

Real-time Machine Learning in the Cross-Section of Stock Returns*

Bin Li
Wuhan University

Alberto Rossi
Georgetown University

Xuemin (Sterling) Yan
Lehigh University

Lingling Zheng
Renmin University of China

September 2, 2022

Abstract

Recent studies document strong performance for machine-learning-based investment strategies. These strategies use anomaly variables discovered ex-post as predictors of stock returns and may not be implementable in real time. We construct real-time machine learning strategies based on a “universe” of fundamental signals. While positive and significant, the out-of-sample performance of these strategies is significantly weaker than those documented by prior studies, especially in value-weighted portfolios. We find similar results when examining a “universe” of past return-based signals. Our results offer a more tempered view of the economic gains associated with machine learning strategies relative to prior literature.

*Bin Li is at the Economics and Management School, Wuhan University, China. Email: binli.whu@whu.edu.cn. Alberto Rossi is at the McDonough School of Business, Georgetown University, U.S.A. Email: agr60@georgetown.edu. Xuemin (Sterling) Yan is at the College of Business, Lehigh University, U.S.A. Email: xuy219@lehigh.edu. Lingling Zheng is at the School of Business, Renmin University of China, China. Email: zhenglingling@rmba.ruc.edu.cn. We thank Marcin Kacperczyk, David Solomon, Allan Timmermann, and seminar and conference participants at the Chinese Finance Annual Meeting, Duke Kunshan University, Hunan University, Lehigh University, Southwest Jiaotong University, Sun Yat-Sen University, Taiwan Finance Association Asset Pricing Conference, and Tongji University for helpful comments and discussions. Bin Li acknowledges financial support from the National Natural Science Foundation of China (No. 71971164). Lingling Zheng acknowledges financial support from the National Natural Science Foundation of China (No. 72122021). All errors are our own.

1 Introduction

Machine learning methods have received considerable attention in the recent asset pricing literature, particularly in the area of return prediction (see, e.g., [Chen, Pelger, and Zhu \(2020\)](#), [Freyberger, Neuhierl, and Weber \(2020\)](#), [Gu, Kelly, and Xiu \(2020\)](#) and [Leippold, Wang, and Zhou \(2021\)](#)). The general conclusions of the existing studies are remarkably similar—machine learning models are superior to traditional models in predicting the cross-section of stock returns. Given the inherent focus of machine learning methods on *out-of-sample* prediction, these studies leave readers with the impression that machine learning methods routinely lead to large improvements in investment performance. Indeed, a common theme among many existing studies is to construct long-short investment strategies based on machine learning forecasts and demonstrate that these strategies are highly profitable.

While prior studies have clearly established the potential for large economic gains to investors using machine learning forecasts, an important issue that has not been fully addressed in the literature is the real-time implementability of machine learning strategies. Specifically, existing studies use published anomaly variables as predictors of stock returns¹ and implicitly assume that they are known to investors at the beginning of the training period, even though most anomalies are discovered years or decades later. While this approach is very natural if the objective is to measure risk premium or estimate the stochastic discount factor, in which case we can take an econometrician’s perspective and analyze data ex-post, such an approach raises the issue of whether the resulting machine learning strategies could have been implemented by real-time investors. Take the asset growth anomaly of [Cooper, Gulen, and Schill \(2008\)](#) as an example. It is unlikely that investors would have been able to single out asset growth as a stock return predictor before that research was published in the mid-2000s. Moreover, published anomalies tend to exhibit strong in-sample performance, in part because of the publication bias ([Harvey, Liu, and Zhu, 2016](#)). Therefore, the economic gains from using machine learning forecasts documented by prior studies are potentially overstated for real-time investors.

In this paper, we examine machine learning strategies based on a “universe” of over 18,000 fundamental signals. Because these signals are constructed from financial statement variables using permutational arguments ([Yan and Zheng, 2017](#)), our strategies are implementable in real-time. In essence, we are trying to imitate the search process of real-time investors. Moreover, examining a universe of fundamental signals, rather than selecting a subset of them based on whether they have

¹There are exceptions. For example, [Martin and Nagel \(2021\)](#) study the in-sample and out-of-sample predictability of an exhaustive list of past return-based predictors.

been published in academic journals, allows us to side-step the issue of data mining and look-ahead bias. As a consequence, our strategies should more accurately reflect the economic benefit of using machine learning forecasts for real-time investors.

We focus on fundamental signals for several reasons. First, fundamental analysis dates back at least to [Graham and Dodd \(1934\)](#), so it is natural to expect investors to consider fundamental signals as a potential class of return predictors (i.e., no look-ahead bias). Second, financial economists have long emphasized the importance of economic intuition behind predictors of expected returns ([Fama and French, 1996](#); [Cochrane, 2011](#)), especially in a machine learning environment ([Arnott, Harvey, and Markowitz, 2019](#)). Fundamental signals are inherently related to firm cash flows and valuations and therefore have stronger economic foundations than most other classes of predictors. Third, one can construct a “universe” of fundamental signals using permutational arguments ([Yan and Zheng, 2017](#)). This is important because real-time investors have no way of knowing which signals turn out to be significant ex-post, so they have to learn from a universe.

Our analysis is also motivated by [Martin and Nagel \(2021\)](#). Standard asset pricing models assume rational expectations, i.e., investors know the model or the data generating process. [Martin and Nagel \(2021\)](#) argue that, in the age of Big Data, this is unrealistic and that investors face a high-dimensional prediction problem instead. In their model, investors use shrinkage methods to learn from a large number of potential predictor variables in real time. They show that in-sample predictability arises almost surely in their setting, even in the absence of data mining or market inefficiency. In contrast, there is no out-of-sample predictability because investors optimally use the information available to them in real-time. Therefore, a central prediction of [Martin and Nagel \(2021\)](#) is that there should be a substantial wedge between in-sample and out-of-sample predictability. We follow [Martin and Nagel \(2021\)](#) and analyze data from the perspective of real-time investors. Moreover, we test their prediction by comparing the in-sample and out-of-sample performance of a universe of fundamental signals.

The primary machine learning method we use is Boosted Regression Trees (BRT). We focus on BRT for several reasons. First, previous studies have shown that BRT exhibit strong predictive performance in finance applications. [Gu, Kelly, and Xiu \(2020\)](#), for example, show that BRT and neural networks are the two best-performing machine learning methods in predicting stock returns. Second, BRT are ideally suited for handling large, high-dimensional data sets because of their computational efficiency. This is important for us because our predictor set, which contains more than 18,000 signals, is much

larger than those examined by previous studies. Third, BRT are robust to missing values, outliers, and the addition of irrelevant input variables. Finally, BRT are not “black boxes” as many other machine learning methods and they are instead known for their interpretability.

We follow [Gu, Kelly, and Xiu \(2020\)](#) and partition our sample period 1963-2019 into a training period, a cross-validation period, and an out-of-sample test period. We form long-short portfolios based on machine learning predicted returns, i.e., buying stocks with high predicted returns and shorting stocks with low predicted returns. We find that the equal-weighted long-short portfolio generates an average return of 0.92% per month (t -statistic=6.33) and a Sharpe ratio of 0.92 during the out-of-sample period 1987-2019. The performance of the value-weighted long-short portfolio is much weaker, earning an average return of 0.40% per month (t -statistic=2.36) and exhibiting a Sharpe ratio of 0.29.

The long-short returns and Sharpe ratios for our machine learning strategies, although statistically significant, are considerably lower than those documented by prior studies. [Gu, Kelly, and Xiu \(2020\)](#), for example, show that the long-short portfolios formed based on neural network forecasts earn an average return of 3.27% per month and a Sharpe ratio of 2.45 in equal-weighted portfolios and an average return of 2.12% per month, and a Sharpe ratio of 1.35 in value-weighted portfolios. Similarly, [Chen, Pelger, and Zhu \(2020\)](#) and [Freyberger, Neuhierl, and Weber \(2020\)](#) report that the hedge portfolios constructed based on their models deliver an out-of-sample Sharpe ratio of 2.6 and 2.75, respectively. Thus, compared to the previous literature, our results indicate that the economic gains to real-time investors from using machine learning forecasts are much more modest.

Institutional investors are more likely to have the resources and sophistication to use machine learning methods. Previous studies (e.g., [Gompers and Metrick \(2001\)](#)) have shown that institutional investors prefer large, liquid stocks because these stocks are more investable. To evaluate whether our machine learning strategies are profitable among large stocks, we repeat our analysis for subsamples of stocks sorted by firm size. We find that the out-of-sample performance of our machine learning strategies is statistically significant among small stocks but only marginally significant and, in some cases, insignificant among large stocks. The weak evidence of out-of-sample predictability among large stocks suggests that the economic benefit of using machine learning forecasts may be even more limited for institutional investors.

The relatively weak performance of our machine learning strategies is not specific to BRT. We find even weaker evidence of out-of-sample predictability using neural network forecasts. Specifically,

while the long-short returns are generally significant in equal-weighted portfolios, they are insignificant in value-weighted portfolios. Consistent with [Gu, Kelly, and Xiu \(2020\)](#), we find that shallow learning performs better than deep learning in neural networks. We also repeat our analysis using a rolling-window approach instead of a recursive one. If the relations between fundamental signals and future stock returns are unstable over time, then the rolling-window approach should perform better. Contrary to this argument, we find that our machine learning strategies perform slightly worse under the rolling-window approach than under the recursive-window approach.

One might argue that the modest performance of our machine learning strategies is due to our large universe of fundamental signals. In particular, if most of these signals are uninformative about future stock returns, machine learning strategies based on our universe could be sub-optimal. As noted earlier, BRT are robust to the addition of irrelevant predictor variables, so this is not a big concern for us. Nevertheless, to examine whether the performance of our machine learning strategies is hampered by the large size of our predictor set, we construct various subsets of our universe based on the prominence of the underlying accounting variables. Overall, consistent with our expectation, we find no evidence that machine learning strategies based on smaller universes of fundamental signals perform significantly better.

In contrast to the modest out-of-sample predictability, our fundamental signals exhibit strong in-sample predictability. For ease of comparison with the out-of-sample analysis, we perform our in-sample analysis using the same machine learning method (i.e., BRT) and for the same sample period (i.e., 1987-2019). We find that the equal-weighted long-short return is over 3% per month, and the value-weighted long-short return is more than 2%, both of which are highly statistically significant. These in-sample long-short returns are several times larger than their out-of-sample counterparts. In addition, we find that the in-sample R^2 ranges from 1.21% to 1.65%, while the out-of-sample R^2 is only 0.28%. As predicted by [Martin and Nagel \(2021\)](#), we find a substantial wedge between in-sample and out-of-sample predictability.

Our analyses so far have focused on fundamental signals. The main reason for this focus is that we can construct a “universe” of fundamental signals ([Yan and Zheng, 2017](#)). Past return-based signals are another important class of signals that is well suited for our analysis. In particular, we follow [Martin and Nagel \(2021\)](#) and construct a universe of technical signals based on past 120 months’ monthly stock returns and monthly squared returns. Investors have long used fundamental and technical analysis to assess firm value and predict stock returns. By also examining technical signals, we are

able to study the two most important classes of return predictors that real-time investors could have considered decades ago.

As in our analysis of fundamental signals, we continue to use BRT as the primary machine learning method. We find that the machine-learning strategy based on technical signals earns an average return of 0.94% per month (t -statistic=6.33) and exhibits a Sharpe ratio of 0.97 in equal-weighted portfolios. The performance of value-weighted portfolios is significantly weaker. The average long-short return is 0.56% per month (t -statistic=3.34), while the Sharpe ratio is 0.45. The results based on neural network forecasts are much worse. The long-short returns for shallow learning strategies are marginally significant, while those for deep learning strategies are completely insignificant, regardless of whether we examine equal- or value-weighted portfolios. Overall, our analyses based on technical signals paint a similar picture to that based on fundamental signals. That is, the performance of real-time machine learning strategies is positive and significant, but economically and statistically weaker than those reported by the prior literature, particularly when we examine value-weighted portfolios or use neural networks to generate return forecasts.

We emphasize that our study is not intended as a direct comparison with previous studies. We use BRT as the primary machine learning method and neural networks as an alternative. In comparison, [Gu, Kelly, and Xiu \(2020\)](#) examine a comprehensive set of machine learning methods including but not limited to BRT and neural networks, while [Freyberger, Neuhierl, and Weber \(2020\)](#) and [Chen, Pelger, and Zhu \(2020\)](#) use group LASSO and generative adversarial networks (GAN), respectively. In addition, our strategy focuses on fundamental signals and technical signals, due primarily to our desire to construct a universe, while prior studies also include other types of predictor variables. These methodological and data differences might have contributed to the performance difference between our machine learning strategies and those examined by prior studies.² It is also important to note that although the performance of our strategies is significantly weaker than those documented by prior literature, these real-time machine learning strategies are nevertheless profitable, particularly in equal-weighted portfolios, suggesting that real-time investors do benefit from using machine learning forecasts.

Our paper builds on and contributes to the recent literature employing machine learning methods in empirical asset pricing. [Gu, Kelly, and Xiu \(2020\)](#) use machine learning methods to measure risk

²We believe the most crucial difference between our approach and prior methodologies is that prior studies use anomaly variables discovered ex-post as predictors of stock returns, whereas we include in our predictor set a universe of fundamental signals and technical signals that are known ex-ante. Our machine learning strategies are not only implementable in real-time but also free from data-mining concerns or look-ahead biases.

premium and show that machine learning models, particularly trees and neural networks, significantly outperform linear regression models in predicting stock returns.³ [Chen, Pelger, and Zhu \(2020\)](#) estimate the SDF by using deep learning neural networks and show that their model outperforms all other benchmark models. [Freyberger, Neuhierl, and Weber \(2020\)](#) use the adaptive group LASSO for model selection and show that their model exhibits superior out-of-sample performance. [Kozak, Nagel, and Santosh \(2020\)](#) use shrinkage and selection methods to construct an SDF that summarizes the joint explanatory power of a large cross-section of return predictors.⁴ These studies have established the potential for large economic gains to investors using machine learning strategies. We complement the existing studies by taking the perspective of real-time investors. We contribute to the literature by constructing real-time-implementable machine learning strategies and showing that they are less profitable than those considered by prior literature.

Our paper is closely related to [Avramov, Kaplanski, and Subrahmanyam \(2021\)](#). They use machine learning methods to construct a fundamental index from 105 signals that are based on deviations of accounting variables from their recent averages and show that this fundamental deviation index significantly predicts future stock returns. [Avramov, Kaplanski, and Subrahmanyam \(2021\)](#) differ from our study in several ways. First, they use quarterly accounting data and their sample period is correspondingly shorter. Second, they use LASSO and regression regularization methods, whereas we focus on BRT and neural networks. Third and more importantly, they examine a specific category of fundamental signals, which does not include popular financial ratios such as return on assets, asset turnover, and liquidity ratio. It is not obvious that real-time investors ex-ante would be able to single out this particular category of fundamental signals while ignoring the other categories in their predictor set.

Our paper is also related to [Avramov, Cheng, and Metzker \(2020\)](#), who show that much of the profitability of machine learning-based investment strategies is derived from difficult-to-arbitrage stocks or during periods when limits-to-arbitrage are elevated. Our finding that the out-of-sample predictability is significantly weaker among large stocks is consistent with [Avramov, Cheng, and Metzker \(2020\)](#). Limits-to-arbitrage, however, is not our primary focus. Our main argument is that machine learning strategies that use subsequently discovered anomaly variables as predictors, including those considered

³The objective of our study is not to compare machine learning models with traditional linear regression models. Rather, our focus is on the *real-time* performance of machine learning strategies.

⁴For additional studies that use machine learning methods in asset pricing, please also see, e.g., [Rapach, Strauss, and Zhou \(2013\)](#), [Chinco, Clark-Joseph, and Ye \(2019\)](#), [Feng, Polson, and Xu \(2020\)](#), [Bryzgalova, Pelger, and Zhu \(2020\)](#), [Bianchi, Büchner, and Tamoni \(2021\)](#), [Cong, Tang, Wang, and Zhang \(2021\)](#), and [Leippold, Wang, and Zhou \(2021\)](#).

by [Avramov, Cheng, and Metzker \(2020\)](#), may not be implementable in real-time.

Finally, our paper is related to [Arnott, Harvey, and Markowitz \(2019\)](#) and [Israel, Kelly, and Moskowitz \(2020\)](#), who caution that machine learning methods may not work as well in finance as in some other disciplines. In particular, machine learning methods face three significant challenges in finance applications: the lack of data (on the time series dimension), the low signal-to-noise ratio, and the adaptive nature of financial markets. The modest performance of our real-time machine learning strategies could be a manifestation of these challenges faced by market professionals and investors.

The rest of our paper proceeds as follows. Section 2 describes our data, sample, and methods. Section 3 and Section 4 present the empirical results. Section 6 concludes.

2 Data, Sample and Methods

This section describes the stock sample and associated fundamental signals we employ in our main analysis. We then describe the cross-sectional prediction problem underlying the portfolio strategies we generate and the main empirical method we use—Boosted Regression Trees (BRT). Third, we describe how we implement our machine learning strategy.

