: International evidence. *Journal of Banking & Finance*, 34(11), 2573–2586.

Metrick, A. (1999). Performance evaluation with transactions data: The stock selection of investment newsletters. *Journal of Finance*, *54*(5), 1743-1775.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: the role of technical indicators. *Management Science*, *60*(7), 1772-1791.

Neely, C., Weller, P., & Dittmar, R. (1997). Is technical analysis in the foreign exchange market profitable? A genetic programming approach. *Journal of Financial and Quantitative Analysis*, 32(4), 405-426.

Novy-Marx, R., & Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *Review of Financial Studies*, 29(1), 104-147.

Park, C. H., & Irwin, S. H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys*, *21*(4), 786–826.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S. B., ... & Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, 38(3), 705-871.

Rad, H., Low, R. K. Y., Miffre, J., & Faff, R. (2023). The commodity risk premium and neural networks. *Journal of Empirical Finance*, 101433.

Schwert, G. W. (2003). Anomalies and market efficiency. *Handbook of the Economics of Finance*, 1, 939-974.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2), 288–302.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance*, 70(5), 1903-1948.

Stickel, S. E. (1992). Reputation and performance among security analysts. *Journal of Finance*, *47*(5), 1811-1836.

Sullivan, R., Timmermann, A., & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance*, 54(5), 1647–1691.

Timmermann, A. (2006). Forecast combinations. *Handbook of Economic Forecasting*, *1*, 135-196.

Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value? *Journal of Finance*, *51*(1), 137–167.

Xiong, X., Meng, Y., Joseph, N. L., & Shen, D. (2020). Stock mispricing, hard-to-value stocks and the influence of internet stock message boards. *International Review of Financial An*alysis, 72, 101576.

Yamamoto, R. (2012). Intraday technical analysis of individual stocks on the Tokyo Stock Exchange. *Journal of Banking & Finance*, *36*(11), 3033–3047.

Yan, X., & Zheng, L. (2017). Fundamental analysis and the cross-section of stock returns: A data-mining approach. *Review of Financial Studies*, 30(4), 1382–1423.

Zaremba, A., Umutlu, M., & Maydybura, A. (2020). Where have the profits gone? Market efficiency and the disappearing equity anomalies in country and industry returns. *Journal of Banking & Finance*, 121, 105966.

Zhang, X. F. (2006). Information uncertainty and stock returns. *Journal of Finance*, 61(1), 105-137.

Zhu, Y., & Zhou, G. (2009). Technical analysis: An asset allocation perspective on the use of moving averages. *Journal of Financial Economics*, 92(3), 519–544.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301-320.

Figure 1. Predictive R^2 Coefficients of Models Using Accounting and Market Data

The figure presents the out-of-sample R^2 coefficient (in %) for machine learning models based on market and accounting information. The monthly return predictions are based on the combination model, which aggregates seven individual models: ordinary least squares, partial least squares, least absolute shrinkage and selection operator, elastic net, gradient-boosted regression trees, random forest, and feed-forward neural network. The models are trained using the complete set of 131 stock characteristics from Jensen et al. (2023), or the subsets of the 44 market and 87 accounting variables only. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022.

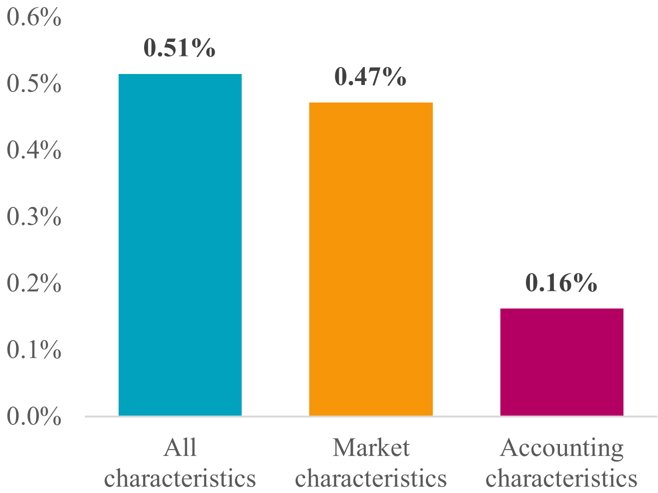
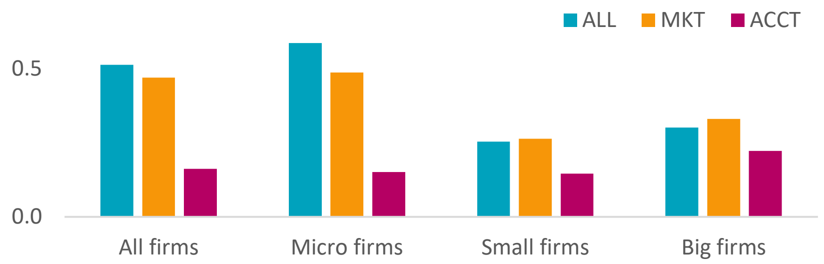


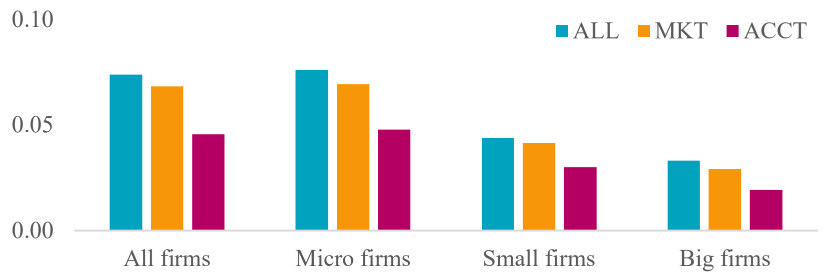
Figure 2. Prediction Performance of Market and Accounting Characteristics

The figure presents the out-of-sample predictive R² coefficients (R^2_{OOS}, Panel A), as well as average cross-sectional Pearson product-moment correlation correlation coefficients (Panel B) and Spearman rank-based correlation coefficients (Panel A). The predictions come from a combination (COMB) forecasting model integrating seven individual machine learning models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are trained using the complete set of 131 stock characteristics from Jensen et al. (2023) (ALL) or the subsets of the 44 market (MKT) and 87 accounting (ACCT) variables. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022. The tests are run separately for the micro firms, comprising the smallest companies representing in total 3% of the aggregate market capitalization; small firms, accounting for the subsequent 7%; and big firms, i.e., the largest companies in the market representing 90% of the total market capitalization.

Panel A: Out-of-sample predictive R² coefficients (R^2_{OOS})



Panel B: Average cross-sectional Pearson product-moment correction coefficient



Panel C: Average cross-sectional Spearman rank-based correction coefficient

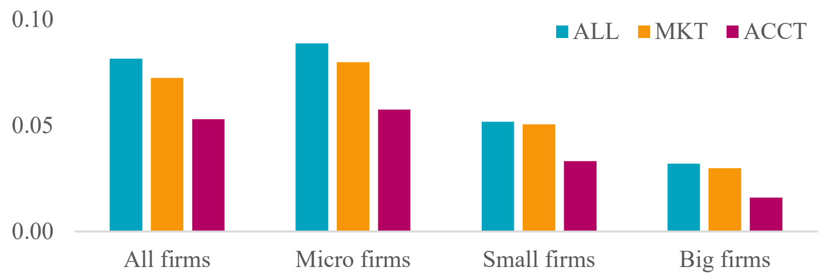


Figure 3. Characteristic Importance per Category

The figure displays the aggregate importance of market and accounting characteristics for the return predictions. The reported values are averages across individual models used in the study. The variable importance (VI) is calculated as the reduction of the overall out-of-sample predictive R^2 resulting from setting all values of a given variable to zero in the training sample. VI is averaged across all the training samples and rescaled to 1. The VIs per category is aggregated as in Bali et al. (2022). The variable classification is available in Appendix A. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022. The tests are run separately for the micro firms, comprising the smallest companies representing in total 3% of the aggregate market capitalization; small firms, accounting for the subsequent 7%; and big firms, i.e., the largest companies in the market representing 90% of the total market capitalization.