2.1 Stock Sample and Associated Fundamental Signals

We obtain monthly stock returns, share price, SIC code, and shares outstanding from the Center for Research in Security Prices (CRSP) and annual accounting data from Compustat. Our sample consists of the NYSE, AMEX, and NASDAQ common stocks (with a CRSP share code of 10 or 11) with the necessary data to construct fundamental signals and compute subsequent stock returns. We exclude financial stocks, i.e., those with a one-digit SIC code of 6. We also remove stocks with a share price lower than \$1. To mitigate backfilling biases, we require that a firm be listed on Compustat for two years before it is included in our sample ([Fama and French, 1993](#)). We obtain [Fama and French \(1996, 2015\)](#) factors and the momentum factor from Kenneth French’s website and [Hou, Xue, and Zhang \(2015\)](#) q -factors from Lu Zhang’s website.⁵ Our sample spans from July 1963 to June 2019, and our sample consists of 15,035 stocks.

We construct the universe of fundamental signals for our sample of stocks following [Yan and](#)

⁵Kenneth French’s data library is located at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. The q -factors can be downloaded from <http://global-q.org/index.html>.

Zheng (2017).⁶ We start with 240 accounting variables (listed in Appendix B) and compute, for each variable, a total of 76 signals (listed in Appendix C). These signals are obtained by taking the original accounting variables and transforming them by computing changes, ratios, and other potentially economically meaningful transformations. The final number of fundamental signals we include in our analysis is 18,113, which is slightly smaller than 18,240 (240×76) because not all combinations of the accounting variables result in meaningful signals, and some of the combinations are redundant. For brevity, we refer the readers to Yan and Zheng (2017) for complete details regarding selecting accounting variables and constructing fundamental signals.

2.2 Methodology

2.2.1 Prediction Equation. We predict the cross-section of stock returns using the following specification:

$$R_{i,t+1} = f(\mathbf{x}_{i,t}|\theta) + \epsilon_{i,t+1}$$

where $R_{i,t+1}$ denotes annual excess return for stock i from July of year t to June of year $t + 1$, $\mathbf{x}_{i,t}$ denotes a vector of variables used to predict the cross section of returns, and θ denotes the parameters for the prediction function f . Stocks are indexed as $i = 1, \dots, N$ and years are indexed by $t = 1, \dots, T$.

The vector of predictive variables includes the 18,113 fundamental signals described earlier. To make sure the accounting information is publicly available to investors, we follow Fama and French (1992) and pair accounting variables in year $t - 1$ with stock returns from July of year t to June of year $t + 1$. We follow Gu, Kelly, and Xiu (2020) and transform all fundamental signals as follows. We first rank all non-missing fundamental signals each year and then scale their ranks to the interval $[-1, +1]$. By construction, the cross-sectional median of the transformed fundamental signals is zero.

We predict annual excess returns for two reasons. First, our fundamental signals are constructed from annual financial statements and are updated annually. Second, the number of signals considered in our study is substantially larger than those in prior studies. Predicting annual returns is computationally more efficient than predicting monthly returns.⁷

⁶We prefer to use an existing universe of fundamental signals instead of constructing one specifically for this study. Chordia, Goyal, and Saretto (2020) extend Yan and Zheng (2017) and construct a universe of over 2 million fundamental signals. We do not use this universe because real-time investors are unlikely to have the computing power to deal with many predictive variables in a machine learning context.

⁷We conduct most of our empirical analyses on a high-performance cluster of 14 computing nodes, each of which is equipped with 128GB of RAM.

2.2.2 Boosted Regression Trees. Our baseline specification includes 18,113 fundamental signals. Due to this high dimensionality, traditional parametric methods tend to overfit the model, while semi- and non-parametric methods suffer from the curse of dimensionality. Hence, we model the prediction function f using machine learning methods. We choose the “off-the-shelf” machine learning tool called Boosted Regression Trees (BRT), in particular, the LightGBM implementation (Ke, Meng, Finley, Wang, Chen, Ma, Ye, and Liu, 2017) for our baseline analysis.

We choose BRT as our primary machine learning method for several reasons. First, BRT routinely rank among the very best machine learning algorithms in both finance and non-finance applications.⁸ Second, BRT can handle large data sets with high dimensionality without overfitting, because they simultaneously perform subsampling, model combination, and shrinkage. Third, BRT are robust to missing values and outliers (Hastie, Tibshirani, and Friedman, 2009). In particular, BRT are invariant under all monotone transformations of the individual input variables, making the forecasts generated robust to extreme values. Fourth, BRT are robust to the addition of irrelevant input variables (Friedman, 2001; Hastie, Tibshirani, and Friedman, 2009), because the underlying Classification and Regression Trees (CART) algorithm is designed to perform variable selection. Finally, because BRT are rooted in the CART framework, they possess good interpretability. For example, BRT return the rank and relative importance of all the potential regressors available, known as relative influence measures.⁹ This feature distinguishes BRT from harder-to-interpret methods such as neural networks.

Regression Trees

A regression tree is built through a process known as binary recursive partitioning, which is an iterative process that splits the data into partitions or branches. Suppose we have P potential predictor (“state”) variables and a single dependent variable over T observations, i.e., (x_t, y_{t+1}) for $t = 1, 2, \dots, T$, with $x_t = (x_{t1}, x_{t2}, \dots, x_{tp})$. Fitting a regression tree requires deciding (i) which predictor variables to use to split the sample space and (ii) which split points to use. The regression trees we use employ recursive binary partitions, so the fit of a regression tree can be written as an additive model:

$$f(x) = \sum_{j=1}^J c_j I\{x \in S_j\},$$

⁸See a list of Machine Learning Challenge winning solutions on the LightGBM’s website at <https://github.com/microsoft/LightGBM/tree/master/examples>.

⁹To conserve space, we provide a description of the relative influence measures in Appendix E. We also implement the relative influence measure on our data and report the results in Appendix E.

where S_j , $j = 1, \dots, J$ are the regions we split the space spanned by the predictor variables into, $I\{\}$ is an indicator variable, and c_j is the constant used to model the dependent variable in each region. If the L^2 norm criterion function is adopted, the optimal constant is $\hat{c}_j = \text{mean}(y_{t+1}|x_t \in S_j)$.

The globally optimal splitting point is difficult to determine, particularly in cases where the number of state variables is large. Hence, we use a sequential greedy algorithm. Using the full set of data, the algorithm considers a splitting variable p and a split point s so as to construct half-planes,

$$S_1(p, s) = \{X|X_p \leq s\} \quad \text{and} \quad S_2(p, s) = \{X|X_p > s\},$$

that minimize the sum of squared residuals:

$$\min_{p,s} \left[\min_{c_1} \sum_{x_t \in S_1(p,s)} (y_{t+1} - c_1)^2 + \min_{c_2} \sum_{x_t \in S_2(p,s)} (y_{t+1} - c_2)^2 \right]. \quad (1)$$

For a given choice of p and s , the fitted values, \hat{c}_1 and \hat{c}_2 , are

$$\begin{aligned} \hat{c}_1 &= \frac{1}{\sum_{t=1}^T I\{x_t \in S_1(p, s)\}} \sum_{t=1}^T y_{t+1} I\{x_t \in S_1(p, s)\}, \\ \hat{c}_2 &= \frac{1}{\sum_{t=1}^T I\{x_t \in S_2(p, s)\}} \sum_{t=1}^T y_{t+1} I\{x_t \in S_2(p, s)\}. \end{aligned} \quad (2)$$

The best splitting pair (p, s) in the first iteration can be determined by searching through each of the predictor variables, $p = 1, \dots, P$. Given the best partition from the first step, the data is then partitioned into two additional states and the splitting process is repeated for each of the subsequent partitions. Predictor variables that are never used to split the sample space do not influence the fit of the model, so the choice of splitting variable effectively performs variable selection.

Regression trees are ideally suited for handling high-dimensional data sets, incorporating multiway interactions among predictors, and capturing non-linear relations between predictors and the predicted variable. However, the approach is sequential, and successive splits are performed on fewer and fewer observations, increasing the risk of fitting idiosyncratic data patterns. Furthermore, there is no guarantee that the sequential splitting algorithm leads to the globally optimal solution. To deal with these problems, we next consider a regularization method known as boosting.

Boosting

Boosting is based on the idea that combining a series of simple prediction models can lead to more accurate forecasts than those available from any individual model. Boosting algorithms iteratively re-weight data used in the initial fit by adding new trees in a way that increases the weight on observations modeled poorly by the existing collection of trees. From above, recall that a regression tree can be written as:

$$\mathcal{T}(x; \{S_j, c_j\}_{j=1}^J) = \sum_{j=1}^J c_j I\{x \in S_j\}. \quad (3)$$

A boosted regression tree is simply the sum of regression trees:

$$f_B(x) = \sum_{b=1}^B \mathcal{T}_b(x; \{S_{b,j}, c_{b,j}\}_{j=1}^J),$$

where $\mathcal{T}_b(x; \{S_{b,j}, c_{b,j}\}_{j=1}^J)$ is the regression tree used in the b -th boosting iteration and B is the number of boosting iterations. Given the model fitted up to the $(b-1)$ -th boosting iteration, $f_{b-1}(x)$, the subsequent boosting iteration seeks to find parameters $\{S_{j,b}, c_{j,b}\}_{j=1}^J$ for the next tree to solve a problem of the form

$$\{\hat{S}_{j,b}, \hat{c}_{j,b}\}_{j=1}^J = \min_{\{S_{j,b}, c_{j,b}\}_{j=1}^J} \sum_{t=0}^{T-1} [y_{t+1} - (f_{b-1}(x_t) + \mathcal{T}_b(x_t; \{S_{j,b}, c_{j,b}\}_{j=1}^J))]^2.$$

For a given set of state definitions (“splits”), $S_{j,b}$, $j = 1, \dots, J$, the optimal constants, $c_{j,b}$, in each state are derived iteratively from the solution to the problem

$$\begin{aligned} \hat{c}_{j,b} &= \min_{c_{j,b}} \sum_{x_t \in S_{j,b}} [y_{t+1} - (f_{b-1}(x_t) + c_{j,b})]^2 \\ &= \min_{c_{j,b}} \sum_{x_t \in S_{j,b}} [e_{t+1,b-1} - c_{j,b}]^2, \end{aligned} \quad (4)$$

where $e_{t+1,b-1} = y_{t+1} - f_{b-1}(x_t)$ is the empirical error after $b-1$ boosting iterations. The solution to this problem is the regression tree that most reduces the average of the squared residuals $\sum_{t=1}^T e_{t+1,b-1}^2$, and $\hat{c}_{j,b}$ is the mean of the residuals in the j th state.

Forecasts are simple to generate from this approach. The boosted regression tree is first estimated using data from $t = 1, \dots, t^*$. Then, the forecast of y_{t^*+1} is based on the model estimates and the

value of the predictor variable at time t^* , x_{t^*} . Boosting makes it more attractive to employ small trees (characterized by few terminal nodes) at each boosting iteration, reducing the risk that the regression trees will overfit. Moreover, by summing over a sequence of trees, boosting performs a type of model averaging that increases the stability and accuracy of the forecasts.

2.3 Implementation

We implement our BRT model by following [Gu, Kelly, and Xiu \(2020\)](#). We divide our sample period (1963-2019) into 12 years of training sample (1963-1974), 12 years of validation sample (1975-1986), and the remaining 33 years (1987-2019) for out-of-sample testing. We begin the out-of-sample period in 1987 in order to align with [Gu, Kelly, and Xiu \(2020\)](#).

We refit our model every year because our fundamental signals are updated annually. Each time we refit the model, we increase the training sample by one year while maintaining the length of the validation period at 12 years. This recursive window approach allows for the incorporation of all available information in generating forecasts. Every year, we generate return forecasts for all the stocks in our sample. We then construct decile portfolios based on the predicted returns. We hold these portfolios for 12 months and rebalance them every year. Our long-short strategy goes long in the decile portfolio with the highest BRT expected returns and short in the decile portfolio with the lowest BRT predicted returns.

To generate return forecasts, we need to estimate the model's parameters using the training data and specify two key hyper-parameters, i.e., the number of boosting iterations and the BRT shrinkage parameter (also known as the learning rate). To choose these two hyper-parameters, we adopt the commonly used grid search with validation procedure ([Hastie, Tibshirani, and Friedman, 2009](#); [Gu, Kelly, and Xiu, 2020](#)).¹⁰ We leave all other tuning parameters at their LightGBM default values.

Specifically, we first use the training sample to estimate the model under each set of hyper-parameter values. We then use the hyper-parameters that show the best performance during the validation period to re-estimate the final model. For example, suppose we want to forecast the cross-section of stock returns for 1987. We fit models under different hyper-parameter values during the training period 1963-1974 and then use the validation period 1975-1986 to gauge the performance of these trained models. We choose the hyper-parameters that deliver the best performance during the validation period and then use these hyper-parameters to re-estimate the final model for the combined

¹⁰Our grid for the number of boosting iterations is $\{100, 250, 500, 750, 1000\}$, while our grid for the learning rate is $\{0.01, 0.05, 0.1\}$.

training and validation period 1963-1986. When we move forward and forecast the cross-section of stock returns for 1988, our validation period rolls forward by one year and stays at 12 years, i.e., 1976-1987, while our training period increases by one year and goes from 1963 to 1975 (13 years).¹¹

Our fundamental signals contain missing values. Although BRT can handle missing values, we pre-process the missing values to make BRT forecasts comparable to other machine learning methods that cannot handle missing values. Specifically, we follow the approach of [Gu, Kelly, and Xiu \(2020\)](#) and replace missing values with the cross-sectional median. Recall that we have normalized all non-missing fundamental signals to the $[-1, +1]$ interval by using their cross-sectional ranks. By construction, the cross-sectional median of the transformed signals is zero. We, therefore, assign all missing values as zero.¹²

Performance Evaluation

Each year we sort all sample stocks into deciles based on BRT predicted returns, construct equal- and value-weighted portfolios, and focus on the long-short strategy that buys stocks in the top decile and shorts stocks in the bottom decile. We estimate CAPM 1-factor, Fama-French 3-factor, Carhart 4-factor, Fama-French 5-factor, Fama-French 5-factor + Momentum factor, and Q factor models by running the following time-series regressions:

$$\begin{aligned}
 r_t &= \alpha + \beta MKT_t + \epsilon_t \\
 r_t &= \alpha + \beta MKT_t + s SMB_t + h HML_t + \epsilon_t \\
 r_t &= \alpha + \beta MKT_t + s SMB_t + h HML_t + u UMD_t + \epsilon_t \\
 r_t &= \alpha + \beta MKT_t + s SMB_t + h HML_t + r RMW_t + c CMA_t + \epsilon_t \\
 r_t &= \alpha + \beta MKT_t + s SMB_t + h HML_t + r RMW_t + c CMA_t + u UMD_t + \epsilon_t \\
 r_t &= \alpha + \beta MKT_t + s SMB_t + r ROE_t + i IA_t + \epsilon_t
 \end{aligned} \tag{5}$$

where r_t is the long-short portfolio return based on BRT-generated forecasts for month t , and MKT, SMB, HML, UMD, RMW, CMA, ROE, and IA are market, size, value, momentum, profitability, investment (FF5), return on equity, and investment (Q) factors ([Carhart, 1997](#); [Fama and French, 2015](#); [Hou, Xue, and Zhang, 2015](#)). We focus on the alpha estimates and their t -statistics estimated using [Newey and West \(1987\)](#) standard errors.

¹¹We show in Section 4.4 that our main results are robust to alternative training and validation periods.

¹²The performance of the BRT portfolios is similar without pre-processing the missing values.

3 Main Results

In this section, we report the main results of our paper. We start by reporting in Section 3.1 the baseline results that compute the out-of-sample realized returns for BRT portfolios. We then report in Section 3.2 the abnormal performance of the BRT portfolios that control for various risk factors. Finally, in Section 3.3 we compute the in-sample performance of BRT portfolios and then compare it with the out-of-sample performance.

3.1 Baseline Results

Table 1 shows the results of our baseline analysis. As stated earlier, we sort stocks into deciles each year based on one-year-ahead BRT predicted returns. We then construct a long-short portfolio that buys stocks with the highest BRT predicted returns and sells stocks with the lowest BRT predicted returns. We track the performance of these portfolios for 12 months. Following Gu, Kelly, and Xiu (2020) we report in Table 1 the BRT predicted returns (i.e., the sorting variable), the average realized returns, the standard deviation of realized returns, and the Sharpe ratios of BRT-sorted portfolios.

The left panel of Table 1 focuses on equally weighted portfolios. The first column shows the BRT predicted return, which is by construction monotonically increasing from decile 1 (-0.05% per month) to decile 10 (1.69% per month). The second column reports the out-of-sample average realized return for each portfolio: our primary variable of interest. We find that the performance of BRT portfolios increases nearly monotonically from decile 1 (0.09%) to decile 10 (1.01%). The long-short portfolio earns an average return of 0.92% per month (or 11.04% per year), with a highly significant t -statistic of 6.33.

The standard deviation of the realized returns is U-shaped across the BRT decile portfolios, i.e., the portfolios with the lowest and the highest BRT predicted returns have higher volatilities than the other portfolios. Not surprisingly, we find that the long-short portfolio has a much lower volatility than the long-only portfolios. Finally, the last column of the left panel reports the annualized Sharpe ratio, which ranges from 0.04 to 0.63 across the ten BRT decile portfolios. The Sharpe ratio of the long-short portfolio is much higher at 0.92, which is primarily driven by the lower volatility of the long-short portfolio.

The empirical asset pricing literature has gradually moved away from equal-weighted portfolios and towards value-weighted portfolios (Hou, Xue, and Zhang, 2020) because equally weighted portfolios

tend to overweight small-cap stocks that can be harder to trade. We report in the right panel of Table 1 the value-weighted portfolio returns. The BRT predicted return is again by construction monotonically increasing from decile 1 (0.00%) to decile 10 (1.61%). More important, the realized average portfolio return also increases from decile 1 (0.39%) to decile 10 (0.79%), although the relation is far from monotonic. The spread between decile 10 and decile 1 is 0.40% per month, or 4.8% per year.¹³ Even though this spread is statistically significant at the 5% level, its magnitude is less than half of the spread for equally weighted portfolios. The Sharpe ratio exhibits a similar pattern, higher for decile 10 (0.45) than for decile 1 (0.21). The Sharpe ratio for the long-short portfolio is underwhelming at 0.29. As a comparison, the Sharpe ratio for the market portfolio over the same period is 0.45. Therefore, the out-of-sample performance of our real-time machine learning strategies is economically modest.

Overall, we show in Table 1 that long-short portfolios formed based on BRT forecasts earn statistically significant returns, especially in equal-weighted portfolios. The magnitude of the long-short performance, however, is much lower than those documented in the prior literature. For example, the BRT models in Gu, Kelly, and Xiu (2020) achieve an equally weighted monthly long-short portfolio return of 2.14% per month and a Sharpe ratio of 1.73. The corresponding numbers for value-weighted portfolios are 0.99% per month and a Sharpe ratio of 0.81.¹⁴ Similarly, Chen, Pelger, and Zhu (2020) report an out-of-sample Sharpe ratio of 2.60 and Freyberger, Neuhierl, and Weber (2020) report that their model delivers an out-of-sample Sharpe ratio of 2.75. To sum up, while our results indicate that machine learning methods show promise in predicting stock returns, they are less extreme than those presented in the prior literature. The main difference between our paper and prior studies is that we employ a universe of fundamental signals that could have been employed *ex-ante* by a real time investor. The conditioning information set we adopt is therefore free from data-mining concerns and look-ahead biases. Once we control for a more realistic information set, our results indicate that the economic gains to real-time investors from using machine learning methods are substantially smaller

¹³These returns are before trading costs. Novy-Marx and Velikov (2016) report that the round-trip transaction cost for typical value-weighted anomaly strategies exceeds 50 basis points. The corresponding number for equal-weighted anomaly strategies is at least two times higher. We use these estimates along with the turnover rate of BRT portfolios to evaluate the profitability of our machine learning strategies after trading costs. We find that the long-short return is about 7.5% per year for equal-weighted BRT portfolios after trading cost, down from 11% per year before cost. The long-short return for value-weighted BRT portfolios is about 3.8% per year after trading cost, down from 4.8% per year before cost. Overall, trading costs further reduce the profitability and significance of machine learning trading strategies. For ease of comparison with prior literature (e.g., Chen, Pelger, and Zhu (2020); Freyberger, Neuhierl, and Weber (2020); Gu, Kelly, and Xiu (2020)), we report before-trading cost performance of our machine learning strategies in all tables.

¹⁴We note that we implement our BRT model using LightGBM, while Gu, Kelly, and Xiu (2020) implement using scikit-learn. When we implement our model using scikit-learn in conjunction with our fundamental signals, we obtain even less significant results than what we currently report in the paper.

than previously documented.

3.2 Controlling for Common Risk Factors

The results in Table 1 do not control for exposures to risk. It could be that the long-short portfolios based on BRT forecasts have positive and significant returns because they are exposed to well-known sources of risk such as value or profitability. Table 2 shows the risk-adjusted performance of our BRT portfolios once we control for risk exposures using the six models described in Section 2.3. Irrespective of whether we use the CAPM model (columns 1-2), the Fama-French 3-factor model (columns 3-4), the Carhart 4-factor model (columns 5-6), the Fama-French 5-factor model (columns 7-8), the Fama-French 5-factor model augmented with momentum (columns 9-10) or the q -factor model (columns 11-12), we find that portfolios with higher BRT predicted returns have higher average realized risk-adjusted returns. Taking the Carhart 4-factor model as an example, we find that the alpha of decile 1 is negative and significant at -0.42% per month (t -statistic=-2.26), while the alpha of decile 10 is 0.59% per month (t -statistic=3.63). The resulting long-short portfolio has a monthly alpha of 1.01% and is statistically significant with a t -statistic of 6.15.

The results for value-weighted risk-adjusted returns are weaker than the equal-weighted results—in line with the findings in Table 1. Across the various risk-adjustment models, the monthly abnormal performance ranges from a minimum of 0.46% (5.52% annualized) for the CAPM to a maximum of 0.80% (9.60% annualized) for the Fama-French 5-factor model with momentum. In all cases, the alphas of the long-short portfolios are statistically different from zero.

Consistent with the findings reported in Section 3.1, our results suggest that machine learning tools indeed can help predict stock returns. Still, the degree of predictability is significantly lower than what has been reported in the literature once we use as covariates the universe of signals that investors could have constructed in real-time and not the ones that have shown to be successful *ex-post* in predicting the cross-section of stock returns.

3.3 In-sample analysis

Martin and Nagel (2021) develop a model in which investors face a high-dimensional prediction problem and use shrinkage methods to learn from a large number of potential predictors. Martin and Nagel (2021) show that in-sample predictability arises naturally in their setting even in the absence of data mining or market inefficiency. In contrast, there is no out-of-sample predictability because

investors optimally use the information in real-time. Therefore, the central prediction of [Martin and Nagel \(2021\)](#) is that there should be a substantial wedge between in-sample and out-of-sample predictability.

We test this prediction by comparing the in-sample and out-of-sample performance of a universe of fundamental signals. In Sections 3.1 and 3.2, we have already examined the out-of-sample performance of our universe of fundamental signals. We find that the long-short portfolios based on BRT forecasts earn positive and significant returns out of sample. This finding does not necessarily contradict [Martin and Nagel \(2021\)](#), who assume investors are risk-neutral and Bayesian learners with objectively correct beliefs. That is, by construction, there is no room for out-of-sample predictability associated with risk premia or biased beliefs in [Martin and Nagel \(2021\)](#).

In our analysis, we instead focus on [Martin and Nagel's \(2021\)](#) prediction that there should be a significant degradation from in-sample to out-of-sample performance. For ease of comparison, the sample period for our in-sample analysis is the same as that for the out-of-sample test (i.e., 1987-2019). To conduct the in-sample analysis, we fit our BRT model using the full set of 18,113 fundamental signals and use the fitted model to predict each year's returns. There is no consensus on how in-sample tests should be conducted in a machine learning context. For robustness, we perform our test in two ways. In the first, we aim to maintain comparability with the out-of-sample analysis in Section 3.1 and select the optimal hyper-parameters using data from 1963 to 1986. We then use these hyper-parameters to perform an in-sample analysis for 1987-2019. In the second, we align our analysis with [Martin and Nagel \(2021\)](#) and select the optimal hyper-parameters using leave-one-year-out cross-validation over 1987-2019. The procedure uses a particular year t as the validation period and the remaining years as training periods. We choose the combination of hyper-parameters with the highest average validation performance across all years. Finally, we retrain and test the model for the entire 1987-2019 period using the optimally-selected hyper-parameters.

We employ two performance metrics for our analyses. The first is the R^2 of the predictive regression. The second is the long-short portfolio return. Panel A of Table 3 reports the means and standard deviations of the two in-sample R^2 s (corresponding to the two in-sample approaches described above) and the out-of-sample R^2 . Panels B and C maintain the same structure but report the average monthly returns of the equal-weighted (Panel B) and value-weighted (Panel C) long-short portfolios along with their t -statistics.

Consistent with the predictions of [Martin and Nagel \(2021\)](#), we find much greater in-sample

predictability than out-of-sample predictability. Starting with Panel A, the in-sample R^2 in the first and second columns suggest that BRT forecasts explain between 1.21% and 1.65% of the return variation. In comparison, the out-of-sample results indicate a much lower R^2 of 0.28%.

Turning to our second performance metric, we find that the equal- and value-weighted average long-short returns for the first in-sample analysis are 3.21% per month (t -statistics=6.43) and 2.02% per month (t -statistics=4.42), respectively. The results for the second in-sample approach are qualitatively similar and quantitatively higher, with equal- and value-weighted long-short returns of 4.37% and 3.10%, respectively, and highly statistically significant. These in-sample results are several times larger than their out-of-sample counterparts, which are 0.92% for the equal-weighted portfolios and 0.40% for the value-weighted portfolios. We also examine risk-adjusted long-short returns. For brevity, we present the detailed results in Appendix D. Across different models, the equal-weighted long-short alphas for the in-sample analyses range from 2.82% per month to 3.38% per month. The value-weighted risk-adjusted returns range from 1.70% to 2.22%. These numbers are, once again, several times higher than their out-of-sample counterparts reported in Table 2.

Figure 1 plots the times-series of in-sample and out-of-sample return predictability over the period 1987-2019. In Panel A, we compute cross-sectional predictive R^2 s, as we do in Table 3, but instead of averaging across all periods, we compute and plot 12-month moving averages. We depict the results for the two in-sample specifications in orange and red, respectively, and the out-of-sample results in blue. We also plot in dark grey the two-standard-error band around the out-of-sample R^2 s. Panel B plots the BRT long-short portfolio returns¹⁵ for the two in-sample specifications and the out-of-sample long-short returns, adopting the same structure as Panel A. Both panels show that the in-sample predictability is consistently higher than the out-of-sample predictability. The gap between in-sample and out-of-sample predictability is often substantial, e.g., in the early 2000s. Overall, across both performance metrics, we find a significant degradation from in-sample performance to out-of-sample performance, consistent with the predictions of the theoretical model developed in [Martin and Nagel \(2021\)](#).

To summarize, we develop a test of [Martin and Nagel \(2021\)](#). In their model, cash flow growth rates are predictable by a large number of firm characteristics. Fundamental signals are ideally suited for this analysis because of their close relations to cash flows. [Martin and Nagel \(2021\)](#) caution that one should not use the set of published predictors in the academic literature to test their model because

¹⁵We plot equally weighted long-short returns in Figure 1. The chart looks similar if we plot value-weighted long-short returns instead.

some of these predictors may have already been data-mined ex-post. We follow their suggestion and use in our test a universe of fundamental signals, which are free from data-mining concerns. Our empirical results indicate significant degradation from in-sample predictability to out-of-sample predictability, which supports the prediction of [Martin and Nagel \(2021\)](#).

4 Additional Results

In this section, we provide several extensions of our baseline analysis. Section 4.1 examines the performance of BRT long-short portfolios across large and small stocks. Section 4.2 uses an alternative machine learning method, i.e., neural networks. Section 4.3 employs rolling windows instead of recursive windows in estimating the BRT model. Section 4.4 studies whether our results are robust to alternative training and validation periods. Section 4.5 examines different subsets of our universe of fundamental signals. Finally, Section 4.6 investigates whether the performance of BRT portfolios varies with economic and market conditions.

4.1 Focusing on Stocks with Different Market Capitalizations

The strategies we examine in this paper are more likely to be implemented by institutional investors rather than individual investors because they are the ones with the resources and sophistication to use machine learning methods. Previous studies (e.g., [Gompers and Metrick \(2001\)](#)) have shown that institutional investors prefer large-capitalization stocks because they are more liquid and more investable. To evaluate whether the profitability of our machine learning strategies varies across stocks with different capitalizations, each year we divide our sample stocks into four groups: 1) the top 1,000 stocks by market capitalization; 2) the top 2,000 stocks by market capitalization; 3) the bottom 2,000 stocks by market capitalization; 4) the bottom 1,000 stocks by market capitalization, and report the results of our baseline analysis on these four groups of stocks in Table 4.¹⁶

The top panel of Table 4 reports the results for equal-weighted portfolios. We find that the raw and risk-adjusted long-short returns are positive and significant across all size groups. More importantly, the long-short performance increases from the largest to the smallest stocks. Specifically, the long-short return is 0.54% per month for the largest 1,000 stocks, 0.84% for the largest 2,000 stocks, 0.91% for the smallest 2,000 stocks, and 0.98% for the smallest 1,000 stocks. The lower predictive performance

¹⁶[Gu, Kelly, and Xiu \(2020\)](#) also examine the predictability for large stocks (top 1,000 stocks by market capitalization) and small stocks (bottom 1,000 stocks by market capitalization) separately.

for the largest stocks is not surprising. These stocks are likely to incorporate new information more quickly and are hence less likely to be predictable using machine learning algorithms.

The results for value-weighted portfolios are weaker, particularly among the largest stocks. For example, the average long-short return is only 0.23% (t -statistic=0.78) for the largest 1,000 stocks, and 0.31% (t -statistic=1.26) for the largest 2,000 stocks, both of which are statistically insignificant. However, the long-short returns for large stocks do become marginally significant when we control for risks using the Fama-French 5-factor model, the Fama-French 5-factor augmented with momentum, and the q -factor model. In comparison, the average long-short returns among small stocks are economically and statistically significant whether we examine raw or risk-adjusted returns.

Overall, the results in Table 4 indicate that the long-short performance of BRT portfolios is weaker for large stocks than for small stocks. This finding suggests that machine learning methods are better at predicting the returns of smaller stocks, for which news is incorporated more slowly into asset prices. As institutional investors are reluctant to trade smaller capitalization stocks because they are less liquid and less scalable, our results suggest that the benefits of machine learning strategies may be even more limited for institutional investors.

4.2 Neural Networks

In our baseline analysis, we use BRT, which is one of the most powerful machine learning methods for stock return predictions. Nevertheless, one might be concerned that our main results are model-specific and may not extend to other machine learning methods. To ensure this is not the case, we extend our analysis to Neural Networks (NNs) mainly because—together with boosted regression trees—NNs are among the top performers when it comes to return prediction (Gu, Kelly, and Xiu, 2020; Bianchi, Büchner, and Tamoni, 2021). We follow Gu, Kelly, and Xiu (2020) and conduct our analysis using NNs with 1 to 5 hidden layers.

Our results, reported in Table 5, reveal several important findings. First, the equal-weighted long-short returns based on NNs are generally significant, while the value-weighted long-short returns are insignificant. Second, among equal-weighted portfolios, we find that shallow NNs perform better than deep NNs. For example, NNs with 1 and 2 hidden layers achieve long-short returns of 0.74% (t -statistics=3.72) and 0.82% per month (t -statistics=4.53), respectively. NNs with 3 or 4 hidden layers exhibit much lower but still significant long-short returns, while NNs with 5 hidden layers generate insignificant long-short portfolio returns. This finding is consistent with Gu, Kelly, and Xiu

(2020), who show that shallow learning performs better than deep learning. Third, the performance of long-short portfolios based on neural network forecasts is much weaker than those documented by prior machine learning studies, particularly for value-weighted portfolios. Gu, Kelly, and Xiu (2020), for example, show that the long-short portfolios formed based on neural network forecasts earn an average return of 3.27% per month in equal-weighted portfolios and an average return of 2.12% in value-weighted portfolios. Overall, similar to BRT, our results based on neural networks suggest that the real-time performance of machine learning strategies is more modest than that portrayed by prior studies.

4.3 Rolling Windows

We use recursive windows in our baseline specification to align ourselves with the majority of the literature (e.g., Gu, Kelly, and Xiu (2020)). Recursive windows allow for incorporating all available information in generating forecasts, but they can lead to poor forecasts if the data generating process changes over time. An alternative to recursive windows is rolling windows that generate forecasts based on less information and hence are potentially less precise but are more robust to time variations in the relation between fundamental signals and returns. If the relation between the fundamental signals and stock returns is time-varying, rolling windows may improve the predictive power of machine learning algorithms. To assess this possibility, we repeat our main analysis using the rolling window approach described below.

We set the initial estimation period to 24 years so that our out-of-sample test period starts from 1987, the same as in the recursive window approach. To select the optimal hyper-parameters, we split the 24 years into training and validation periods following our baseline specification. In particular, our training period is 12 years and the validation period is 12 years.¹⁷ After obtaining the optimal hyper-parameters, we re-estimate the final model using the 24-year window. Each year we refit the model by moving the 24-year window forward by one year. The estimation period is fixed at 24 years under the rolling window approach. In comparison, under the recursive window approach, the estimation period expands as we roll forward.