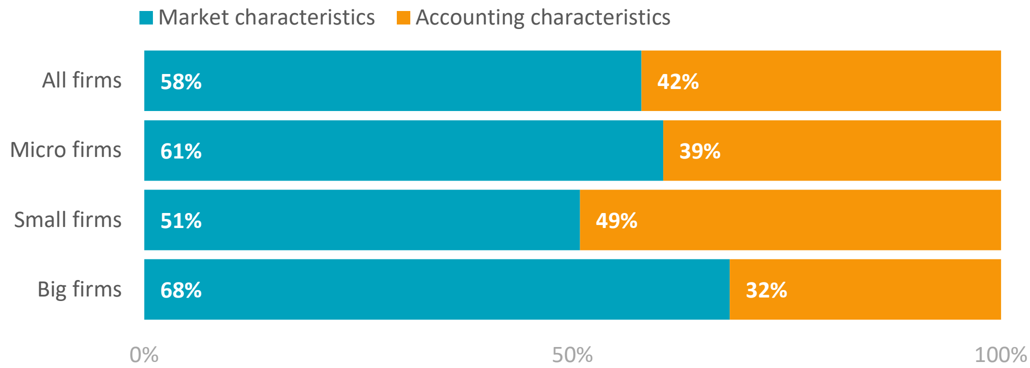
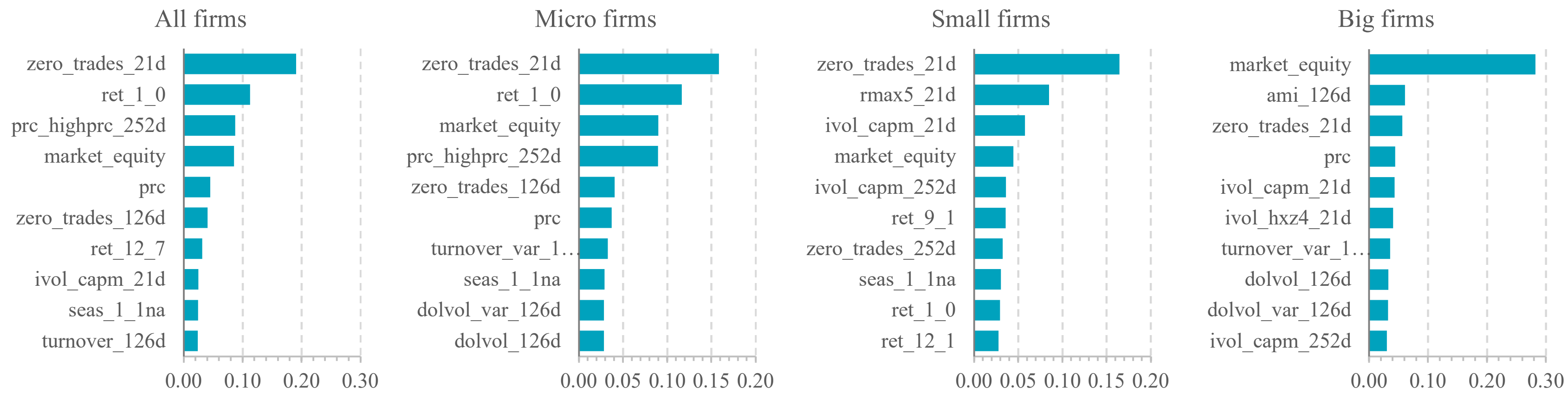


Figure 4. Importance of Top Variables in Different Models

The figure presents the average variable importance of the top 10 characteristics across the individual machine learning models used in the study. The individual exhibits display the reduction in R^2 from setting all values of a given variable to zero in the training sample. The numbers are averaged across all the training samples, and the values for all considered variables are rescaled to sum to 1. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022. The tests are run separately for the micro firms, comprising the smallest companies representing in total 3% of the aggregate market capitalization; small firms, accounting for the subsequent 7%; and big firms, i.e., the largest companies in the market representing 90% of the total market capitalization. Panels A and B concentrate on the models trained using market and accounting characteristics only. The variable classification is available in Appendix A.

Panel A: Models trained based on market characteristics



Panel B: Models trained based on accounting characteristics

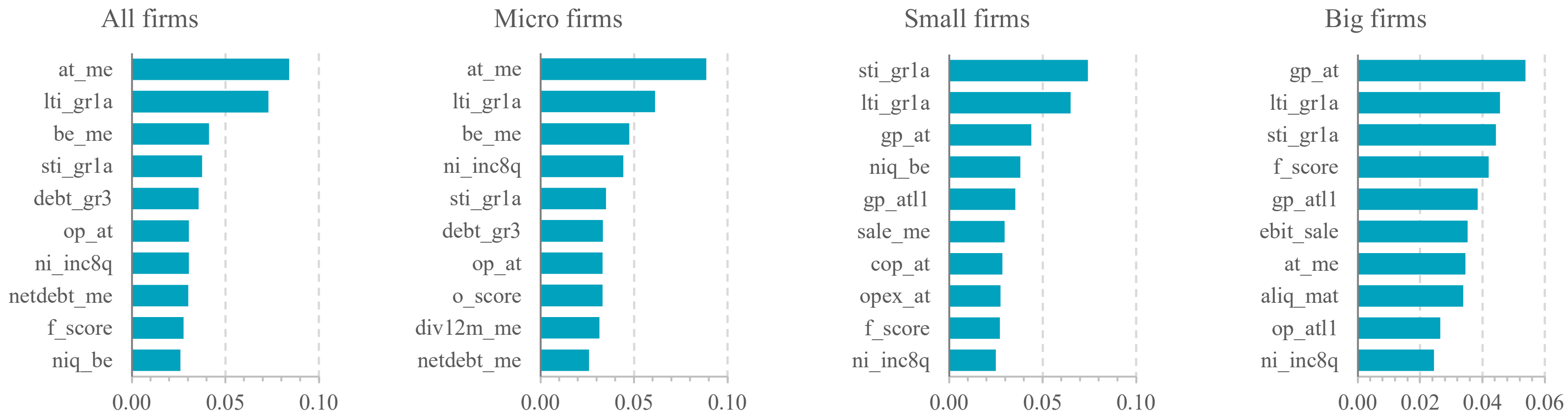


Figure 5. Predictive R^2 Coefficients for the Sample Excluding Low-Priced Stocks

The figure presents the out-of-sample predictive R^2 coefficients (R^2_{OOS}). The predictions come from a combination (COMB) forecasting model integrating seven individual machine learning models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are trained using the complete set of 131 stock characteristics from Jensen et al. (2023) (ALL) or the subsets of the 44 market (MKT) and 87 accounting (ACCT) variables. The sample comprises NYSE, AMEX, and NASDAQ firms, and excludes all stocks with the share price below $1 at the end of last month. The study period begins in January 1972, and the testing period runs from January 1987 to December 2022. The tests are run separately for the micro firms, comprising the smallest companies representing in total 3% of the aggregate market capitalization; small firms, accounting for the subsequent 7%; and big firms, i.e., the largest companies in the market representing 90% of the total market capitalization.

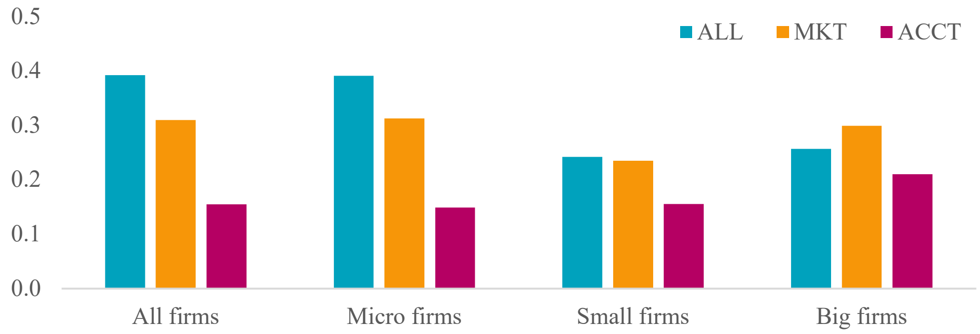


Figure 6. Sharpe Ratios of Long-Short Portfolios

The figure presents the annualized Sharpe ratios on long-short strategies based on machine learning models. The strategies buy (sell) a decile of stocks with the highest (lowest) expected return. The return predictions are made based on eight different models: ordinary least squares (OLS), partial least squares (PLS), least absolute shrinkage and selection operator (LASSO), elastic net (ENET), gradient boosted regression trees (GBRT), random forest (RF), feed-forward neural network (FFNN), and forecast combination (COMB). The models are trained using the complete set of 131 stock characteristics from Jensen et al. (2023) (ALL) or the subsets of the 44 market (MKT) and 87 accounting (ACCT) variables only. The portfolios are value-weighted and rebalanced monthly. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022.