Table 6 presents the performance of BRT portfolios for the rolling window approach. We find that the equally weighted portfolios achieve a long-short return of 0.80% per month (t -statistic=4.36) and a Sharpe ratio of 0.71. These numbers are lower than their counterparts for the recursive window

¹⁷We have considered several alternative training and validation periods and find our results to be qualitatively similar.

approach. Specifically, in Table 1 we report that the equal-weighted portfolios exhibit a long-short return of 0.92% (t -statistic=6.33) and a Sharpe ratio of 0.92. The risk-adjusted returns for the rolling window approach are also correspondingly lower than those for the recursive window approach. The results for value-weighted portfolios paint a similar picture. For example, the average long-short return is 0.33% (t -statistic=1.33) under the rolling window approach, compared to the 0.40% (t -statistic=2.36) under the recursive window approach. Overall, we find that the performance of BRT portfolios is somewhat weaker for the rolling window approach than for the recursive window approach.

4.4 Alternative Training and Validation Periods

In our baseline specification, we use an initial training period of 12 years and a validation period of 12 years. In comparison, Gu, Kelly, and Xiu (2020) employ an initial training period of 18 years and a validation period of 12 years. As explained earlier, we choose an initial training period of 12 years because we want to start our out-of-sample test period in 1987, the same as in Gu, Kelly, and Xiu (2020). In this section, we examine whether our results are robust to our choices of the initial training period and validation period. Specifically, we consider nine alternative specifications in which the initial training period varies from 10 to 18 years, while the validation period varies from 10 to 14 years. We examine the performance of BRT portfolios under each of these alternative specifications.

Table 7 presents the results. The top panel reports the results for equal-weighted portfolios, while the bottom panel reports the results for value-weighted portfolios. For convenience, we reproduce the results for our baseline specification in the first row of each panel. Our baseline specification is denoted as “12+12”, meaning 12 years of initial training period and 12 years of validation period. We denote the alternative specifications similarly. For example, “18+12” means 18 years of initial training and 12 years of validation period.

Overall, our results are highly robust across all alternative specifications. For example, the equal-weighted long-short returns range from 0.87% to 1.02% across the alternative specifications, compared to 0.92% for the baseline specification. Similarly, the value-weighted long-short returns range from 0.37% to 0.55% across the alternative specifications, compared to 0.40% for the baseline specification. The level of statistical significance for the long-short returns is also similar between the baseline and alternative specifications. Finally, the results on risk-adjusted returns are also robust to alternative specifications of initial training and validation periods.

4.5 Results Obtained on Subsets of the Fundamental Signals

Our baseline analysis employs a large number of predictor variables. Specifically, we construct a universe of 18,113 fundamental signals based on permutations of 240 accounting variables and 76 financial ratio configurations. One may argue that not all of these signals are actually considered by real-time investors and that the inclusion of these signals weakens the out-of-sample performance of our machine learning strategies. As noted earlier, BRT are known to be robust to the inclusion of irrelevant predictors. Nevertheless, to explore whether our relatively weak out-of-sample performance is driven by the large number of signals in our universe, we repeat our analysis on various subsets of the fundamental signals used in our baseline analysis. We consider two categories of subsets of signals. In the first, we consider subsets of the 240 accounting variables. Specifically, we rank all 240 accounting variables based on the percentage of missing values. We then repeat our analysis using subsets of accounting variables with the lowest percentage of missing values. The basic idea is that accounting variables with fewer missing values are likely to be more important and more value-relevant. In the second category, we consider subsets of the 76 financial ratio categories that are more important and more likely to be considered by real-time investors.

4.5.1 Results using Subsets of the 240 Accounting Variables. In this section, we re-compute our baseline results for subsets of 240 accounting variables ranked based on the percentage of missing values across all stocks for the period 1963-2019. Some accounting variables are missing for all firms before a certain year. For example, all cash flow statement variables are missing before 1988. Including the years for which these variables were missing would artificially inflate their missing value proportions, so when computing the missing rate for an accounting variable, we exclude those years for which the variable is missing for all stocks. Appendix B reports the missing rates for all the 240 accounting variables in our data. Consistent with our expectation, we find that accounting variables with fewer missing values tend to be more important variables. For example, only 0.01% of the “total assets” are missing. For “total sales”, the missing rate is also extremely low at 0.05%. In comparison, 71.1% of the “non-recurring discontinued operations” are missing.

The first row in each panel of Table 8 shows the BRT performance based on the fundamental signals constructed using the 30 accounting variables with the fewest missing values. The second row expands the set to 60 accounting variables with the lowest missing value rates. Each of the remaining rows increases the number of accounting variables by 30 compared to the previous row. The last

row includes all 240 accounting variables, i.e., the entire universe of signals examined in our baseline analysis.

Examining equally weighted portfolio returns reported in the top panel, we find that the long-short performance is the worst when we include only the 30 accounting variables with the lowest missing value rate in our universe. Specifically, the equally weighted long-short return is 0.30% per month, statistically insignificant with a t -statistic of 1.03. As we increase the number of accounting variables in the subset, the portfolio performance generally improves. For example, the long-short portfolio based on 60 accounting variables delivers an average monthly return of 0.72% (t -statistic=2.34), while the one based on 210 accounting variables has an average return of 1.29% (t -statistic=6.55). The best performance is achieved when the number of accounting variables equals 180 or 210. We do observe a decline in performance as we increase the number of regressors from 210 to 240.

The bottom panel of Table 8 reports the results for value-weighted portfolios which, consistent with the baseline results, are weaker than the equal-weighted portfolios. For the same subset of accounting variables, equal-weighted portfolios consistently outperform value-weighted portfolios in both raw returns and risk-adjusted returns. More importantly, similar to the equal-weighted results discussed above, we find that the performance of value-weighted portfolios generally increases with the size of the subset and peaks when the number of accounting variables is in the 180-210 range.

Overall, our results indicate that increasing the number of accounting variables included in the analysis enlarges the conditioning information set that can be exploited by real-time investors and generally increases the profitability of the machine learning strategies. This finding is inconsistent with the argument that the relatively weak performance of our machine learning strategies is due to the large size of our universe.

4.5.2 Results using Subsets of Financial Ratio Configurations. In this section, we repeat our baseline analysis on several subsets of the 76 financial ratio configurations. In constructing the universe of fundamental signals, we follow [Yan and Zheng \(2017\)](#) and use 15 base variables (Y) in addition to the 240 accounting variables (X). We refer the readers to [Yan and Zheng \(2017\)](#) for more details. We consider two subsets of Y's based on the importance of such base variables. The first subset includes the three most commonly used base variables—i.e., total assets, total sales, and market cap—which we term “Y3.” The second subset (termed “Y5”) includes two additional important base variables, i.e., total liability and shareholder’s equity.

We also split the 76 financial ratio configurations into five categories based on their functional forms. The first category (P1) includes the ratios of accounting variables to base variables (i.e., the ratios #1 to #15 in Appendix C). The second category (P2) includes the change of ratios in the first category (i.e., the ratios #16 to #30). The third category (P3) contains the percentage change of ratios in the first category, or the ratios #31 to #45). The fourth category (P4) contains changes in accounting variables scaled by lagged base variables (i.e., the ratios #46 to #60). The fifth category (P5) includes the difference between the percentage changes in both accounting variables and base variables (i.e., the ratios #61 to #75).¹⁸

The top panel of Table 9 shows the BRT equal-weighted portfolio performance on the two subsets of the base variables and the five subsets of the financial ratio configurations. In contrast, the bottom panel shows the results for value-weighted portfolios. The equal-weighted long-short portfolio returns are positive and statistically significant regardless of which subset we examine. For example, when we use the three most important base variables (Y3), BRT achieve a significant long-short return of 0.82% per month, with an associated t -statistics of 4.10. Adding two additional base variables (Y5) leads to a long-short portfolio return of 0.72%, also statistically significant. Similar results hold for subsets from P1 to P5, where the average long-short returns range from 0.51% to 0.87%. It is important to note that the long-short returns for all seven subsets of financial ratios are lower than that for the full universe of fundamental signals (0.92%).

The bottom panel of Table 9 repeats the exercise for value-weighted returns. As in our baseline analysis, value-weighted returns are weaker than equal-weighted returns and often lack statistical significance. We also find that the value-weighted long-short returns for subsets of financial ratios are generally lower than that for the full universe of fundamental signals. The only exception is the subset Y5. Overall, Table 9 shows little systematic evidence that the performance of our machine learning strategies would be much better had we considered a significantly smaller universe of fundamental signals.

4.6 Testing for Time-varying Predictability

In Table 10, we examine whether the profitability of BRT strategies varies with economic and market conditions. Specifically, we split our sample period based on investor sentiment,¹⁹ the VIX index also

¹⁸We note that for each category, we also include the percentage change of the accounting variable itself, i.e., the ratio #76 in Appendix C.

¹⁹We obtain the investor sentiment's data from Wurgler's website at <http://people.stern.nyu.edu/jwurgler/>.

known as the “fear-gauge”, market liquidity (Pástor and Stambaugh, 2003), business cycle indicators as published by NBER, and market state—proxied by the cumulative market returns over the previous 24 months. We also divide our sample period into two halves (1987-2003 and 2003-2019) to examine whether the predictability declines over time.

Panel A shows the long-short portfolio returns for high- and low-sentiment periods. When examining equal-weighted returns, we find significant predictability during both high- and low-sentiment periods. In contrast, value-weighted long-short returns are only marginally significant during low-sentiment periods and insignificant during high-sentiment periods. Whether we look at equal- or value-weighted returns, the difference in long-short returns between high- and low-sentiment periods is statistically insignificant. We find similar results in Panel B, where we divide the sample period into high- and low-VIX periods, and in Panel C, where we divide periods into high- and low-liquidity periods. In each panel, we find significant equal-weighted returns across both subperiods. The value-weighted returns, however, are either insignificant or marginally significant. As in Panel A, we find little significant evidence of differential predictability across subperiods.

In Panel D, we find that the BRT strategies are more profitable during recessions than during expansions. The difference in long-short returns is economically significant, 0.54% for equal-weighted portfolios and 0.66% for value-weighted portfolios, but statistically insignificant. In Panel E, we split the sample period into UP and DOWN market states based on previous 24-month cumulative market returns. We find that the long-short return is higher during UP state than during DOWN state. Specifically, the equal-weighted long-short return is 1.23% during UP state and 0.63% during DOWN state. Similarly, the value-weighted long-short return is 0.78% during UP state and 0.03% during DOWN state. The differences in long-short returns between the UP and DOWN states are economically large and statistically marginally significant.

In Panel F, we find a similar level of predictability during the first and second half of our sample period. Specifically, the equal-weighted long-short return is 1.05% during the first half of our sample period and 0.80% during the second half. Similarly, the value-weighted long-short return is 0.32% during the first half and 0.48% during the second half. The difference in long-short returns between the two halves is economically small and statistically insignificant.

Overall, the results in Table 10 indicate that the return predictability implied by our real-time machine learning strategies does not change significantly with investor sentiment, market volatility, or market liquidity. However, there is some evidence that the profitability of our BRT strategies varies

systematically with the business cycle and market state. Finally, we find no evidence that the return predictability differs significantly across the two halves of our sample period.

5 Technical Signals

Our analyses so far have focused on fundamental signals. A key reason for this focus is that we are able to construct a “universe” of fundamental signals. Real-time investors do not know what signals turn out to be significant ex-post, so they have to learn from a universe of potential strategies. Another class of signals ideally suited for our analysis is past return-based signals. In this section, we follow [Martin and Nagel \(2021\)](#) and construct a universe of past return-based signals and then repeat our main analyses. Specifically, we include in our universe the monthly returns and monthly squared returns during the past 120 months, excluding the most recent month.²⁰ Therefore, we have $119 \times 2 = 238$ past return-based signals for this analysis. For ease of exposition, we also refer to these signals as technical signals.²¹

Our stock sample for this analysis consists of the NYSE, AMEX, and NASDAQ common stocks (with a CRSP share code of 10 or 11) with valid past return data. We exclude those stocks with a share price lower than \$1 at the end of month $t - 1$. Our final sample includes 25,631 stocks. Our sample period is from July 1926 to December 2019.

As in our analysis of fundamental signals, we use BRT as the primary machine learning method. We continue to employ a recursive window approach to train and test our model. Our initial training period spans from July 1926 to June 1938 and our initial validation period is from July 1938 to June 1950. Our out-of-sample test starts from July 1950, and rolls forward one month at a time. As we roll forward, our training period expands, while our validation period stays as 12 years (or 144 months). We predict next month’s stock return in our analysis.

As in our analysis of fundamental signals, we form long-short portfolios of stocks based on the BRT predicted returns. Specifically, we long stocks with the highest BRT predicted returns and short stocks with the lowest BRT predicted returns. We track the performance of these portfolios for one month and compute the return spread between the long and short portfolios. For performance evaluation,

²⁰ [Martin and Nagel \(2021\)](#) exclude the most recent month to avoid microstructure effects.

²¹ We do not combine the “universe” of fundamental signals and the “universe” of technical signals in our analysis for three reasons. First, the number of technical signals in our universe is much smaller than the number of fundamental signals. Second, the sample period for our technical signals, which extends back to 1926, is much longer than that for the fundamental signals. Third, we predict one-month-ahead stock returns for technical signals. Due to computational constraints, we predict annual stock returns for fundamental signals.

we report alphas for the long-short portfolio using the CAPM, the Fama-French three-factor model, and the Carhart four-factor model. We do not report Fama-French five-factor alphas or HXZ Q-factor alphas because the factors data are available only after July 1963 and January 1967, respectively.

Table 11 shows the main results for the technical signals. As in Table 1, we report the BRT predicted returns (i.e. the sorting variable), the average realized returns, the standard deviation of realized returns, and the Sharpe ratios of BRT-sorted portfolios. In addition, we also present the CAPM, Fama-French 3-factor, and Carhart 4-factor alphas. We report results for both equal-weighted and value-weighted portfolios.

The top panel of Table 11 presents the results for equally weighted portfolios. We find that the long-short portfolio earns an average return of 0.94% per month (t -statistic=6.33). The annual Sharpe ratio of the long-short portfolio is 0.97. The performance of the value-weighted portfolios is more modest. The average long-short return is 0.56% per month, with a t -statistic of 3.34. The Sharpe ratio of the value-weighted long-short portfolio is 0.45, lower than that of the market portfolio. The CAPM, Fama-French 3-factor, and Carhart 4-factor alphas of the long-short portfolio continue to be positive and statistically significant. It is worth noting that the Carhart 4-factor alpha is substantially lower than the average long-short return as well as the CAPM and Fama-French 3-factor alphas. Specifically, the 4-factor alpha for the long-short portfolio is 0.33% per month, with a t -statistic of 2.20. The smaller Carhart alpha is not surprising because much of the predictive ability of past returns is related to the momentum effect of Jegadeesh and Titman (1993).

Table 12 compares the in-sample and out-of-sample performance for technical signals. We show that in contrast to the modest out-of-sample predictability, technical signals exhibit much greater in-sample predictability. For example, the in-sample R^2 ranges from 1% to 1.45%, compared to the out-of-sample R^2 of 0.50%. The in-sample long-short returns, i.e., 1.92-3.14% for equal-weighted portfolios and 1.32-1.90% for value-weighted portfolios, are also much higher than their out-of-sample counterparts. These results are similar to what we find for fundamental signals and are once again consistent with the predictions of Martin and Nagel (2021) that there should be a substantial wedge between in-sample and out-of-sample predictability.

In Table 13 we repeat the main analysis by using neural network forecasts. We find that the performance of real-time machine learning strategies based on neural network forecasts is substantially weaker than that of BRT. Similar to our results for fundamental signals, we find that shallow learning, i.e., networks with 1 or 2 hidden layers perform relatively better than deep learning, i.e., those with

3, 4, or 5 hidden layers. Specifically, the long-short returns for shallow learning strategies are positive and statistically significant, although they are much lower than those based on BRT. In comparison, the long-short returns for deep learning strategies are completely insignificant, regardless of whether we examine equal- or value-weighted portfolios.

Overall, our results based on technical signals are broadly consistent with those based on fundamental signals. Specifically, we find significant long-short returns for our machine learning strategies, suggesting that real-time investors benefit from using machine learning forecasts. However, the performance of these real-time machine learning strategies is considerably weaker than those reported in the prior literature, particularly when we examine value-weighted long-short portfolios or use neural networks to generate return forecasts.

6 Conclusions

Recent studies document strong performance for machine learning-based investment strategies. Our analyses paint a more conservative picture of the practical value of machine learning strategies for real-time investors. The machine learning strategies examined by prior studies use subsequently discovered anomaly variables as predictors of stock returns and cannot be implemented in real-time. We construct machine learning strategies based on a “universe” of fundamental signals. The out-of-sample performance of our strategies is positive and significant, but considerably weaker than those documented by previous studies, particularly in value-weighted portfolios. We find similar results examining a universe of past return-based signals. We also find significant degradation from in-sample performance to out-of-sample performance, supporting the predictions of [Martin and Nagel \(2021\)](#). Overall, our results indicate that machine learning strategies are profitable, but the economic gains to real-time investors from using machine learning forecasts are more modest than previously thought.

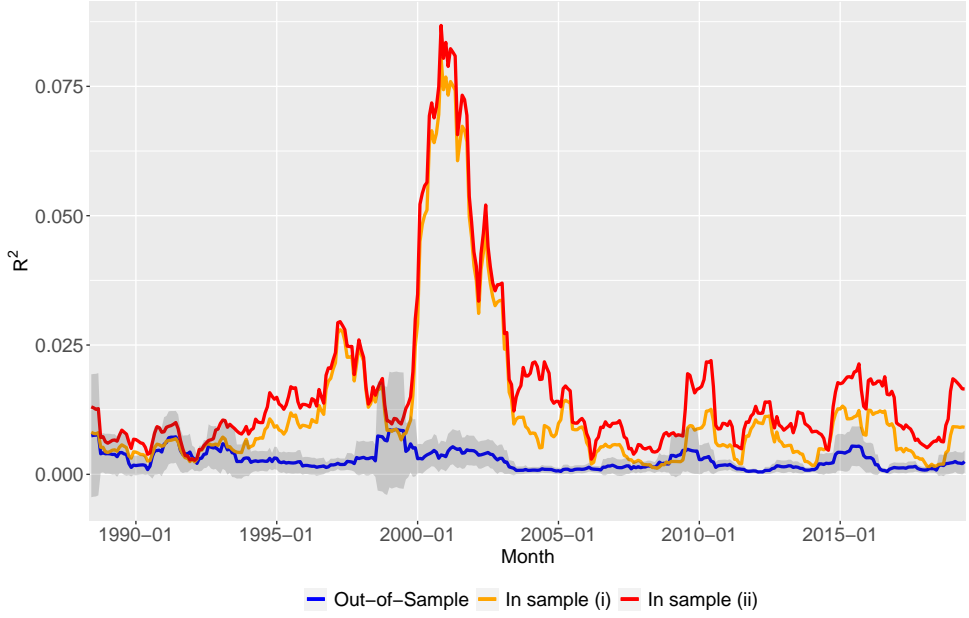
References

- Arnott, R., C. R. Harvey, and H. Markowitz. 2019. A Backtesting Protocol in the Era of Machine Learning. *Journal of Financial Data Science* 1:64–74.
- Avramov, D., S. Cheng, and L. Metzker. 2020. Machine Learning versus Economic Restrictions: Evidence from Stock Return Predictability. SSRN Scholarly Paper ID 3450322.
- Avramov, D., G. Kaplanski, and A. Subrahmanyam. 2021. Postfundamentals Price Drift in Capital Markets: A Regression Regularization Perspective. *Management Science* forthcoming.
- Bianchi, D., M. Büchner, and A. Tamoni. 2021. Bond Risk Premiums with Machine Learning. *Review of Financial Studies* 34:1046–89.
- Breiman, L., J. Friedman, C. J. Stone, and R. Olshen. 1984. *Classification and Regression Trees*. Chapman and Hall/CRC.
- Bryzgalova, S., M. Pelger, and J. Zhu. 2020. Forest through the Trees: Building Cross-Sections of Stock Returns. Working Paper.
- Carhart, M. M. 1997. On Persistence in Mutual Fund Performance. *Journal of Finance* 52:57–82.
- Chen, L., M. Pelger, and J. Zhu. 2020. Deep Learning in Asset Pricing. SSRN Scholarly Paper ID 3350138, Rochester, NY.
- Chinco, A., A. D. Clark-Joseph, and M. Ye. 2019. Sparse Signals in the Cross-Section of Returns. *Journal of Finance* 74:449–92.
- Chordia, T., A. Goyal, and A. Saretto. 2020. Anomalies and False Rejections. *Review of Financial Studies* 33:2134–79.
- Cochrane, J. H. 2011. Presidential Address: Discount Rates. *Journal of Finance* 66:1047–108.
- Cong, L. W., K. Tang, J. Wang, and Y. Zhang. 2021. AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI. SSRN Scholarly Paper ID 3554486.
- Cooper, M. J., H. Gulen, and M. J. Schill. 2008. Asset Growth and the Cross-Section of Stock Returns. *Journal of Finance* 63:1609–51.

- Fama, E. F., and K. R. French. 1992. The Cross-Section of Expected Stock Returns. *Journal of Finance* 47:427–65.
- . 1993. Common Risk Factors In The Returns On Stocks And Bonds. *Journal of Financial Economics* 33:3–56.
- . 1996. Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance* 51:55–84.
- . 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116:1–22.
- Feng, G., N. Polson, and J. Xu. 2020. Deep Learning in Characteristics-Sorted Factor Models. SSRN Scholarly Paper ID 3243683.
- Freyberger, J., A. Neuhierl, and M. Weber. 2020. Dissecting Characteristics Nonparametrically. *Review of Financial Studies* 33:2326–77.
- Friedman, J. H. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* 29:1189–232.
- Gompers, P. A., and A. Metrick. 2001. Institutional Investors and Equity Prices. *Quarterly Journal of Economics* 116:229–59.
- Graham, B., and D. L. F. Dodd. 1934. *Security analysis*. McGraw-Hill, New York.
- Gu, S., B. T. Kelly, and D. Xiu. 2020. Empirical Asset Pricing via Machine Learning. *Review of Financial Studies* 33:2223–73.
- Harvey, C. R., Y. Liu, and H. Zhu. 2016. ... and the Cross-Section of Expected Returns. *Review of Financial Studies* 29:5–68.
- Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- Hou, K., C. Xue, and L. Zhang. 2015. Digesting Anomalies: An Investment Approach. *Review of Financial Studies* 28:650–705.
- . 2020. Replicating Anomalies. *Review of Financial Studies* 33:2019–133.
- Israel, R., B. T. Kelly, and T. J. Moskowitz. 2020. Can Machines “Learn” Finance? *Journal of Investment Management* 18:23–36.

- Jegadeesh, N., and S. Titman. 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance* 48:65–91.
- Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems* 30, 3146–54.
- Kozak, S., S. Nagel, and S. Santosh. 2020. Shrinking the Cross-Section. *Journal of Financial Economics* 135:271–92.
- Leippold, M., Q. Wang, and W. Zhou. 2021. Machine Learning in the Chinese Stock Market. *Journal of Financial Economics* forthcoming.
- Martin, I., and S. Nagel. 2021. Market Efficiency in the Age of Big Data. *Journal of Financial Economics* forthcoming.
- Newey, W. K., and K. D. West. 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:703–8.
- Novy-Marx, R., and M. Velikov. 2016. A Taxonomy of Anomalies and Their Trading Costs. *Review of Financial Studies* 29:104–47.
- Pástor, L., and R. F. Stambaugh. 2003. Liquidity Risk and Expected Stock Returns. *Journal of Political Economy* 111:642–85.
- Rapach, D. E., J. K. Strauss, and G. Zhou. 2013. International Stock Return Predictability: What Is the Role of the United States? *Journal of Finance* 68:1633–62.
- Yan, X. S., and L. Zheng. 2017. Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach. *Review of Financial Studies* 30:1382–423.

(a) In-Sample and Out-of-Sample R^2



(b) In-Sample and Out-of-Sample Long-Short Return

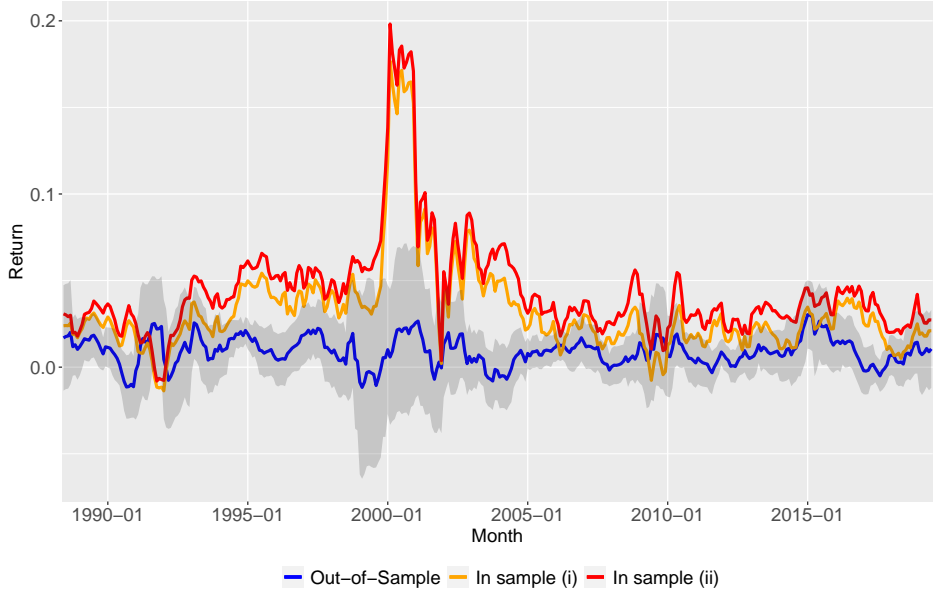


Figure 1: In- and Out-of-Sample Return Predictability

This figure shows the in- and out-of-sample BRT return predictability. Panel (a) plots the 12-month moving averages of in- and out-of-sample R^2 . In-sample (i) denotes the in-sample results with optimal hyper-parameters chosen over training period 1962 - 1986. For In-sample (ii) the optimal hyper-parameters are chosen using leave-one-year-out cross validation over the test period 1987-2019, following [Martin and Nagel \(2021\)](#). Panel (b) shows the in- and out-of-sample returns of a long-short portfolio strategy based on BRT forecasts.

Table 1: Performance of Portfolios Sorted by BRT Predicted Returns

Rank	Equal Weight					Value Weight				
	Pred	Avg	t-stat	SD	SR	Pred	Avg	t-stat	SD	SR
1 (Low)	-0.05	0.09	0.22	7.93	0.04	0.00	0.39	1.25	6.26	0.21
2	0.30	0.50	1.57	6.35	0.27	0.30	0.62	2.56	5.48	0.39
3	0.50	0.71	2.35	6.18	0.40	0.51	0.59	2.63	4.78	0.42
4	0.64	0.83	2.85	5.87	0.49	0.64	0.74	3.19	4.62	0.55
5	0.73	0.81	2.85	5.61	0.50	0.73	0.60	2.30	4.62	0.45
6	0.80	0.92	3.15	5.66	0.56	0.80	0.64	2.91	4.52	0.49
7	0.88	0.98	3.26	5.72	0.59	0.88	0.67	2.64	4.92	0.47
8	0.97	1.01	3.25	5.58	0.63	0.97	0.50	1.85	4.83	0.36
9	1.11	0.95	2.80	6.03	0.55	1.11	0.62	2.13	5.15	0.42
10 (High)	1.69	1.01	2.76	6.89	0.51	1.61	0.79	2.47	6.00	0.45
10-1	1.74	0.92	6.33	3.51	0.92	1.61	0.40	2.36	4.75	0.29

This table reports the excess returns of decile portfolios sorted by BRT predicted returns from 1987 to 2019. We predict stock annual excess returns using 18,113 fundamental signals (as described in Section 2.1). We use a recursive window approach and select the optimal hyper-parameters using a cross-validation approach. Our initial estimation period is 1963-1986. The first 12 years is the training period and the second 12 years is the validation period. As we roll forward, the training period expands while the validation period stays at 12 years. The left panel reports equal-weighted portfolio results. In the first column of this panel we report the average predicted monthly returns from the BRT model (*Pred*). The second and third columns report the average realized monthly excess returns (*Avg*) and associated *t*-statistics (*t-stat*), computed using [Newey and West \(1987\)](#) standard errors with 12 lags. Finally, in the fourth and fifth column we report the portfolios' return standard deviations (*SD*) and Sharpe ratios (*SR*), respectively. The right panel reports the same results using value-weighted portfolio returns. All returns are expressed in percent per month.

Table 2: Risk-adjusted Performance of Portfolios Sorted by BRT Predicted Returns

Rank	Equal Weight											
	CAPM		FF3		Carhart		FF5		FF5+Mom		Q	
	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat
1 (Low)	-0.77	-3.16	-0.70	-4.21	-0.42	-2.26	-0.27	-1.34	-0.10	-0.47	0.01	0.03
2	-0.24	-1.18	-0.24	-2.06	-0.05	-0.41	-0.10	-0.68	0.03	0.20	0.11	0.70
3	-0.03	-0.15	-0.05	-0.50	0.10	0.90	0.01	0.09	0.12	0.92	0.20	1.34
4	0.11	0.59	0.08	0.92	0.23	2.13	0.11	1.05	0.21	2.00	0.27	2.15
5	0.12	0.68	0.08	1.06	0.20	2.41	0.06	0.68	0.14	1.73	0.20	1.70
6	0.23	1.16	0.16	1.78	0.31	3.38	0.12	1.16	0.23	2.56	0.26	1.93
7	0.28	1.42	0.24	2.68	0.37	3.72	0.27	2.39	0.36	3.37	0.39	2.88
8	0.32	1.68	0.29	2.85	0.43	3.91	0.32	2.76	0.42	3.67	0.46	3.38
9	0.22	1.19	0.20	1.62	0.42	3.71	0.38	2.35	0.52	3.90	0.55	3.39
10 (High)	0.23	1.03	0.29	1.76	0.59	3.63	0.65	3.06	0.84	4.35	0.86	4.07
10-1	1.00	6.44	0.99	6.23	1.01	6.15	0.92	4.83	0.94	4.91	0.85	4.45

Rank	Value Weight											
	CAPM		FF3		Carhart		FF5		FF5+Mom		Q	
	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat
1 (Low)	-0.38	-2.48	-0.28	-2.38	-0.34	-2.89	-0.14	-1.14	-0.19	-1.66	-0.11	-0.68
2	-0.06	-0.45	0.02	0.16	0.00	0.04	0.06	0.39	0.04	0.30	0.12	0.82
3	-0.04	-0.49	-0.02	-0.18	-0.02	-0.26	-0.07	-0.72	-0.07	-0.75	0.00	-0.01
4	0.12	1.40	0.15	1.78	0.09	1.14	0.06	0.67	0.03	0.32	0.08	0.82
5	-0.02	-0.29	-0.02	-0.25	0.00	-0.01	-0.11	-1.39	-0.09	-1.20	-0.07	-0.88
6	0.05	0.54	0.01	0.15	0.03	0.35	-0.20	-2.37	-0.17	-1.88	-0.16	-2.12
7	0.03	0.32	0.00	-0.02	0.02	0.24	-0.14	-1.32	-0.11	-1.06	-0.11	-1.02
8	-0.12	-1.11	-0.12	-1.03	-0.05	-0.49	-0.16	-1.32	-0.11	-0.98	-0.08	-0.63
9	-0.03	-0.21	0.04	0.27	0.08	0.76	0.13	1.21	0.16	1.51	0.23	1.77
10 (High)	0.08	0.37	0.18	1.13	0.32	2.25	0.54	3.67	0.61	4.11	0.58	3.99
10-1	0.46	2.16	0.46	2.61	0.66	3.72	0.68	3.11	0.80	3.79	0.69	2.96

This table shows the risk-adjusted performance of the BRT portfolios based on the CAPM model, the Fama-French 3-factor model, the Carhart 4-factor model, the Fama-French 5-factor model, the Fama-French 5-factor model augmented with momentum factor, and the q -factor model. The top panel reports results for equal-weighted portfolios. The bottom reports results for value-weighted portfolios. All returns are expressed in percent per month.

Table 3: In- and Out-of-sample Return Predictability

Panel A: In- and Out-of-Sample R^2			
	In-Sample (i)	In-Sample (ii)	Out-of-Sample
Mean	1.21	1.65	0.28
S.D.	2.42	2.71	0.60
Panel B: Equal-weighted Long-short Returns			
	In-Sample (i)	In-Sample (ii)	Out-of-Sample
Mean	3.21	4.37	0.92
t -stat	6.43	8.07	6.33
Panel C: Value-weighted Long-short Returns			
	In-Sample (i)	In-Sample (ii)	Out-of-Sample
Mean	2.02	3.10	0.40
t -stat	4.42	6.26	2.36

This table reports the in- and out-of-sample performance of BRT forecasts. The first and second columns of Panel A report the means and standard deviations of the in-sample R^2 . In the first column, the optimal hyper-parameters are determined using 1962-1986 as the training period. In the second column, the optimal hyper-parameters are determined using a leave-one-year-out cross-validation procedure over the period 1987-2019, following [Martin and Nagel \(2021\)](#). The third column reports the out-of-sample R^2 . Panels B and C maintain the same structure but report the average monthly returns and their t -statistics associated with equal-weighted (Panel B) and value-weighted (Panel C) long-short portfolios. All returns and R^2 are expressed in percent.