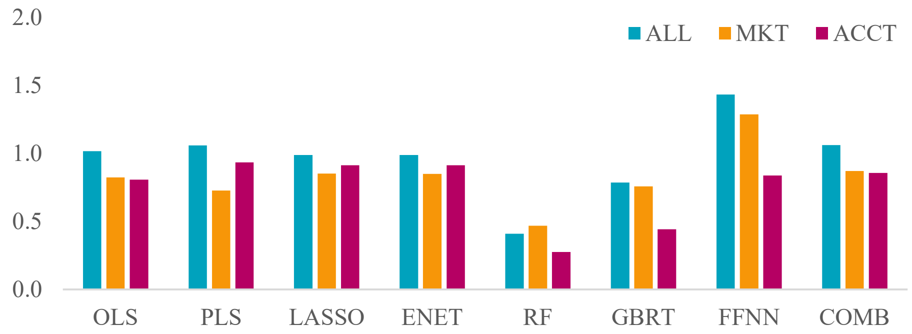
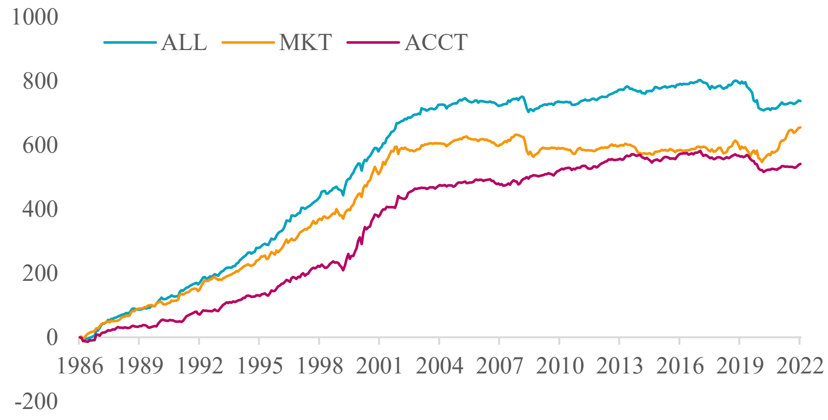


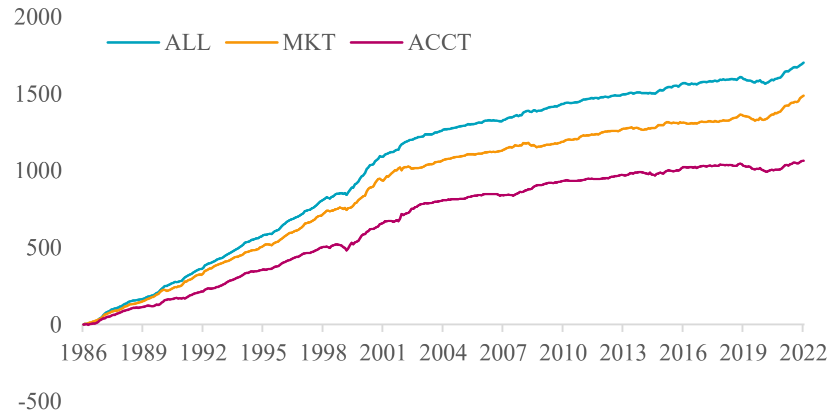
Figure 7. Cumulative Returns over Time

The figure presents cumulative returns on long-short strategies based on machine learning models. The strategies buy (sell) a decile of stocks with the highest (lowest) expected return. The predictions come from a combination (COMB) forecasting model integrating seven individual machine learning models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are trained using the complete set of 131 stock characteristics from Jensen et al. (2023) (ALL) or the subsets of the 44 market (MKT) and 87 accounting (ACCT) variables only. The portfolios are value-weighted and rebalanced monthly. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022. The returns are ported for all stocks in the sample (Panel A), as well as for subgroups of micro, small, and big firms (Panels B, C, and D, respectively. All the tests are run separately for the micro firms, comprising the smallest companies representing in total 3% of the aggregate market capitalization, small firms, accounting for the subsequent 7%, and big firms, i.e., the largest companies in the market representing 90% of the total market capitalization.

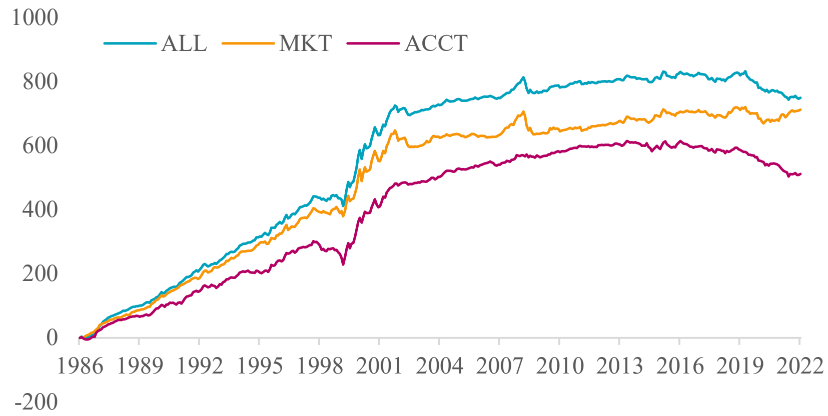
Panel A: All firms



Panel B: Micro firms



Panel C: Small firms



Panel D: Big firms

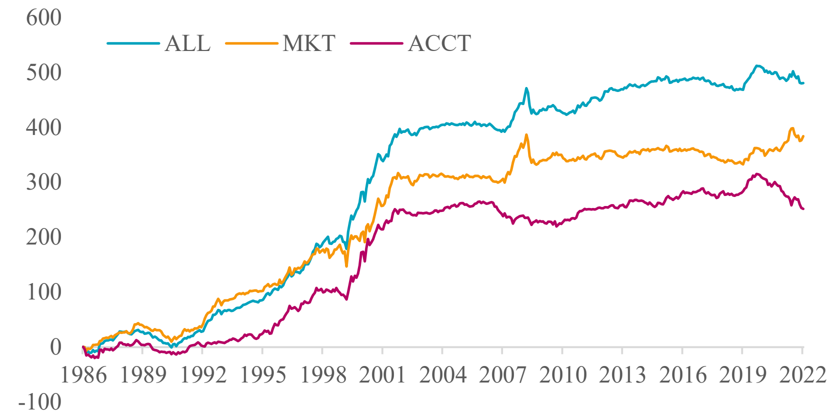


Figure 8. Sharpe Ratios in International Markets

The figure presents the annualized Sharpe ratios on long-short portfolios formed on predictions using different types of information implemented in international stock markets. The return predictions are based on a forecast combination model (COMB), which aggregates seven individual models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are supplied with three types of stock characteristics: all features from Jensen et al. (2022) (ALL), the subset of features based on market data only (MKT), and features using accounting data (ACCT). The strategies buy (sell) a decile of stocks with the highest (lowest) expected return. The portfolios are value-weighted and rebalanced monthly. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022.

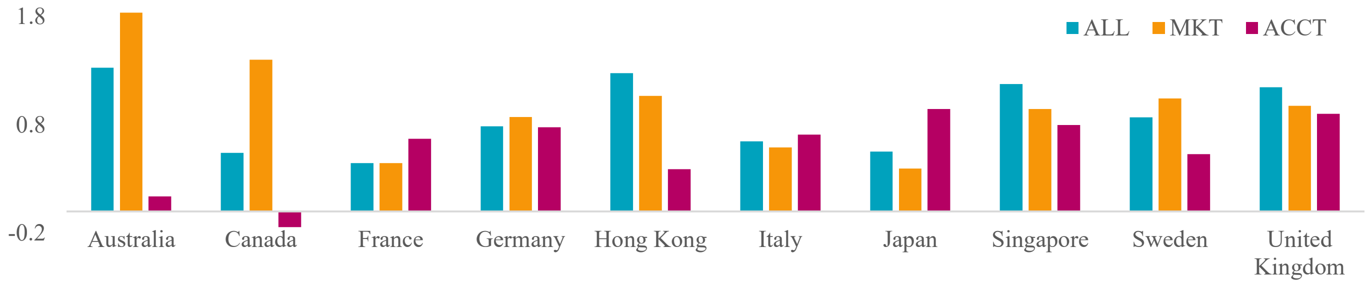


Table 1. Summary Statistics for the Research Sample

The table presents the structure of the research sample size. *Big firms* are the largest companies, which account for 90% of the total market capitalization; *Small firms* represent the subsequent 7% of the market capitalization; *Micro firms* are the smallest stocks, accounting for the remaining 3% of the market. The exhibit displays the average firm size (in USD million), the average number of companies, and the average total market capitalization of a given size class (USD billion). The values in italics illustrate the percentage structure. The sample comprises NYSE, AMEX, and NASDAQ firms, and the presented values concern the testing period from January 1987 to December 2022.