Table 4: Performance of Portfolios Sorted by BRT Predicted Returns – Subsets of Stocks Based on Market Capitalization

Sub-Samples	Equal Weight														
	Returns		SR	CAPM		FF3		Carhart		FF5		FF5+MOM		Q	
	Avg	t-stat		alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat
Top 1000	0.54	2.04	0.42	0.59	2.35	0.67	2.36	0.61	2.64	0.94	2.76	0.88	2.90	0.94	2.64
Top 2000	0.84	3.66	0.79	0.92	4.01	0.94	3.78	0.89	4.09	1.06	3.65	1.02	3.77	0.96	3.25
Bottom 2000	0.91	3.48	0.67	0.98	4.04	0.98	3.77	1.04	4.02	0.93	3.17	0.97	3.42	0.91	3.23
Bottom 1000	0.98	5.65	0.98	1.06	6.13	1.06	5.82	1.06	5.73	0.98	4.85	0.98	4.81	0.91	4.63

Sub-Samples	Value Weight														
	Returns		SR	CAPM		FF3		Carhart		FF5		FF5+MOM		Q	
	Avg	t-stat		alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat
Top 1000	0.23	0.78	0.16	0.22	0.66	0.34	1.15	0.37	1.42	0.74	2.37	0.73	2.50	0.81	2.43
Top 2000	0.31	1.26	0.24	0.32	1.15	0.39	1.53	0.50	2.12	0.70	2.47	0.75	2.71	0.73	2.63
Bottom 2000	0.72	2.59	0.50	0.75	2.88	0.77	2.88	0.86	2.95	0.76	2.35	0.82	2.47	0.73	2.23
Bottom 1000	0.95	5.15	0.88	0.98	5.32	1.00	5.13	1.03	5.08	1.06	4.42	1.07	4.39	0.99	4.33

This table shows the performance of BRT portfolios on stocks with different market capitalization. We report results for the 1,000 and 2,000 stocks with the largest market capitalization as well as for the 1,000 and 2,000 stocks with the smallest market capitalization. All results are computed following the procedure described in Table 1. The first three columns report average monthly returns for the long-short portfolios as well as the associated Sharpe Ratios (SR). The remaining columns report risk-adjusted returns—see Table 2 for details. The top panel reports results for equal-weighted returns. The bottom panel reports results for value-weighted returns. All returns are expressed in percent per month.

Table 5: Performance of Portfolios Sorted by NN Predicted Returns

Method	Returns		SR	Equal Weight						Q					
	Avg	t-stat		CAPM		FF3		Carhart			FF5		FF5+MOM		
				alpha	t-stat	alpha	t-stat	alpha	t-stat		alpha	t-stat	alpha	t-stat	alpha
NN1	0.74	3.72	0.70	0.77	3.44	0.73	3.89	0.89	5.34	0.87	3.99	0.97	4.84	0.85	3.74
NN2	0.82	4.53	0.80	0.99	5.33	0.90	5.54	0.75	4.47	0.54	3.57	0.46	2.86	0.45	2.62
NN3	0.43	2.42	0.50	0.49	2.62	0.54	3.06	0.53	3.07	0.57	3.51	0.56	3.11	0.57	3.04
NN4	0.47	2.32	0.51	0.42	1.74	0.46	2.20	0.56	2.92	0.71	3.67	0.76	3.97	0.74	4.21
NN5	0.11	0.54	0.11	0.05	0.22	0.10	0.47	0.21	1.11	0.19	1.06	0.27	1.47	0.35	1.88

Method	Returns		SR	Value Weight						Q					
	Avg	t-stat		CAPM		FF3		Carhart			FF5		FF5+MOM		
				alpha	t-stat	alpha	t-stat	alpha	t-stat		alpha	t-stat	alpha	t-stat	alpha
NN1	-0.04	-0.17	-0.03	-0.04	-0.16	-0.05	-0.25	0.19	0.94	0.12	0.42	0.28	1.23	0.19	0.71
NN2	0.06	0.27	0.05	0.24	0.87	0.11	0.52	0.02	0.11	-0.22	-1.13	-0.25	-1.30	-0.24	-1.11
NN3	0.04	0.17	0.03	0.11	0.48	0.10	0.41	0.19	0.86	-0.07	-0.28	0.02	0.07	0.01	0.05
NN4	0.21	0.63	0.16	0.13	0.33	0.24	0.73	0.21	0.72	0.56	2.12	0.52	2.02	0.54	1.81
NN5	0.20	0.76	0.16	0.11	0.36	0.15	0.56	0.16	0.64	0.14	0.67	0.16	0.75	0.25	1.06

This table shows the performance of portfolios sorted by Neural Network (NN) predicted returns. We consider NN models with hidden layers that range from 1 through 5. The first three columns report average monthly returns for the long-short portfolios as well as the associated Sharpe Ratios. The remaining columns report risk-adjusted returns—see Table 2 for details. The top panel reports results for equal-weighted returns. The bottom panel reports results for value-weighted returns. All returns are expressed in percent per month.

Table 6: Performance of Portfolios Sorted by BRT Predicted Returns – Rolling-window Approach

Rank	Equal Weight															
	High-Low		SR		CAPM		FF3		Carhart		FF5		FF5+MOM		Q	
	mean	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat
1 (Low)	0.23	0.56	0.10	0.10	-0.61	-2.30	-0.57	-2.92	-0.35	-1.84	-0.22	-1.41	-0.08	-0.49	0.02	0.11
2	0.64	1.87	0.33	0.33	-0.14	-0.61	-0.15	-1.09	0.08	0.51	0.03	0.18	0.18	1.26	0.26	1.49
3	0.74	2.26	0.41	0.41	0.00	-0.02	-0.02	-0.19	0.15	1.10	0.09	0.65	0.20	1.51	0.27	1.77
4	0.82	2.68	0.47	0.47	0.09	0.44	0.05	0.50	0.18	1.88	0.10	0.92	0.19	1.97	0.27	2.52
5	0.80	2.81	0.49	0.49	0.12	0.59	0.08	0.81	0.21	2.12	0.05	0.47	0.15	1.48	0.21	1.61
6	0.85	3.09	0.54	0.54	0.17	0.96	0.11	1.45	0.26	2.92	0.10	1.03	0.20	2.21	0.22	1.57
7	0.91	3.10	0.57	0.57	0.23	1.22	0.18	1.89	0.32	3.19	0.19	1.68	0.28	2.65	0.32	2.40
8	0.91	2.98	0.55	0.55	0.22	1.22	0.18	1.59	0.39	3.49	0.23	1.48	0.37	2.56	0.40	1.96
9	0.90	2.79	0.49	0.49	0.15	0.82	0.16	1.22	0.42	2.84	0.36	1.74	0.53	2.63	0.55	2.13
10 (High)	1.03	2.71	0.53	0.53	0.26	1.07	0.32	1.80	0.54	3.09	0.62	2.96	0.76	3.65	0.80	3.70
10-1	0.80	4.36	0.71	0.71	0.87	4.11	0.88	4.24	0.88	4.32	0.84	4.13	0.84	3.89	0.78	3.36

Rank	Value Weight															
	High-Low		SR		CAPM		FF3		Carhart		FF5		FF5+MOM		Q	
	mean	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat
1 (Low)	0.40	1.12	0.20	0.20	-0.39	-1.68	-0.27	-1.12	-0.30	-1.57	-0.03	-0.15	-0.06	-0.38	0.04	0.18
2	0.57	1.79	0.33	0.33	-0.14	-0.66	-0.04	-0.21	-0.02	-0.09	0.09	0.55	0.10	0.65	0.18	0.95
3	0.62	2.13	0.38	0.38	-0.06	-0.35	0.01	0.03	-0.05	-0.35	0.06	0.49	0.02	0.14	0.09	0.61
4	0.60	2.58	0.44	0.44	-0.02	-0.15	-0.01	-0.15	-0.02	-0.20	-0.10	-1.13	-0.09	-1.09	-0.07	-0.78
5	0.59	2.62	0.45	0.45	-0.01	-0.09	0.02	0.23	0.05	0.68	0.00	-0.05	0.02	0.26	0.04	0.41
6	0.77	3.31	0.56	0.56	0.13	1.29	0.14	1.41	0.14	1.54	0.00	0.04	0.02	0.18	0.07	0.75
7	0.51	1.92	0.37	0.37	-0.13	-1.63	-0.13	-1.58	-0.09	-1.16	-0.22	-2.60	-0.19	-2.38	-0.15	-1.52
8	0.67	2.91	0.51	0.51	0.08	0.91	0.09	0.91	0.14	1.62	-0.02	-0.24	0.02	0.27	0.09	0.88
9	0.53	1.73	0.37	0.37	-0.11	-0.65	-0.07	-0.53	-0.04	-0.37	-0.08	-0.69	-0.06	-0.55	-0.03	-0.28
10 (High)	0.73	2.13	0.43	0.43	0.02	0.09	0.10	0.71	0.20	1.43	0.30	2.32	0.36	2.71	0.38	2.88
10-1	0.33	1.33	0.21	0.21	0.40	1.33	0.37	1.27	0.50	2.03	0.33	1.42	0.42	1.85	0.34	1.31

This table shows the performance of BRT portfolios using rolling windows instead of recursive windows. All results are computed following the procedure described in Table 1 except that our estimation period is fixed at 24 years. In particular, we select the optimal hyper-parameters using the most recent 12 years as our validation period and the remaining 12 years as our training period. After obtaining the optimal hyper-parameters, we retrain our model using the rolling window of 24 years. The first three columns report average monthly returns for the decile portfolios and the associated Sharpe Ratios. The remaining columns report risk-adjusted returns—see Table 2 for details. The top panel reports results for equal-weighted returns. The bottom panel reports results for value-weighted returns. All returns are expressed in percent per month.

Table 7: Performance of Portfolios Sorted by BRT Predicted Returns – Alternative Training and Validation Periods

Training +Validation	Returns		SR	Equal Weight				Q							
	Avg	<i>t-stat</i>		CAPM	FF3	Carhart	FF5	FF5+MOM	alpha	<i>t-stat</i>	alpha	<i>t-stat</i>			
12+12	0.92	6.33	0.92	1.00	6.44	0.99	6.23	1.00	6.15	0.93	4.83	0.94	4.91	0.85	4.45
10+10	1.02	6.73	1.06	1.09	6.87	1.09	6.40	1.04	6.53	1.12	5.62	1.09	5.64	1.01	5.19
12+10	1.01	6.46	1.05	1.08	6.65	1.09	6.33	1.06	6.49	1.11	5.55	1.09	5.58	1.01	5.11
14+10	1.01	6.11	1.04	1.07	6.25	1.08	5.99	1.06	6.04	1.09	5.22	1.07	5.21	1.01	4.86
14+12	0.87	5.79	0.86	0.92	5.96	0.91	5.82	0.91	5.74	0.80	4.56	0.81	4.65	0.72	4.11
16+12	0.88	5.81	0.88	0.95	6.32	0.97	6.55	0.95	6.39	0.87	5.12	0.86	5.08	0.79	4.72
18+12	0.89	5.69	0.90	0.96	6.28	0.97	6.48	0.96	6.42	0.86	5.22	0.87	5.23	0.81	4.96
14+14	0.99	5.68	1.03	1.09	6.68	1.13	6.81	1.05	7.07	1.10	5.55	1.04	5.62	1.01	5.20
16+14	1.00	5.47	1.05	1.10	6.46	1.12	6.60	1.06	6.83	1.09	5.48	1.04	5.54	1.00	5.24
18+14	0.93	4.96	0.95	1.03	5.94	1.05	6.05	0.98	6.39	1.01	4.98	0.96	5.07	0.93	4.79

Training +Validation	Returns		SR	Value Weight				Q							
	Avg	<i>t-stat</i>		CAPM	FF3	Carhart	FF5	FF5+MOM	alpha	<i>t-stat</i>	alpha	<i>t-stat</i>			
12+12	0.40	2.36	0.29	0.46	2.16	0.46	2.61	0.66	3.72	0.68	3.11	0.80	3.79	0.69	2.96
10+10	0.41	2.55	0.32	0.45	2.21	0.48	2.81	0.58	3.46	0.75	3.59	0.81	3.67	0.76	3.54
12+10	0.41	2.57	0.32	0.45	2.17	0.47	2.70	0.61	3.55	0.78	3.78	0.85	4.00	0.79	3.86
14+10	0.40	2.36	0.31	0.43	2.00	0.47	2.53	0.60	3.33	0.76	3.52	0.84	3.79	0.77	3.58
14+12	0.37	2.11	0.27	0.39	1.80	0.40	2.24	0.60	3.32	0.57	2.72	0.71	3.31	0.59	2.61
16+12	0.39	2.17	0.29	0.44	1.99	0.48	3.00	0.66	3.97	0.67	3.57	0.80	3.97	0.69	3.64
18+12	0.41	2.15	0.30	0.45	1.93	0.48	2.82	0.64	3.71	0.66	3.34	0.77	3.78	0.66	3.42
14+14	0.48	2.41	0.37	0.56	2.27	0.64	3.46	0.71	4.08	0.92	4.68	0.96	4.38	0.93	4.68
16+14	0.55	2.62	0.41	0.60	2.36	0.67	3.46	0.73	4.03	0.94	4.56	0.97	4.39	0.93	4.48
18+14	0.51	2.34	0.38	0.56	2.11	0.63	3.14	0.69	3.64	0.91	4.17	0.93	4.03	0.88	4.11

This table shows the performance of BRT portfolios using alternative initial training period and validation periods. We report results for different specifications that vary in the number of years used for initial training and validation. For example, “14+10” refers to the specification that uses an initial training period of 14 years and a validation period of 10 years. All results are computed following the procedure described in Table 1. The first three columns report average monthly returns for the long-short portfolios as well as the associated Sharpe Ratios. The remaining columns report risk-adjusted returns—see Table 2 for details. The top panel reports results for equal-weighted returns. The bottom panel reports results for value-weighted returns. All returns are expressed in percent per month.

Table 8: Performance of Portfolios Sorted by BRT Predicted Returns – Subsets of Accounting Variables

# of X	Returns		Equal Weight						Q					
	Avg	t-stat	CAPM		FF3		Carhart		FF5+MOM					
			alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat		
30	0.30	1.03	0.33	1.10	0.24	0.88	0.46	1.85	0.27	0.84	0.42	1.42	0.32	0.86
60	0.72	2.34	0.80	2.58	0.71	2.39	0.78	3.07	0.54	1.71	0.60	2.11	0.52	1.57
90	0.74	3.22	0.92	3.84	0.83	4.34	0.67	3.42	0.67	2.99	0.57	2.46	0.52	2.00
120	0.56	2.93	0.65	3.58	0.65	3.63	0.63	3.14	0.76	3.34	0.73	2.99	0.71	2.89
150	1.00	6.09	1.16	7.33	1.09	7.39	1.07	6.76	0.92	5.26	0.91	5.09	0.84	4.33
180	1.29	6.56	1.43	7.45	1.38	7.24	1.29	6.96	1.16	5.67	1.12	5.41	1.13	5.09
210	1.29	6.55	1.46	7.28	1.39	7.32	1.26	6.65	1.05	5.21	0.99	4.63	0.98	4.25
240	0.92	6.33	1.00	6.44	0.99	6.23	1.00	6.15	0.93	4.83	0.94	4.91	0.85	4.45

# of X	Returns		Value Weight						Q					
	Avg	t-stat	CAPM		FF3		Carhart		FF5+MOM					
			alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat		
30	-0.12	-0.36	-0.01	-0.03	-0.10	-0.31	0.05	0.18	-0.06	-0.16	0.04	0.11	0.04	0.09
60	-0.09	-0.24	-0.03	-0.07	-0.16	-0.44	-0.03	-0.10	-0.28	-0.76	-0.17	-0.50	-0.25	-0.68
90	0.09	0.33	0.32	1.08	0.24	0.85	0.08	0.34	0.23	0.77	0.12	0.43	0.19	0.58
120	0.28	0.99	0.43	1.52	0.42	1.45	0.33	1.31	0.63	1.95	0.55	1.74	0.54	1.68
150	0.34	1.47	0.59	2.65	0.47	2.22	0.54	2.70	0.26	1.04	0.32	1.34	0.33	1.13
180	0.55	1.90	0.82	2.92	0.69	2.72	0.81	3.61	0.40	1.39	0.51	2.02	0.54	1.75
210	0.58	2.29	0.85	3.35	0.71	3.40	0.82	3.99	0.35	1.58	0.46	2.19	0.53	2.06
240	0.40	2.36	0.46	2.16	0.46	2.61	0.66	3.72	0.68	3.11	0.80	3.79	0.69	2.96

This table shows the performance of BRT portfolios when we vary the number of accounting variables (X) used to construct the fundamental signals. In particular, we progressively increase the number of fundamental variables used in the analysis from 30 to 240 on the basis of their missing rates. All results are computed following the procedure described in Table 1. The first three columns report average monthly returns for the long-short portfolios and the associated Sharpe Ratios. The remaining columns report risk-adjusted returns—see Table 2 for details. The top panel reports results for equal-weighted returns. The bottom panel reports results for value-weighted returns. All returns are expressed in percent per month.

Table 9: Performance of Portfolios Sorted by BRT Predicted Returns – Subsets of Financial Ratio Configurations

# of Y	Returns		SR	CAPM		FF3		Equal Weight Carhart		FF5		FF5+MOM		Q	
	Avg	t-stat		alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat
Y3	0.82	4.10	0.79	0.95	4.91	0.91	4.84	0.77	4.21	0.69	3.58	0.61	3.21	0.59	3.12
Y5	0.72	3.31	0.62	0.90	4.35	0.84	4.26	0.68	3.50	0.64	3.00	0.54	2.57	0.49	2.28
P1	0.69	3.33	0.63	0.79	3.70	0.74	3.51	0.62	3.53	0.43	2.13	0.36	2.02	0.37	1.94
P2	0.79	4.80	0.73	0.86	4.49	0.78	4.63	0.69	4.06	0.57	3.11	0.52	2.56	0.46	2.07
P3	0.85	4.50	0.89	1.01	5.55	0.90	5.81	0.81	5.01	0.63	4.95	0.59	4.32	0.61	4.08
P4	0.51	2.85	0.50	0.63	3.61	0.59	3.54	0.58	3.43	0.41	2.23	0.41	2.25	0.36	1.77
P5	0.87	4.28	0.86	1.01	5.30	0.91	5.40	0.83	4.32	0.61	4.43	0.58	3.82	0.62	3.64

# of Y	Returns		SR	CAPM		FF3		Value Weight Carhart		FF5		FF5+MOM		Q	
	Avg	t-stat		alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat
Y3	0.38	1.82	0.28	0.54	2.35	0.47	2.15	0.41	2.20	0.33	1.72	0.30	1.53	0.32	1.67
Y5	0.62	2.48	0.43	0.81	3.22	0.74	3.29	0.65	3.19	0.59	2.89	0.54	2.54	0.53	2.46
P1	0.31	1.20	0.23	0.40	1.47	0.40	1.42	0.28	1.13	0.21	0.91	0.14	0.65	0.26	1.04
P2	0.18	1.03	0.16	0.32	1.80	0.23	1.39	0.32	1.79	0.09	0.47	0.16	0.86	0.14	0.69
P3	0.15	0.69	0.12	0.39	2.05	0.29	1.44	0.18	0.88	0.07	0.34	0.02	0.08	0.01	0.03
P4	0.22	1.10	0.20	0.36	1.80	0.31	1.56	0.29	1.45	0.19	0.89	0.18	0.85	0.20	0.96
P5	0.31	1.18	0.23	0.57	2.52	0.44	2.09	0.39	1.59	0.02	0.12	0.03	0.13	0.07	0.32

This table shows the BRT performance of different subsets of financial ratio configurations. Row “Y3” builds signals using three base variables, i.e., total assets, total sales, and market capitalization. Row “Y5” builds signals using five base variables, i.e., total assets, total sales, market capitalization, total liability, and shareholder’s equity. Row “P1” denotes the financial ratios configuration #1-15 and #76; “P2” denotes #16-30 and #76; “P3” denotes #31-45 and #76; “P4” denotes #46-60 and #76; “P5” denotes #61-76. All results are computed following the procedure described in Table 1. The first three columns report average monthly returns for the long-short portfolios and the associated Sharpe Ratios. The remaining columns report risk-adjusted returns—see Table 2 for details. The top panel reports results for equal-weighted returns. The bottom panel reports results for value-weighted returns. All returns are expressed in percent per month.

Table 10: Predictability of Stock Returns across Different Economic and Market Conditions

Panel A. Sentiment					
Equal Weight			Value Weight		
High	Low	Diff	High	Low	Diff
1.12	0.80	0.32	0.37	0.52	-0.15
(4.31)	(3.18)	(0.89)	(0.91)	(1.86)	(-0.31)

Panel B. VIX					
Equal Weight			Value Weight		
High	Low	Diff	High	Low	Diff
0.74	0.99	-0.25	0.15	0.58	-0.43
(2.29)	(5.14)	(-0.68)	(0.34)	(2.30)	(-0.84)

Panel C. Liquidity					
Equal Weight			Value Weight		
High	Low	Diff	High	Low	Diff
0.97	0.88	0.09	0.64	0.16	0.48
(4.22)	(3.21)	(0.24)	(2.22)	(0.42)	(-0.98)

Panel D. Business Cycle					
Equal Weight			Value Weight		
Recession	Expansion	Diff	Recession	Expansion	Diff
1.42	0.88	0.54	1.00	0.34	0.66
(1.88)	(4.82)	(0.70)	(0.72)	(1.50)	(0.47)

Panel E. Past Market Returns					
Equal Weight			Value Weight		
UP	DOWN	Diff	UP	DOWN	Diff
1.23	0.63	0.60	0.78	0.03	0.75
(4.37)	(2.84)	(1.69)	(2.19)	(0.08)	(1.55)

Panel F. Calendar Subperiods					
Equal Weight			Value Weight		
1987-2003	2003-2019	Diff	1987-2003	2003-2019	Diff
1.05	0.80	0.25	0.32	0.48	-0.16
(3.41)	(4.40)	(0.69)	(0.76)	(1.97)	(-0.34)

This table reports the average long-short portfolio returns from the BRT model across subperiods sorted by economic and market conditions. For sentiment, VIX, liquidity and past market returns we split the sample period into two subperiods based on the median value of the conditioning variable. We then compute the performance of the BRT model across the different subperiods. In Panel D, we split the sample into recession and expansion based on the NBER recession indicator. In Panel F, we split the sample period into two halves: 1987-2003 and 2003-2019. In all cases, we report the performance of the long-short BRT portfolio in each subperiod. We also report whether the difference in long-short performance across different subperiods is statistically significant. All returns are expressed in percent per month.

Table 11: Performance of Portfolios Sorted by BRT Predicted Returns based on Technical Signals

Rank	Pred	Avg	t-stat	SD	SR	Equal Weighted		FF3		Carhart	
						CAPM	t-stat	alpha	t-stat	alpha	t-stat
1 (Low)	-0.01	0.23	0.79	7.17	0.11	-0.63	-3.40	-0.63	-5.50	-0.35	-3.21
2	0.04	0.39	1.60	6.25	0.22	-0.39	-2.78	-0.47	-6.31	-0.22	-2.83
3	0.06	0.58	2.65	5.54	0.36	-0.13	-1.02	-0.26	-4.28	-0.08	-1.23
4	0.07	0.74	3.74	5.14	0.50	0.07	0.59	-0.06	-1.17	0.05	0.86
5	0.07	0.78	4.07	4.89	0.55	0.12	1.14	-0.01	-0.35	0.06	1.34
6	0.08	0.79	4.18	4.82	0.57	0.14	1.34	0.00	0.09	0.06	1.74
7	0.08	0.87	4.56	4.83	0.62	0.21	2.07	0.09	1.83	0.12	2.70
8	0.09	0.97	5.02	5.00	0.67	0.30	2.83	0.17	3.91	0.19	4.22
9	0.10	1.02	4.98	5.31	0.67	0.32	2.77	0.19	3.85	0.19	3.92
10 (High)	0.13	1.17	4.89	5.94	0.68	0.42	2.76	0.29	4.14	0.27	3.64
10-1	0.14	0.94	6.33	3.34	0.97	1.05	6.60	0.92	7.21	0.62	5.08

Rank	Pred	Avg	t-stat	SD	SR	Value Weighted		FF3		Carhart	
						CAPM	t-stat	alpha	t-stat	alpha	t-stat
1 (Low)	0.00	0.38	1.39	7.00	0.19	-0.52	-3.07	-0.41	-3.35	-0.20	-1.49
2	0.04	0.44	2.04	5.88	0.26	-0.36	-3.25	-0.31	-3.43	-0.12	-1.54
3	0.06	0.51	2.78	5.12	0.35	-0.19	-2.86	-0.21	-3.29	-0.06	-0.87
4	0.07	0.60	3.35	4.78	0.43	-0.08	-1.33	-0.12	-2.04	-0.01	-0.18
5	0.07	0.62	3.89	4.42	0.48	-0.01	-0.17	-0.05	-0.84	-0.01	-0.15
6	0.08	0.57	3.57	4.28	0.46	-0.05	-1.13	-0.08	-1.76	-0.07	-1.43
7	0.08	0.71	4.50	4.32	0.57	0.09	2.11	0.08	1.82	0.05	1.04
8	0.09	0.70	4.45	4.38	0.55	0.06	1.46	0.05	1.14	0.01	0.25
9	0.10	0.81	4.79	4.73	0.60	0.14	2.58	0.13	2.54	0.05	0.87
10 (High)	0.12	0.95	4.92	5.19	0.63	0.22	2.54	0.24	3.01	0.13	1.91
10-1	0.13	0.56	3.34	4.30	0.45	0.74	4.39	0.65	4.41	0.33	2.20

This table reports the excess returns of decile portfolios sorted by BRT predicted returns based on technical signals from 1950 to 2019. We predict stock monthly excess returns using 238 technical signals (as described in Section 5). We use a recursive window approach and select the optimal hyper-parameters using a cross-validation approach. Our initial estimation period is 1926-1950. The first 12 years (or 144 months) is the training period, and the second 12 years (or 144 months) is the validation period. As we roll forward, the training period expands while the validation period stays at 12 years. The upper panel reports equal-weighted portfolio results. In this panel's first column, we report the average predicted monthly returns from the BRT model (*Pred*). The second and third columns report the average realized monthly excess returns (*Avg*) and associated *t*-statistics (*t-stat*), computed using [Newey and West \(1987\)](#) standard errors with 12 lags. Finally, in the fourth and fifth column we report the portfolios' return standard deviations (*SD*) and Sharpe ratios (*SR*), respectively. The remaining columns show the risk-adjusted performance of the BRT portfolios based on the CAPM model, the Fama-French 3-factor model, and the Carhart 4-factor model. The lower panel reports the results for value-weighted portfolio returns. All returns are expressed in percent per month.

Table 12: In- and Out-of-sample Return Predictability on Technical Signals

Panel A: In- and Out-of-Sample R^2			
	In-Sample (i)	In-Sample (ii)	Out-of-Sample
Mean	1.00	1.45	0.50
S.D.	2.23	3.76	0.95

Panel B: Equal-weighted Long-short Returns			
	In-Sample (i)	In-Sample (ii)	Out-of-Sample
Mean	1.92	3.14	0.94
t -stat	10.98	14.30	6.33

Panel C: Value-weighted Long-short Returns			
	In-Sample (i)	In-Sample (ii)	Out-of-Sample
Mean	1.32	1.90	0.56
t -stat	7.04	9.52	3.34

This table reports the in- and out-of-sample performance of BRT forecasts based on technical signals. The first and second columns of Panel A report the means and standard deviations of the in-sample R^2 . In the first column, the optimal hyper-parameters are determined using 1926-1950 as the training period. In the second column, the optimal hyper-parameters are determined using a leave-one-year-out cross validation procedure over the period 1950-2019, following [Martin and Nagel \(2021\)](#). The third column reports the out-of-sample R^2 . Panels B and C maintain the same structure but report the average monthly returns and their t -statistics associated with equal-weighted (Panel B) and value-weighted (Panel C) long-short portfolios. All returns and R^2 are expressed in percent.

Table 13: Performance of Portfolios Sorted by NN Predicted Returns on Technical Predictors

Method	Returns		SR	Equal Weighted CAPM		FF3		Carhart	
	<i>Avg</i>	<i>t-stat</i>		<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>
NN1	0.51	3.50	0.47	0.56	3.40	0.56	3.93	0.14	1.04
NN2	0.22	1.83	0.25	0.28	2.09	0.27	2.42	-0.01	-0.05
NN3	0.19	1.77	0.23	0.17	1.64	0.24	2.31	0.05	0.40
NN4	0.07	0.69	0.08	0.03	0.25	0.11	1.23	-0.06	-0.61
NN5	-0.15	-0.98	-0.19	-0.24	-1.45	-0.14	-1.03	-0.24	-1.38

Method	Returns		SR	Value Weighted CAPM		FF3		Carhart	
	<i>Avg</i>	<i>t-stat</i>		<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>
NN1	0.36	2.16	0.28	0.41	2.21	0.44	2.71	-0.05	-0.33
NN2	0.30	2.16	0.26	0.35	2.46	0.36	2.67	-0.05	-0.37
NN3	0.31	2.07	0.27	0.21	1.54	0.30	2.32	0.04	0.30
NN4	0.25	1.94	0.22	0.14	1.04	0.21	2.08	-0.04	-0.38
NN5	0.13	0.85	0.12	-0.04	-0.29	0.05	0.43	-0.07	-0.50

This table shows the performance of portfolios sorted by Neural Network (NN) predicted returns based on technical signals. We consider NN models with hidden layers that range from 1 through 5. The first three columns report average monthly returns for the long-short portfolios and the associated Sharpe Ratios (SR), and the remaining columns report risk-adjusted returns. The top panel reports results for equal-weighted returns. The bottom panel reports results for value-weighted returns. All returns are expressed in percent per month.

Appendix A: Grids of Hyper-Parameters for Cross Validation

BRT	NN
# of iteration $\in \{100, 250, 500, 750, 1000\}$	L1 penalty $\lambda_1 \in \{10^{-5}, 10^{-3}\}$
learning rate $\in \{0.01, 0.05, 0.1\}$	Learning Rate LR $\in \{0.001, 0.01\}$
	Batch Size = 10000
	Epochs = 100
	Patience = 5
	Adam Para. = Default

This table shows the grids of hyper-parameters used in the cross validation of Boosted Regression Trees (BRT) and Neural Networks (NN).

Appendix B: List of Accounting Variables

#	Variable	Description	Missing Rate	Start Year
1	ACCHG	Accounting changes - cumulative effect	39.29%	1988
2	ACO	Current assets other total	0.76%	1963
3	ACOX	Current assets other sundry	2.20%	1963
4	ACT	Current assets - total	2.13%	1963
5	AM	Amortization of intangibles	33.03%	1965
6	AO	Assets – other	0.06%	1963
7	AOLOCH	Assets and liabilities other net change	38.36%	1988
8	AOX	Assets – other - sundry	2.22%	1963
9	AP	Accounts payable – trade	4.88%	1963
10	APALCH	Accounts payable & accrued liabilities increase/decrease	53.14%	1988
11	AQC	Acquisitions	12.98%	1972
12	AQI	Acquisitions income contribution	32.50%	1975
13	AQS	Acquisitions sales contribution	32.26%	1975
14	AT	Assets – total	0.01%	1963
15	BAST	Average short-term borrowing	74.28%	1978
16	CAPS	Capital surplus/share premium reserve	2.08%	1963
17	CAPX	Capital expenditure	2.18%	1963
18	CAPXV	Capital expenditure PPE Schedule V	1.39%	1963
19	CEQ	Common/ordinary equity - total	1.54%	1963
20	CEQL	Common equity liquidation value	1.62%	1963
21	CEQT	Common equity tangible	1.64%	1963
22	CH	Cash	12.33%	1963
23	CHE	Cash and short-term investments	0.72%	1963
24	CHECH	Cash and cash equivalents increase/decrease	28.77%	1972
25	CLD2	Capitalized leases - due in 2nd year	46.55%	1985
26	CLD3	Capitalized leases - due in 3rdyear	46.44%	1985
27	CLD4	Capitalized leases - due in 4thyear	46.18%	1985
28	CLD5	Capitalized leases - due in 5thyear	46.15%	1985
29	COGS	Cost of goods sold	0.09%	1963
30	CSTK	Common/ordinary stock (capital)	1.96%	1963
31	CSTKCV	Common stock-carrying value	28.31%	1963
32	CSTKE	Common stock equivalents – dollar savings	0.06%	1963
33	DC	Deferred charges	28.45%	1965
34	DCLO	Debt capitalized lease obligations	10.08%	1965
35	DCOM	Deferred compensation	72.02%	1980
36	DCPSTK	Convertible debt and stock	2.85%	1963
37	DCVSR	Debt senior convertible	9.89%	1970
38	DCVSUB	Debt subordinated convertible	11.96%	1970
39	DCVT	Debt – convertible	5.80%	1963
40	DD	Debt debentures	10.55%	1965
41	DD1	Long-term debt due in one year	5.05%	1963
42	DD2	Debt Due in 2nd Year	23.27%	1974
43	DD3	Debt Due in 3rd Year	23.32%	1974
44	DD4	Debt Due in 4th Year	23.16%	1974
45	DD5	Debt Due in 5th Year	24.04%	1974
46	DFS	Debt finance subsidiary	79.68%	1992
47	DFXA	Depreciation of tangible fixed assets	65.07%	1970
48	DILADJ	Dilution adjustment	62.54%	1994
49	DILAVX	Dilution available excluding extraordinary items	62.54%	1994
50	DLC	Debt in current liabilities - total	0.72%	1963
51	DLCCH	Current debt changes	60.86%	1974
52	DLTIS	Long-term debt issuance	10.50%	1972