	All firms	Micro firms	Small firms	Big firms
Average firm size [USD million]	3,314	158	1,187	16,199
Average number of firms	5,549	3,456	1,101	992
[%]		*62*	*20*	*18*
Average total market value [USD billion]	18,388	548	1,307	16,070
[%]		*3*	*7*	*90*

Table 2. Returns on Portfolios Based on Market and Accounting Information

The table presents the performance of decile portfolios based on predictions using different types of information. The predictions come from a combination (COMB) forecasting model integrating seven individual machine learning models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are supplied with three types of stock characteristics: 131 features from Jensen et al. (2022) (ALL), the subset of 44 features based on market data only (MKT), and 87 features using accounting data (ACCT). *High* (*Low*) denotes the portfolio with the highest (lowest) predicted return, and *High-Low* indicates the long-long-short portfolio that buys (sells) the top (bottom) decile. The table reports mean monthly returns (R), monthly return standard deviation (SD), annualized Sharpe ratio, the alpha from the Fama and French (2018) six-factor model (α), and the average market value of a company in the portfolio. The portfolios are value-weighted (Panels A and B, respectively) and rebalanced monthly. All returns and alphas are expressed in percentage terms. The values in parentheses are Newey and West's (1987) adjusted *t*-statistics. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022.

	Panel B.1: All variables (ALL)					Panel B.2: Market variables (MKT)					Panel B.3: Accounting variables (ACCT)				
	R	SD	SR	α	MV	R	SD	SR	α	MV	R	SD	SR	α	MV
Low	-0.28	7.56	-0.13	-0.78	1.16	-0.33	7.84	-0.14	-0.76	1.18	0.11	7.36	0.05	-0.37	1.42
2	0.18	5.79	0.11	-0.32	2.36	0.30	6.13	0.17	-0.20	2.39	0.35	5.41	0.23	-0.23	2.90
3	0.44	5.21	0.29	-0.15	3.19	0.54	5.63	0.33	-0.09	3.17	0.52	4.96	0.37	-0.10	3.79
4	0.54	4.88	0.39	-0.04	3.83	0.59	5.12	0.40	0.00	3.72	0.73	4.70	0.54	0.10	4.49
5	0.68	4.56	0.52	0.01	4.14	0.58	4.81	0.42	-0.06	4.08	0.69	4.66	0.52	0.02	4.63
6	0.71	4.65	0.53	-0.02	4.29	0.76	4.71	0.56	0.09	4.25	0.74	4.69	0.54	0.02	4.35
7	0.87	4.65	0.65	0.12	4.19	0.82	4.60	0.62	0.13	4.32	0.85	4.63	0.64	0.16	3.92
8	0.92	4.80	0.67	0.19	4.06	0.70	4.83	0.50	-0.12	4.10	0.82	4.67	0.61	0.06	3.62
9	1.10	5.08	0.75	0.24	3.66	0.99	4.95	0.69	0.23	3.68	0.91	5.21	0.60	0.08	2.82
High	1.43	6.10	0.81	0.56	2.26	1.19	5.33	0.77	0.43	2.24	1.36	6.21	0.76	0.48	1.20
High-Low	1.71	5.57	1.06	1.34		1.52	6.04	0.87	1.20		1.25	5.06	0.86	0.85	
	(5.06)			*(5.45)*		*(4.46)*			*(4.91)*		*(4.57)*			*(4.11)*	

Table 3. Returns on Portfolios Based on Market and Accounting Information across Firm Size

The table reports the mean monthly returns on long-short portfolios based on predictions using different types of information. The strategies buy (sell) a decile of stocks with the highest (lowest) expected return. The return predictions are based on a forecast combination model (COMB), which aggregates seven individual models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are supplied with three types of stock characteristics: all 131 features from Jensen et al. (2022) (ALL), the subset of 44 features based on market data only (MKT), and 87 features using accounting data (ACCT). *High* (*Low*) denotes the portfolio with the highest (lowest) predicted return, and *High-Low* indicates the long-long-short portfolio that buys (sells) the top (bottom) decile. α is the alpha from the six-factor model of Fama and French (2018). The portfolios are value-weighted and rebalanced monthly. All returns and alphas are expressed in percentage terms. The values in parentheses are Newey and West's (1987) adjusted t-statistics. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022. All the tests are run separately for the micro firms, comprising the smallest companies representing in total 3% of the aggregate market capitalization, small firms, accounting for the subsequent 7%, and big firms, i.e., the largest companies in the market representing 90% of the total market capitalization.

	Micro-caps			Small-caps			Big-caps		
	ALL	MKT	ACCT	ALL	MKT	ACCT	ALL	MKT	ACCT
Low	-1.63	-1.45	-0.80	-0.40	-0.33	-0.18	0.02	0.19	0.20
2	-0.25	-0.19	0.04	0.09	0.28	0.20	0.35	0.46	0.56
3	0.25	0.39	0.36	0.26	0.45	0.59	0.52	0.60	0.66
4	0.42	0.59	0.52	0.66	0.56	0.75	0.66	0.59	0.67
5	0.73	0.76	0.65	0.81	0.70	0.80	0.62	0.64	0.75
6	0.86	0.83	0.85	0.85	0.90	0.82	0.79	0.76	0.68
7	1.11	1.09	0.99	0.93	0.83	0.87	0.76	0.79	0.76
8	1.28	1.22	1.01	1.06	1.03	0.96	0.86	0.88	0.72
9	1.58	1.25	1.26	1.19	1.06	1.01	0.87	0.91	0.70
High	2.31	1.99	1.67	1.33	1.32	1.01	1.13	1.08	0.78
High-Low	3.94	3.44	2.47	1.73	1.65	1.18	1.11	0.89	0.58
	(10.91)	*(10.10)*	*(8.31)*	*(4.52)*	*(4.56)*	*(3.67)*	*(3.97)*	*(3.33)*	*(2.45)*
α	3.63	3.25	2.26	1.72	1.59	1.18	0.98	0.62	0.59
	(16.36)	*(14.39)*	*(11.53)*	*(6.77)*	*(6.58)*	*(5.26)*	*(4.51)*	*(2.75)*	*(3.05)*

Table 4. Decile Portfolios Based on Market and Accounting Information: Exclusion of Low-Priced Stocks

The table presents the performance of decile portfolios based on predictions using different types of information. The predictions come from a combination (COMB) forecasting model integrating seven individual machine learning models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are supplied with three types of stock characteristics: 131 features from Jensen et al. (2022) (ALL), the subset of 44 features based on market data only (MKT), and 87 features using accounting data (ACCT). *High* (*Low*) denotes the portfolio with the highest (lowest) predicted return, and *High-Low* indicates the long-long-short portfolio that buys (sells) the top (bottom) decile. The table reports mean monthly returns (R), monthly return standard deviation (SD), annualized Sharpe ratio, the alpha from the Fama and French (2018) six-factor model (α), and the average market value of a company in the portfolio. The portfolios are value-weighted and rebalanced monthly. All returns and alphas are expressed in percentage terms. The values in parentheses are Newey and West's (1987) adjusted *t*-statistics. The sample comprises NYSE, AMEX, and NASDAQ firms, and excludes all stocks with the share price below $1 at the end of last month. The study period begins in January 1972, and the testing period runs from January 1987 to December 2022.

	Panel B.1: All variables (ALL)					Panel B.2: Market variables (MKT)					Panel B.3: Accounting variables (ACCT)				
	R	SD	SR	α	MV	R	SD	SR	α	MV	R	SD	SR	α	MV
Low	-0.61	7.94	-0.27	-1.06	0.86	-0.69	9.06	-0.26	-1.13	0.79	-0.12	7.74	-0.05	-0.60	0.86
2	0.18	6.18	0.10	-0.33	2.04	0.24	6.69	0.12	-0.28	1.80	0.29	5.98	0.17	-0.21	2.15
3	0.32	5.41	0.21	-0.24	3.13	0.50	5.94	0.29	-0.12	2.86	0.46	5.39	0.30	-0.18	3.20
4	0.49	4.84	0.35	-0.12	3.96	0.70	5.29	0.46	0.13	3.87	0.57	4.80	0.41	-0.10	3.84
5	0.76	4.59	0.57	0.14	4.33	0.60	4.87	0.43	-0.03	4.42	0.80	4.66	0.59	0.13	4.23
6	0.75	4.69	0.56	0.08	4.49	0.70	4.67	0.52	0.07	4.73	0.73	4.59	0.55	0.09	4.46
7	0.75	4.64	0.56	-0.01	4.64	0.68	4.55	0.52	-0.02	4.78	0.76	4.59	0.57	0.04	4.66
8	0.89	4.69	0.65	0.12	4.43	0.78	4.63	0.58	0.00	4.73	0.83	4.74	0.60	0.11	4.57
9	1.01	4.99	0.70	0.21	4.08	0.94	4.80	0.68	0.17	4.34	0.87	4.74	0.63	0.09	4.36
High	1.31	5.50	0.82	0.43	3.04	1.08	5.23	0.72	0.30	2.67	1.12	5.07	0.76	0.31	2.66
High-Low	1.92	5.66	1.18	1.49		1.77	7.04	0.87	1.43		1.24	5.50	0.78	0.91	
	(6.11)			*(6.80)*		*(4.47)*			*(5.63)*		*(4.16)*			*(4.51)*	

Table 5. Returns on Portfolios Based on Individual Machine Learning Models

The table reports the returns on long-short portfolios based on predictions using different types of information. The strategies buy (sell) a decile of stocks with the highest (lowest) expected return based on one of seven different models: ordinary least squares (OLS), partial least squares (PLS), least absolute shrinkage, and selection operator (LASSO), elastic net (ENET), gradient boosted regression trees (GBRT), random forest (RF), feed-forward neural network (FFNN). The models are supplied with three types of stock characteristics: all 131 features from Jensen et al. (2022) (ALL), the subset of 44 features based on market data only (MKT), and 87 features using accounting data (ACCT). R denotes the mean monthly return, and α is the alpha from the six-factor model of Fama and French (2018). The portfolios are value-weighted and rebalanced monthly. All returns and alphas are expressed in percentage terms. The values in parentheses are Newey and West's (1987) adjusted t-statistics. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022.

	Mean returns (R)			Six-factor model alphas (α)		
	ALL	MKT	ACCT	ALL	MKT	ACCT
OLS	1.35 (4.59)	1.20 (4.17)	1.02 (4.26)	1.05 (4.93)	0.92 (4.06)	0.67 (3.66)
PLS	1.45 (5.26)	1.09 (3.69)	1.12 (5.00)	1.05 (4.62)	0.70 (3.25)	0.76 (4.36)
LASSO	1.33 (4.60)	1.23 (4.28)	1.11 (4.65)	1.00 (4.68)	0.94 (4.29)	0.78 (4.41)
ENET	1.33 (4.62)	1.23 (4.27)	1.11 (4.65)	1.00 (4.69)	0.94 (4.29)	0.78 (4.42)
RF	0.85 (2.12)	1.01 (2.50)	0.47 (1.58)	0.47 (1.85)	1.01 (3.40)	0.15 (0.76)
GBRT	1.29 (4.43)	1.66 (4.20)	0.69 (2.23)	0.90 (4.26)	1.30 (4.57)	0.42 (2.01)
FFNN	2.36 (7.26)	2.31 (7.06)	1.14 (4.35)	2.12 (8.40)	2.09 (7.45)	0.91 (4.41)

Table 6. Returns on Portfolios Based Market and Accounting Information in Subperiods

The table reports the returns on long-short portfolios based on predictions using different types of information. The strategies buy (sell) a decile of stocks with the highest (lowest) expected return. The return predictions are based on a forecast combination model (COMB), which aggregates seven individual models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are supplied with three types of stock characteristics: all 131 features from Jensen et al. (2022) (ALL), the subset of 44 features based on market data only (MKT), and 87 features using accounting data (ACCT). R in Panel A denotes the mean monthly return, and α in Panel B is the alpha from the six-factor model of Fama and French (2018). The portfolios are value-weighted and rebalanced monthly. All returns and alphas are expressed in percentage terms. The values in parentheses are Newey and West's (1987) adjusted t-statistics. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022. All the tests are run separately for the micro firms, comprising the smallest companies representing in total 3% of the aggregate market capitalization, small firms, accounting for the subsequent 7%, and big firms, i.e., the largest companies in the market representing 90% of the total market capitalization. The results are reported for the entire testing period, as well as for equal subperiods from January 1987 to December 2004 and January 2005 to December 2022.

	ALL			MKT			ACC		
	Micro	Small	Big	Micro	Small	Big	Micro	Small	Big
Panel A: Mean returns									
1987-2022	3.94	1.73	1.11	3.44	1.65	0.89	2.47	1.18	0.58
	(10.91)	*(4.52)*	*(3.97)*	*(10.10)*	*(4.56)*	*(3.33)*	*(8.31)*	*(3.67)*	*(2.45)*
1987-2004	5.86	3.36	1.88	4.95	2.89	1.45	3.74	2.33	1.15
	(13.32)	*(6.08)*	*(4.55)*	*(11.81)*	*(5.64)*	*(3.94)*	*(9.15)*	*(4.59)*	*(3.13)*
2005-2022	2.02	0.11	0.35	1.94	0.41	0.33	1.20	0.04	0.01
	(5.81)	*(0.29)*	*(1.05)*	*(4.83)*	*(0.97)*	*(0.90)*	*(3.89)*	*(0.14)*	*(0.05)*
Panel B. Alphas									
1987-2022	3.63	1.72	0.98	3.25	1.59	0.62	2.26	1.18	0.59
	(16.36)	*(6.77)*	*(4.51)*	*(14.39)*	*(6.58)*	*(2.75)*	*(11.53)*	*(5.26)*	*(3.05)*
1987-2004	5.58	3.04	1.50	4.86	2.64	0.86	3.48	2.09	0.93
	(18.82)	*(8.23)*	*(4.74)*	*(15.54)*	*(7.06)*	*(2.53)*	*(12.49)*	*(6.93)*	*(3.68)*
2005-2022	1.72	0.31	0.34	1.68	0.48	0.22	1.04	0.17	0.18
	(6.49)	*(1.10)*	*(1.31)*	*(5.83)*	*(1.72)*	*(0.79)*	*(4.41)*	*(0.63)*	*(0.76)*

Table 7. Trading Costs of Strategies Based Market and Accounting Information

The table presents the impact of trading costs on portfolio returns based on predictions using different types of information. The long-only strategies, subscripted LO, buy a decile of stocks with the highest expected return. The long-short strategies, superscripted LS, additionally sell a decile of stocks with the lowest expected return. The return predictions are based on a forecast combination model (COMB), which aggregates seven individual models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are supplied with three types of stock characteristics: all 131 features from Jensen et al. (2022) (ALL), the subset of 44 features based on market data only (MKT), and 87 features using accounting data (ACCT). The portfolios are value-weighted and rebalanced monthly. R is the mean monthly returns. $TURN$ is the portfolio turnover, interpreted as the average portfolio share replaced each month. BE denotes the breakeven trading costs at which the mean return equals zero. The turnover ratios and mean returns are reported in percentages, while the trading costs are expressed in basis points. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022. All the tests are run separately for the micro firms, comprising the smallest companies representing in total 3% of the aggregate market capitalization, small firms, accounting for the subsequent 7%, and big firms, i.e., the largest companies in the market representing 90% of the total market capitalization.

	All firms			Micro firms			Small firms			Big firms		
	ALL	MKT	ACCT	ALL	MKT	ACCT	ALL	MKT	ACCT	ALL	MKT	ACCT
Panel A: Long-short portfolios												
R_{LS} (%)	1.71	1.52	1.25	3.94	3.44	2.47	1.73	1.65	1.18	1.11	0.89	0.58
$TURN_{LS}$ (%)	184.88	188.08	173.75	161.36	172.00	126.73	124.12	143.99	87.94	162.27	170.47	135.93
BE_{LS} (bp)	46	40	36	122	100	97	70	57	67	34	26	21
Panel B: Long-only portfolios												
R_{LO} (%)	1.43	1.19	1.36	2.31	1.99	1.67	1.33	1.32	1.01	1.13	1.08	0.78
$TURN_{LO}$ (%)	94.18	95.62	87.79	82.06	87.23	66.28	66.59	75.69	45.66	83.98	87.06	72.75
BE_{LO} (bp)	76	62	78	141	114	126	100	87	110	67	62	54

Table 8. Market Versus Accounting Strategies: Mean-Variance Spanning Test

The table presents the results of mean-variances spanning regressions portfolios based on market and accounting strategies on each other. The strategies buy (sell) a decile of stocks with the highest (lowest) expected return. The return predictions are based on a forecast combination model (COMB), which aggregates seven individual models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are supplied with three types of stock characteristics: all 131 features from Jensen et al. (2022) (ALL), the subset of 44 features based on market data only (MKT), and 87 features using accounting data (ACCT). All portfolios are value-weighted and rebalanced monthly. We regress the returns on strategies formed on ALL, MKT, and ACCT on the MKT and ACCT portfolio returns, as indicated in each panel's heading. α denotes the regression intercept, and β_{MKT} and β_{ACCT} are measures of MKT and ACCT exposures. The controls include the returns on the six Fama and French (2018) factors: MKT, SMB, HML, RMW, CMA, and MOM. R^2 is the adjusted coefficient of determination. α *and* R^2 are expressed in percentage terms. The values in parentheses are Newey and West's (1987) adjusted *t*-statistics. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022. All the tests are run separately for the micro firms, comprising the smallest companies representing in total 3% of the aggregate market capitalization, small firms, accounting for the subsequent 7%, and big firms, i.e., the largest companies in the market representing 90% of the total market capitalization.

	All firms		Micro firms		Small firms		Big firms	
	Panel A: Regressions of MKT on ACCT							
α	0.97	0.97	2.77	2.77	0.70	0.94	0.60	0.44
	(3.49)	*(4.01)*	*(10.31)*	*(10.85)*	*(2.81)*	*(4.40)*	*(2.46)*	*(2.00)*
β_{ACCT}	0.44	0.26	0.27	0.22	0.80	0.54	0.49	0.31
	(8.13)	*(4.74)*	*(5.49)*	*(3.91)*	*(18.01)*	*(12.12)*	*(9.06)*	*(5.56)*
Controls	No	Yes	No	Yes	No	Yes	No	Yes
R^2	13.13	39.50	6.33	26.52	42.87	62.97	15.83	38.37
	Panel B: Regressions of ACCT on MKT							
α	0.79	0.62	1.65	1.73	0.30	0.43	0.29	0.45
	(3.36)	*(3.00)*	*(6.15)*	*(7.38)*	*(1.43)*	*(2.12)*	*(1.48)*	*(2.39)*
β_{MKT}	0.31	0.19	0.24	0.16	0.54	0.47	0.32	0.22
	(8.13)	*(4.74)*	*(5.49)*	*(3.91)*	*(18.01)*	*(12.12)*	*(9.06)*	*(5.56)*
Controls	No	Yes	No	Yes	No	Yes	No	Yes
R^2	13.13	38.18	6.33	36.52	42.87	52.00	15.83	32.19
	Panel C: Regressions of ALL on ACCT							
α	0.83	0.73	2.09	1.96	0.55	0.64	0.69	0.59
	(3.89)	*(3.65)*	*(11.13)*	*(10.16)*	*(3.28)*	*(4.15)*	*(3.67)*	*(3.32)*
β_{ACCT}	0.70	0.72	0.75	0.74	1.00	0.92	0.73	0.66
	(17.25)	*(15.65)*	*(21.36)*	*(17.79)*	*(33.28)*	*(28.51)*	*(17.29)*	*(14.94)*
Controls	No	Yes	No	Yes	No	Yes	No	Yes
R^2	40.76	51.70	51.36	55.59	71.97	79.49	40.87	52.94
α	*Panel D: Regressions of ALL on MKT*							
α	0.77	0.68	1.55	1.38	0.31	0.35	0.44	0.52
	(2.19)	*(3.20)*	*(7.65)*	*(7.18)*	*(2.17)*	*(2.29)*	*(3.18)*	*(3.70)*
β_{MKT}	0.62	0.55	0.70	0.69	0.86	0.86	0.76	0.73
	(18.79)	*(13.56)*	*(21.22)*	*(20.47)*	*(41.36)*	*(29.41)*	*(30.91)*	*(24.22)*
Controls	No	Yes	No	Yes	No	Yes	No	Yes
R^2	44.96	46.83	51.05	61.00	79.86	80.32	68.88	69.87
	Panel E: Regressions of ALL on MKT and ACCT							
α	0.37	0.30	0.56	0.32	0.15	0.09	0.31	0.31
	(2.19)	*(1.75)*	*(4.36)*	*(2.35)*	*(1.68)*	*(0.98)*	*(2.86)*	*(2.80)*
β_{MKT}	0.47	0.44	0.55	0.59	0.57	0.58	0.63	0.62
	(16.03)	*(12.99)*	*(26.64)*	*(25.99)*	*(32.73)*	*(28.26)*	*(29.19)*	*(25.57)*
β_{ACCT}	0.50	0.60	0.60	0.61	0.54	0.60	0.42	0.47
	(14.43)	*(15.11)*	*(26.78)*	*(23.28)*	*(25.46)*	*(27.37)*	*(15.66)*	*(16.32)*
Controls	No	Yes	No	Yes	No	Yes	No	Yes
R^2	62.86	65.38	81.63	82.86	91.96	92.88	80.16	81.47

Table 9. Bivariate Portfolio Sorts on Market and Accounting Predictors

The table presents the returns on bivariate portfolio sorts on predictions using types of information. The predictions come from a combination (COMB) forecasting model integrating seven individual machine learning models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are supplied with two types of stock characteristics: 131 features from Jensen et al. (2022) (ALL), the subset of 44 features from Jensen et al. (2022) based on market data only (MKT), and 87 features from the same source using accounting data (ACCT). The stocks are independently sorted into quintiles on predictions based on MKT and ACCT, and *High* (*Low*) denotes the quintile with the highest (lowest) predicted return. *Average* is the average return across all five quintiles. *High-Low* indicates the long-long-short portfolio that buys (sells) the top (bottom) decile. α is the alpha from the Fama and French (2018) six-factor model. The portfolios are value-weighted and rebalanced monthly. All returns and alphas are expressed in percentage terms. The values in parentheses are Newey and West's (1987) adjusted t-statistics. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022. All the tests are run separately for the total sample of firms (Panel A), micro firms, comprising the smallest companies representing in total 3% of the aggregate market capitalization (Panel B), small firms, accounting for the subsequent 7% (Panel C), and big firms, i.e., the largest companies in the market representing 90% of the total market capitalization (Panel D).

Panel A: All firms

	MKT quintiles						MKT strategies	
	Low	2	3	4	High	Average	High-Low	Alpha
ACCT quintiles								
Low	-0.45	0.22	0.24	0.57	0.58	0.23	1.03 *(2.97)*	0.66 *(2.07)*
2	0.09	0.50	0.66	0.71	0.95	0.58	0.86 *(2.98)*	0.70 *(2.63)*
3	0.24	0.59	0.77	0.70	1.15	0.69	0.91 *(3.45)*	0.76 *(3.12)*
4	0.41	0.85	0.71	0.89	1.12	0.79	0.71 *(2.82)*	0.39 *(1.83)*
High	0.69	0.91	0.92	1.05	1.39	0.99	0.70 *(2.26)*	0.57 *(1.93)*
Average	0.20	0.61	0.66	0.78	1.04		0.84 *(3.60)*	0.61 *(3.04)*
Long-short LM strategies								
High-Low	1.14	0.69	0.68	0.49	0.82	0.76		
	(3.46)	*(2.11)*	*(2.95)*	*(2.69)*	*(3.70)*	*(4.12)*		
Alpha	0.73	0.51	0.38	0.18	0.63	0.48		
	(2.51)	*(2.20)*	*(1.83)*	*(0.97)*	*(2.68)*	*(3.48)*		

Panel B: Micro firms

	MKT quintiles						MKT strategies	
	Low	2	3	4	High	Average	High-Low	Alpha
ACCT quintiles								
Low	-1.76	-0.12	0.03	0.47	0.73	-0.13	2.49 *(9.18)*	2.26 *(9.38)*
2	-0.71	0.44	0.74	0.90	1.17	0.51	1.87 *(5.70)*	1.80 *(6.46)*
3	-0.25	0.50	0.87	1.20	1.51	0.77	1.77 *(6.43)*	1.61 *(6.05)*
4	-0.29	0.82	1.07	1.40	1.75	0.95	2.04 *(7.27)*	1.98 *(7.41)*
High	0.15	1.20	1.40	1.72	2.28	1.35	2.14 *(7.69)*	2.03 *(8.05)*
Average	-0.57	0.57	0.82	1.14	1.49		2.06 *(8.21)*	1.94 *(8.94)*
Long-short LM strategies								
High-Low	1.90	1.33	1.37	1.25	1.55	1.48		
	(7.18)	*(5.31)*	*(5.57)*	*(5.65)*	*(6.56)*	*(7.36)*		
Alpha	1.64	1.19	1.23	1.04	1.42	1.30		
	(6.85)	*(5.69)*	*(5.90)*	*(5.02)*	*(6.22)*	*(7.99)*		

Panel C: Small firms

	MKT quintiles						MKT strategies	
	Low	2	3	4	High	Average	High-Low	Alpha
ACCT quintiles								
Low	-0.49	-0.04	0.46	0.52	0.73	0.24	1.22 *(3.29)*	1.15 *(3.49)*
2	0.12	0.52	0.74	0.83	1.15	0.67	1.03 *(3.50)*	0.96 *(3.72)*
3	0.10	0.65	0.79	0.93	1.29	0.75	1.19 *(4.82)*	1.17 *(5.59)*
4	0.37	0.69	0.90	1.15	1.22	0.87	0.85 *(3.19)*	0.85 *(4.18)*
High	0.78	0.73	1.04	1.04	1.20	0.96	0.42 *(1.75)*	0.40 *(1.74)*
Average	0.18	0.51	0.79	0.90	1.12		0.94 *(3.93)*	0.91 *(4.89)*
Long-short LM strategies								
High-Low	1.27	0.78	0.59	0.52	0.47	0.72		
	(3.45)	*(2.99)*	*(2.65)*	*(2.17)*	*(2.17)*	*(3.35)*		
Alpha	1.20	0.93	0.65	0.45	0.46	0.74		
	(3.78)	*(4.34)*	*(3.50)*	*(2.45)*	*(2.16)*	*(4.61)*		

Panel D: Big firms

	MKT quintiles						MKT strategies	
	Low	2	3	4	High	Average	High-Low	Alpha
ACCT quintiles								
Low	-0.02	0.55	0.47	0.62	0.77	0.48	0.79 *(2.83)*	0.72 *(2.63)*
2	0.38	0.58	0.64	0.71	1.01	0.66	0.63 *(2.87)*	0.40 *(2.03)*
3	0.53	0.54	0.66	0.96	1.04	0.75	0.51 *(2.05)*	0.41 *(1.68)*
4	0.57	0.68	0.75	0.75	1.06	0.76	0.48 *(2.07)*	0.31 *(1.39)*
High	0.43	0.56	0.84	0.76	1.09	0.73	0.66 *(2.35)*	0.40 *(1.67)*
Average	0.38	0.58	0.67	0.76	0.99		0.62 *(3.15)*	0.45 *(2.58)*
Long-short LM strategies								
High-Low	0.45	0.01	0.37	0.13	0.32	0.26		
	(1.83)	*(0.05)*	*(2.02)*	*(0.64)*	*(1.78)*	*(1.65)*		
Alpha	0.57	0.03	0.43	0.16	0.25	0.29		
	(2.37)	*(0.15)*	*(2.48)*	*(0.84)*	*(1.49)*	*(2.30)*		

Table 10. Portfolio Sorts: International Evidence

The table presents the performance of long-short portfolios based on predictions using different types of information implemented in international stock markets. The strategies buy (sell) a decile of stocks with the highest (lowest) expected return. The return predictions are based on a forecast combination model (COMB), which aggregates seven individual models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are supplied with three types of stock characteristics: all features from Jensen et al. (2022) (ALL), the subset of features based on market data only (MKT), and features using accounting data (ACCT). The exact number of available features in each country is provided in the table's rightmost section. The reported performance measures include the mean monthly return and alphas from the six-factor model of Fama and French (2018), both expressed in percentage terms. The portfolios are value-weighted and rebalanced monthly. The values in parentheses are Newey and West's (1987) adjusted t-statistics. The bottom row reports cross-country averages along t-statistics calculated as in Amihud et al. (2015). The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022.

	Mean returns			Alphas			# characteristics		
	ALL	MKT	ACCT	ALL	MKT	ACCT	ALL	MKT	ACCT
Australia	2.54 (6.19)	3.69 (7.46)	0.21 (0.56)	2.31 (5.18)	3.39 (7.08)	0.18 (0.48)	105	40	65
Canada	0.98 (2.45)	3.14 (6.18)	-0.18 (-0.62)	0.96 (2.30)	3.17 (6.34)	-0.39 (-1.30)	105	29	76
France	0.77 (1.92)	0.77 (1.92)	0.96 (2.81)	0.85 (2.26)	0.85 (2.26)	0.98 (2.90)	113	42	71
Germany	1.67 (3.13)	2.01 (3.26)	1.10 (3.07)	1.10 (2.11)	1.34 (2.42)	0.84 (2.43)	112	42	70
Hong Kong	2.42 (4.86)	2.22 (3.74)	0.66 (1.33)	2.13 (4.87)	1.73 (4.24)	0.59 (1.57)	84	42	42
Italy	1.58 (2.91)	1.43 (2.20)	1.54 (3.15)	1.45 (2.89)	1.20 (2.60)	1.37 (2.81)	78	42	36
Japan	0.60 (2.15)	0.43 (1.58)	0.96 (4.46)	0.54 (2.05)	0.51 (1.94)	0.80 (3.82)	115	44	71
Singapore	1.63 (5.34)	1.58 (3.89)	1.28 (3.83)	1.68 (4.95)	1.70 (4.96)	1.19 (3.04)	103	33	70
Sweden	1.98 (3.17)	2.47 (3.45)	1.12 (2.25)	1.90 (4.20)	2.25 (5.03)	1.33 (2.69)	99	39	60
United Kingdom	2.08 (4.00)	1.90 (2.89)	1.44 (3.21)	2.00 (5.49)	1.64 (4.75)	1.32 (4.00)	115	37	78
Average	1.62 (7.70)	1.97 (6.25)	0.91 (5.29)	1.49 (7.79)	1.78 (6.05)	0.82 (4.56)	103	39	64

Table 11. Valuation Difficulty and the Strategies Based on Market and Accounting Information

The table presents the returns on portfolio strategies based on different types of information implemented in stocks with different levels of valuation difficulty. We independently group stocks into quintiles based on valuation uncertainty and machine learning predictions. The measure of valuation uncertainty aggregates four characteristics: firm age (*age*), quarterly return on assets (*niq_at*), share turnover (*turnover_126d*), and idiosyncratic risk (*ivol_ff3_21d*) derived from the Fama-French three-factor model. The return predictions come from a forecast combination (COMB) model, which integrates seven individual machine learning models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The prediction models are supplied with two types of stock characteristics: the subset of 44 features from Jensen et al. (2022) based on market data only (Panel A, MKT), and 87 features from the same source using accounting data (Panel B, ACCT). The intersection of the two sets of breakpoints generates 25 portfolios double-sorted portfolios. *High E(R)* (*Low E(R)*) denotes the quintile with the highest (lowest) predicted return, and *High-Low* indicates the long-long-short portfolio that buys (sells) the top (bottom) decile. α is the alpha from the Fama and French (2018) six-factor model. The portfolios are value-weighted and rebalanced monthly. All returns and alphas are expressed in percentage terms. The last row report the return differentials between the long-short strategies implemented in the hard-to-value (the 5th quintile) and easy-to-value (the 1st quintile) stocks. The values in parentheses are Newey and West's (1987) adjusted *t*-statistics. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022.

	Low E(R)	2	3	4	High E(R)	H-L	t-stat$_{\text{H-L}}$	α	t-stat$_{\alpha}$
	Panel A: Forecasts based on market data (MKT)								
Easy-to-value	0.21	0.47	0.57	0.69	0.91	0.70	*(2.63)*	0.62	*(2.34)*
2	0.07	0.56	0.78	0.81	1.14	1.07	*(3.46)*	0.96	*(3.16)*
3	0.32	0.39	0.61	0.65	1.22	0.90	*(2.98)*	0.65	*(2.27)*
4	-0.17	0.46	0.57	0.89	1.10	1.27	*(3.83)*	1.27	*(3.94)*
Hard-to-value	-0.66	0.25	0.48	0.87	1.43	2.09	*(4.77)*	1.82	*(4.43)*
Difference-in-differences						1.39	*(3.17)*	1.20	*(2.99)*
	Panel B: Forecasts based on accounting data (ACCT)								
Easy-to-value	0.29	0.50	0.79	0.77	0.89	0.59	*(2.71)*	0.39	*(1.80)*
2	0.46	0.65	0.78	0.88	1.09	0.63	*(3.03)*	0.35	*(1.62)*
3	0.31	0.44	0.83	0.87	1.25	0.94	*(3.29)*	0.80	*(2.65)*
4	0.12	0.38	0.65	1.00	1.02	0.90	*(3.47)*	0.83	*(2.85)*
Hard-to-value	-0.27	0.44	0.52	0.68	0.57	0.84	*(2.65)*	0.64	*(1.86)*
Difference-in-differences						0.25	*(0.76)*	0.25	*(0.66)*

Table 12. The Strategies Based on Market and Accounting Information and the Time-Varying Market Volatility

The table reports the returns on long-short portfolios based on predictions using different types of information in the regimes of high and low market volatility. The strategies buy (sell) a decile of stocks with the highest (lowest) expected return. The return predictions are based on a forecast combination model (COMB), which aggregates seven individual models: OLS, PLS, LASSO, ENET, RF, GBRT, and FFNN. The models are supplied with three types of stock characteristics: all 131 features from Jensen et al. (2022) (ALL), the subset of 44 features based on market data only (MKT), and 87 features using accounting data (ACCT). The alphas come from the six-factor model of Fama and French (2018). The portfolios are value-weighted and rebalanced monthly. The performance is reported separately for periods of high and low volatility, where high (low) indicates months when the average value-weighted share volatility (*rvol_21d*) is above (below) its all-time median. All returns and alphas are expressed in percentage terms. The values in parentheses are Newey and West's (1987) adjusted *t*-statistics. The sample comprises NYSE, AMEX, and NASDAQ firms; the study period begins in January 1972, and the testing period runs from January 1987 to December 2022.

	Low volatility			High volatility		
	ALL	MKT	ACCT	ALL	MKT	ACCT
Mean return	1.28	0.84	0.76	2.56	2.71	1.71
	(5.39)	*(3.12)*	*(3.56)*	*(4.64)*	*(3.80)*	*(3.20)*
Alpha	0.88	0.48	0.69	1.99	2.28	0.98
	(3.49)	*(1.81)*	*(3.17)*	*(5.54)*	*(5.28)*	*(2.87)*

Appendix A. Stoch Characteristics Based on Market and Accounting Data

The table lists 131 stock characteristics from Jensen et al. (2023) used in our study. Panel A displays the 87 variables that require accounting data to calculate them, while Panel B presents the 44 characteristics based only on market data. The variables' details and the literature references are available from Jensen et al. (2023). The table spans two pages.

Panel A: 87 characteristics based on accounting data

Symbol	Characteristic	Symbol	Characteristic
aliq_at	Liquidity of book assets	lnoa_gr1a	Change in long-term net operating assets
aliq_mat	Liquidity of market assets	lti_gr1a	Change in long-term investments
at_be	Book leverage	ncoa_gr1a	Change in noncurrent operating assets
at_gr1	Asset Growth	ncol_gr1a	Change in noncurrent operating liabilities
at_me	Assets-to-market	netdebt_me	Net debt-to-price
at_turnover	Capital turnover	nfna_gr1a	Change in net financial assets
be_gr1a	Change in common equity	ni_ar1	Earnings persistence
be_me	Book-to-market equity	ni_be	Return on equity
bev_mev	Book-to-market enterprise value	ni_inc8q	# consecutive quart. with earnings increases
capex_abn	Abnormal corporate investment	ni_ivol	Earnings volatility
capx_gr1	CAPEX growth (1 year)	ni_me	Earnings-to-price
capx_gr2	CAPEX growth (2 years)	niq_at	Quarterly return on assets
capx_gr3	CAPEX growth (3 years)	niq_be	Quarterly return on equity
cash_at	Cash-to-assets	niq_su	Standardized earnings surprise
chcsho_12m	Net stock issues	nncoa_gr1a	Change in net noncurrent operating assets
coa_gr1a	Change in current operating assets	noa_at	Net operating assets
col_gr1a	Change in current operating liabilities	noa_gr1a	Change in net operating assets
cop_at	Cash-based operating profits-to-book assets	o_score	Ohlson O-score
cop_atl1	Cash-based oper. profits-to-lagged book assets	oaccruals_at	Operating accruals
cowc_gr1a	Change in current operating working capital	oaccruals_ni	Percent operating accruals
dbnetis_at	Net debt issuance	ocf_at	Operating cash flow to assets
debt_gr3	Growth in book debt (3 years)	ocf_at_chg1	Change in operating cash flow to assets
debt_me	Debt-to-market	ocf_me	Operating cash flow-to-market
dgp_dsale	Change gross margin minus change sales	op_at	Operating profits-to-book assets
div12m_me	Dividend yield	op_atl1	Operating profits-to-lagged book assets
dsale_dinv	Change sales minus change Inventory	ope_be	Operating profits-to-book equity
dsale_drec	Change sales minus change receivables	ope_bel1	Operating profits-to-lagged book equity
dsale_dsga	Change sales minus change SG&A	opex_at	Operating leverage
earnings_variability	Earnings variability	pi_nix	Taxable income-to-book income
ebit_bev	Return on net operating assets	ppeinv_gr1a	Change PPE and Inventory
ebit_sale	Profit margin	sale_bev	Assets turnover
ebitda_mev	Ebitda-to-market enterprise value	sale_emp_gr1	Labor force efficiency
emp_gr1	Hiring rate	sale_gr1	Sales Growth (1 year)
eq_dur	Equity duration	sale_gr3	Sales Growth (3 years)
eqnpo_12m	Equity net payout	sale_me	Sales-to-market
f_score	Pitroski F-score	saleq_gr1	Sales growth (1 quarter)
fcf_me	Free cash flow-to-price	saleq_su	Standardized Revenue surprise
fnl_gr1a	Change in financial liabilities	sti_gr1a	Change in short-term investments
gp_at	Gross profits-to-assets	taccruals_at	Total accruals
gp_atl1	Gross profits-to-lagged assets	taccruals_ni	Percent total accruals
inv_gr1	Inventory growth	tangibility	Asset tangibility
inv_gr1a	Inventory change	tax_gr1a	Tax expense surprise
ival_me	Intrinsic value-to-market	z_score	Altman Z-score
kz_index	Kaplan-Zingales index		

Panel B: 44 characteristics based on market data only

Symbol	Characteristic	Symbol	Characteristic
age	Firm age	resff3_6_1	Residual momentum t-6 to t-1
ami_126d	Amihud Measure	ret_1_0	Short-term reversal
beta_60m	Market Beta	ret_12_1	Price momentum t-12 to t-1
beta_dimson_21d	Dimson beta	ret_12_7	Price momentum t-12 to t-7
betabab_1260d	Frazzini-Pedersen market beta	ret_3_1	Price momentum t-3 to t-1
betadown_252d	Downside beta	ret_6_1	Price momentum t-6 to t-1
bidaskhl_21d	The high-low bid-ask spread	ret_60_12	Long-term reversal
corr_1260d	Market correlation	ret_9_1	Price momentum t-9 to t-1
coskew_21d	Coskewness	rmax1_21d	Maximum daily return
dolvol_126d	Dollar trading volume	rmax5_21d	Highest 5 days of return
dolvol_var_126d	Coefficient of variation for dollar trading volume	rmax5_rvol_21d	Highest 5 days of return scaled by vol.
iskew_capm_21d	Idiosyncratic skewness from the CAPM	rskew_21d	Total skewness
iskew_ff3_21d	Idios. skew. from the Fama-French 3-factor model	rvol_21d	Return volatility
iskew_hxz4_21d	Idiosyncratic skewness from the q-factor model	seas_1_1an	Year 1-lagged return, annual
ivol_capm_21d	Idiosyncratic volatility from the CAPM (21 days)	seas_1_1na	Year 1-lagged return, nonannual
ivol_capm_252d	Idiosyncratic volatility from the CAPM (252 days)	seas_2_5an	Years 2-5 lagged returns, annual
ivol_ff3_21d	Idios. vol. from the Fama-French 3-factor model	seas_2_5na	Years 2-5 lagged returns, nonannual
ivol_hxz4_21d	Idiosyncratic volatility from the q-factor model	turnover_126d	Share turnover
market_equity	Market Equity	turnover_var_126d	Coefficient of variation for share turnover
prc	Price per share	zero_trades_126d	Number of zero trades (6 months)
prc_highprc_252d	Current price to high price over last year	zero_trades_21d	Number of zero trades (1 month)
resff3_12_1	Residual momentum t-12 to t-1	zero_trades_252d	Number of zero trades (12 months)