#	Variable	Description	Missing Rate	Start Year
53	DLTO	Other long-term debt	9.96%	1965
54	DLTP	Long-term debt tied to prime	38.66%	1975
55	DLTR	Long-term debt reduction	9.84%	1972
56	DLTT	Long-term debt - total	0.20%	1963
57	DM	Debt mortgages & other secured	33.76%	1981
58	DN	Debt notes	10.56%	1965
59	DO	Income (loss) from discontinued operations	3.66%	1963
60	DONR	Nonrecurring discontinued operations	71.10%	1994
61	DP	Depreciation and amortization	0.24%	1963
62	DPACT	Depreciation, depletion and amortization	0.44%	1963
63	DPC	Depreciation and amortization (cash flow)	8.59%	1972
64	DPVIEB	Depreciation ending balance (schedule VI)	19.34%	1970
65	DPVIO	Depreciation other changes (schedule VI)	65.12%	1970
66	DPVIR	Depreciation retirements (schedule VI)	65.14%	1970
67	DRC	Deferred revenue current	73.42%	1994
68	DS	Debt-subordinated	9.93%	1965
69	DUDD	Debt unamortized debt discount and other	29.51%	1963
70	DV	Cash dividends (cash flow)	8.55%	1972
71	DVC	Dividends common/ordinary	0.11%	1963
72	DVP	Dividends - preferred/preference	0.06%	1963
73	DVPA	Preferred dividends in arrears	17.95%	1964
74	DVPIBB	Depreciation beginning balance (schedule VI)	60.82%	1970
75	DVT	Dividends - total	0.11%	1963
76	DXD2	Debt (excl capitalized leases) due in 2nd year	49.31%	1985
77	DXD3	Debt (excl capitalized leases) due in 3rd year	49.25%	1985
78	DXD4	Debt (excl capitalized leases) due in 4th year	48.96%	1985
79	DXD5	Debt (excl capitalized leases) due in 5th year	49.36%	1985
80	EBIT	Earnings before interest and taxes	1.36%	1963
81	EBITDA	Earnings before interest	0.21%	1963
82	ESOPCT	ESOP obligation (common) - total	40.69%	1980
83	ESOPDLT	ESOP debt - long term	49.09%	1990
84	ESOPT	Preferred ESOP obligation - total	41.01%	1964
85	ESUB	Equity in earnings - unconsolidated subsidiaries	12.33%	1963
86	ESUBC	Equity in net loss earnings	22.05%	1972
87	EXRE	Exchange rate effect	38.46%	1988
88	FATB	Property, plant, and equipment buildings	51.33%	1985
89	FATC	Property, plant and equipment construction in progress	47.36%	1985
90	FATE	Property, plant, equipment and machinery equipment	53.32%	1985
91	FATL	Property, plant, and equipment leases	57.58%	1985
92	FATN	Property, plant, equipment and natural resources	47.37%	1985
93	FATO	Property, plant, and equipment other	52.84%	1985
94	FATP	Property, plant, equipment and land improvements	51.25%	1985
95	FIAO	Financing activities other	38.35%	1988
96	FINCF	Financing activities net cash flow	38.35%	1988
97	FOPO	Funds from operations other	7.83%	1972
98	FOPOX	Funds from operations - Other excl option tax benefit	76.37%	1992
99	FOPT	Funds from operations total	69.42%	1972
100	FSRCO	Sources of funds other	70.81%	1972
101	FSRCT	Sources of funds total	71.27%	1972
102	FUSEO	Uses of funds other	70.81%	1972
103	FUSET	Uses of funds total	71.61%	1972
104	GDWL	Goodwill	47.13%	1989
105	GP	Gross profit (loss)	0.09%	1963
106	IB	Income before extraordinary items	0.05%	1963

#	Variable	Description	Missing Rate	Start Year
107	IBADJ	IB adjusted for common stock equivalents	0.05%	1963
108	IBC	Income before extraordinary items (cash flow)	7.82%	1972
109	IBCOM	Income before extraordinary items available for common	0.05%	1963
110	ICAPT	Invested capital – total	1.54%	1963
111	IDIT	Interest and related income - total	42.18%	1965
112	INTAN	Intangible assets – total	10.02%	1963
113	INTC	Interest capitalized	16.78%	1963
114	INTPN	Interest paid net	43.82%	1988
115	INVCH	Inventory decrease (increase)	43.46%	1988
116	INVFG	Inventories finished goods	41.28%	1970
117	INVO	Inventories other	52.52%	1984
118	INVRM	Inventories raw materials	40.27%	1969
119	INVT	Inventories – total	1.43%	1963
120	IN VWIP	Inventories work in progress	43.69%	1970
121	ITCB	Investment tax credit (balance sheet)	3.20%	1963
122	ITCI	Investment tax credit (income account)	37.65%	1963
123	IVACO	Investing activities other	38.35%	1988
124	IVAEQ	Investment and advances – equity	9.07%	1963
125	IVAO	Investment and advances other	7.07%	1963
126	IVCH	Increase in investments	13.68%	1972
127	IVNCF	Investing activities net cash flow	38.35%	1988
128	IVST	Short-term investments – total	12.35%	1963
129	IVSTCH	Short-term investments change	48.38%	1988
130	LCO	Current liabilities other total	4.76%	1963
131	LCOX	Current liabilities other sundry	6.10%	1963
132	LCOXDR	Current liabilities-other-excl deferred revenue	72.40%	1994
133	LCT	Current liabilities – total	1.69%	1963
134	LIFR	LIFO reserve	22.04%	1976
135	LO	Liabilities – other – total	0.72%	1963
136	LT	Liabilities – total	0.50%	1963
137	MIB	Minority interest (balance sheet)	6.37%	1963
138	MII	Minority interest (income account)	10.24%	1963
139	MRC1	Rental commitments minimum 1st year	27.85%	1975
140	MRC2	Rental commitments minimum 2nd year	28.34%	1975
141	MRC3	Rental commitments minimum 3rd year	28.46%	1975
142	MRC4	Rental commitments minimum 4th year	28.61%	1975
143	MRC5	Rental commitments minimum 5th year	30.38%	1975
144	MRCT	Rental commitments minimum 5 year total	29.51%	1975
145	MSA	Marketable securities adjustment	18.18%	1976
146	NI	Net income (loss)	0.06%	1963
147	NIADJ	Net income adjusted for common stock equiv.	2.24%	1963
148	NIECI	Net income effect capitalized interest	59.92%	1976
149	NOPI	Non-operating income (expense)	0.10%	1963
150	NOPIO	Non-operating income (expense) other	0.10%	1963
151	NP	Notes payable short-term borrowings	0.80%	1963
152	OANCF	Operating activities net cash flow	38.36%	1988
153	OB	Order backlog	64.22%	1971
154	OIADP	Operating income after depreciation	0.07%	1963
155	PI	Pre-tax income	0.06%	1963
156	PIDOM	Pretax income domestic	74.94%	1981
157	PIFO	Pretax income foreign	75.36%	1981
158	PPEGT	Property, plant, and equipment – total (gross)	0.45%	1963
159	PPENB	Property, plant, and equipment buildings (net)	70.38%	1970
160	PPENC	Property plant equipment construction in progress (net)	65.66%	1970

#	Variable	Description	Missing Rate	Start Year
161	PPENLI	Property plant equipment land and improvements (net)	70.26%	1970
162	PPENME	Property plant equipment machinery and equipment (net)	69.73%	1970
163	PPENNR	Property plant equipment natural resources (net)	69.31%	1970
164	PPENO	Property plant and equipment other (net)	69.26%	1970
165	PPENT	Property, plant, and equipment – total (net)	0.11%	1963
166	PPEVBB	Property plant equipment beginning balance (schedule V)	57.03%	1963
167	PPEVEB	Property, plant, and equipment ending balance	15.25%	1963
168	PPEVO	Property, plant, and equipment other changes (schedule V)	62.50%	1963
169	PPEVR	Property, plant and equipment retirements (schedule V)	62.50%	1963
170	PRSTKC	Purchase of common and preferred stock	12.98%	1972
171	PSTK	Preferred/preference stock (capital) – total	0.24%	1963
172	PSTKC	Preferred stock convertible	4.96%	1963
173	PSTKL	Preferred stock liquidating value	0.05%	1963
174	PSTKN	Preferred/preference stock – non-redeemable	1.48%	1963
175	PSTKR	Preferred/preference stock - redeemable	20.89%	1964
176	PSTKRV	Preferred stock redemption value	0.06%	1963
177	RDIP	In process R&D expense	65.68%	1994
178	RE	Retained earnings	2.04%	1963
179	REA	Retained earnings restatement	10.33%	1970
180	REAJO	Retained earnings other adjustments	30.06%	1983
181	RECCH	Accounts receivable decrease (increase)	41.58%	1988
182	RECCO	Receivables – current – other	3.21%	1963
183	RECD	Receivables – estimated doubtful	29.03%	1970
184	RECT	Receivables – total	1.45%	1963
185	RECTA	Retained earnings cumulative translation adjustment	30.39%	1983
186	RECTR	Receivables – trade	17.96%	1967
187	REUNA	Retained earnings unadjusted	29.89%	1983
188	SALE	Sales/turnover (net)	0.05%	1963
189	SEQ	Stockholders' equity – total	2.24%	1963
190	SIV	Sale of investments	16.24%	1972
191	SPI	Special items	3.93%	1963
192	SPPE	Sale of property	28.92%	1972
193	SPPIV	Sale of property plant equipment investments gain (loss)	38.36%	1988
194	SSTK	Sale of common and preferred stock	9.55%	1972
195	TLCF	Tax loss carry forward	23.48%	1963
196	TSTK	Treasury stock – total (all capital)	16.37%	1970
197	TSTKC	Treasury stock – common	26.38%	1974
198	TSTKP	Treasury stock – preferred	25.51%	1963
199	TXACH	Income taxes accrued increase/decrease	56.69%	1988
200	TXBCO	Excess tax benefit stock options -cash flow	76.43%	1992
201	TXC	Income tax – current	16.78%	1963
202	TXDB	Deferred taxes (balance sheet)	3.34%	1963
203	TXDBA	Deferred tax asset - long term	73.84%	1993
204	TXDBCA	Deferred tax asset - current	73.11%	1994
205	TXDBCL	Deferred tax liability - current	74.46%	1994
206	TXDC	Deferred taxes (cash flow)	10.38%	1972
207	TXDFED	Deferred taxes-federal	48.37%	1985
208	TXDFO	Deferred taxes-foreign	45.98%	1985
209	TXDI	Income tax – deferred	6.99%	1963
210	TXDITC	Deferred taxes and investment tax credit	3.34%	1963
211	TXDS	Deferred taxes-state	48.91%	1985
212	TXFED	Income tax federal	16.78%	1963
213	TXFO	Income tax foreign	19.02%	1970
214	TXNDB	Net deferred tax asset (liab) - total	69.95%	1994

#	Variable	Description	Missing Rate	Start Year
215	TXNDBA	Net deferred tax asset	72.66%	1994
216	TXNDBL	Net deferred tax liability	72.67%	1994
217	TXNDBR	Deferred tax residual	72.05%	1994
218	TXO	Income taxes - other	33.11%	1963
219	TXP	Income tax payable	5.93%	1963
220	TXPD	Income taxes paid	45.36%	1988
221	TXR	Income tax refund	10.40%	1963
222	TXS	Income tax state	17.76%	1963
223	TXT	Income tax total	0.06%	1963
224	TXW	Excise taxes	24.39%	1976
225	WCAP	Working capital (balance sheet)	2.15%	1963
226	WCAPC	Working capital change other increase/decrease	72.51%	1972
227	WCAPCH	Working capital change total	74.62%	1972
228	XACC	Accrued expenses	19.16%	1963
229	XAD	Advertising expense	64.98%	1963
230	XDEPL	Depletion expense (schedule VI)	68.80%	1970
231	XI	Extraordinary items	1.60%	1963
232	XIDO	Extra. items and discontinued operations	0.06%	1963
233	XIDOC	Extra. items and disc. operations (cash flow)	9.44%	1972
234	XINT	Interest and related expenses – total	5.05%	1963
235	XOPR	Operating expenses – total	0.09%	1963
236	XPP	Prepaid expenses	43.96%	1963
237	XPR	Pension and retirement expense	25.03%	1963
238	XRD	Research and development expense	47.01%	1963
239	XRENT	Rental expense	14.34%	1963
240	XSGA	Selling, general and administrative expense	12.13%	1963

This table lists the 240 accounting variables used in this study and their descriptions. Our sample period is 1963-2019. We begin with all accounting variables on the balance sheet, income statement, and cash flow statement included in the annual Compustat database. We exclude all variables with fewer than 20 years of data or fewer than 1,000 firms with non-missing data on average per year. We exclude per-share-based variables such as book value per share and earnings per share. We remove LSE (total liabilities and equity), REVT (total revenue), OIBDP (operating income before depreciation), and XDP (depreciation expense) because they are identical to TA (total assets), SALE (total sale), EBITDA (earnings before interest) and DFXA (depreciation of tangible fixed assets) respectively. Please refer to [Yan and Zheng \(2017\)](#) for more details.

Appendix C: List of Financial Ratios and Configurations

#	Description	#	Description	#	Description	#	Description
1	X/AT	16	Δ in X/AT	31	$\% \Delta$ in X/AT	46	$\Delta X/LAGAT$
2	X/ACT	17	Δ in X/ACT	32	$\% \Delta$ in X/ACT	47	$\Delta X/LAGACT$
3	X/INVT	18	Δ in X/INVT	33	$\% \Delta$ in X/INVT	48	$\Delta X/LAGINVT$
4	X/PPENT	19	Δ in X/PPENT	34	$\% \Delta$ in X/PPENT	49	$\Delta X/LAGPPENT$
5	X/LT	20	Δ in X/LT	35	$\% \Delta$ in X/LT	50	$\Delta X/LAGLT$
6	X/LCT	21	Δ in X/LCT	36	$\% \Delta$ in X/LCT	51	$\Delta X/LAGLCT$
7	X/DLTT	22	Δ in X/DLTT	37	$\% \Delta$ in X/DLTT	52	$\Delta X/LAGDLTT$
8	X/CEQ	23	Δ in X/CEQ	38	$\% \Delta$ in X/CEQ	53	$\Delta X/LAGCEQ$
9	X/SEQ	24	Δ in X/SEQ	39	$\% \Delta$ in X/SEQ	54	$\Delta X/LAGSEQ$
10	X/ICAPT	25	Δ in X/ICAPT	40	$\% \Delta$ in X/ICAPT	55	$\Delta X/LAGICAPT$
11	X/SALE	26	Δ in X/SALE	41	$\% \Delta$ in X/SALE	56	$\Delta X/LAGSALE$
12	X/COGS	27	Δ in X/COGS	42	$\% \Delta$ in X/COGS	57	$\Delta X/LAGCOGS$
13	X/XSGA	28	Δ in X/XSGA	43	$\% \Delta$ in X/XSGA	58	$\Delta X/LAGXSGA$
14	X/EMP	29	Δ in X/EMP	44	$\% \Delta$ in X/EMP	59	$\Delta X/LAGEMP$
15	X/MKTCAP	30	Δ in X/MKTCAP	45	$\% \Delta$ in X/MKTCAP	60	$\Delta X/LAGMKTCAP$
						61	$\% \Delta$ in X - $\% \Delta$ in AT
						62	$\% \Delta$ in X - $\% \Delta$ in ACT
						63	$\% \Delta$ in X - $\% \Delta$ in INVT
						64	$\% \Delta$ in X - $\% \Delta$ in PPENT
						65	$\% \Delta$ in X - $\% \Delta$ in LT
						66	$\% \Delta$ in X - $\% \Delta$ in LCT
						67	$\% \Delta$ in X - $\% \Delta$ in DLTT
						68	$\% \Delta$ in X - $\% \Delta$ in CEQ
						69	$\% \Delta$ in X - $\% \Delta$ in SEQ
						70	$\% \Delta$ in X - $\% \Delta$ in ICAPT
						71	$\% \Delta$ in X - $\% \Delta$ in SALE
						72	$\% \Delta$ in X - $\% \Delta$ in COGS
						73	$\% \Delta$ in X - $\% \Delta$ in XSGA
						74	$\% \Delta$ in X - $\% \Delta$ in EMP
						75	$\% \Delta$ in X - $\% \Delta$ in MKTCAP
						76	$\% \Delta$ in X

This table lists the 76 financial ratios and configurations used in this study. Our sample period is 1963-2019. We begin with all accounting variables on the balance sheet, income statement, and cash flow statement included in the annual Compustat database. We exclude all variables with fewer than 20 years of data or fewer than 1,000 firms with non-missing data on average per year. We exclude per-share-based variables such as book value per share and earnings per share. “X” represents the 240 accounting variables listed in Appendix B. “Y” represents the fifteen base variables, including AT (total assets), ACT (total current assets), INVT (inventory), PPENT (property, plant, and equipment), LT (total liabilities), LCT (total current liabilities), DLTT (long-term debt), CEQ (total common equity), SEQ (stockholders’ equity), ICAPT (total invested capital), SALE (total sale), COGS (cost of goods sold), XSGA (selling, general, and administrative cost), EMP (number of employees), and MKTCAP (market capitalization). Please refer to [Yan and Zheng \(2017\)](#) for more details.

Appendix D: In-sample Risk-adjusted Performance of BRT Portfolios

Returns		Panel A. Choose optimal hyper-parameter using 1963 - 1986											
		CAPM		FF3		Carhart		FF5		FF5+MOM		Q	
<i>Avg</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>
EW	3.21	3.38	6.85	3.33	6.13	2.82	6.98	2.99	4.67	2.67	5.54	2.83	3.90
VW	2.02	4.42	2.22	4.64	2.13	4.30	1.76	4.65	1.94	3.54	1.70	4.15	3.40
		SR	1.56	1.00									

Returns		Panel B. Choose optimal hyper-parameter using leave-one-year-out cross validation on 1987 - 2019											
		CAPM		FF3		Carhart		FF5		FF5+MOM		Q	
<i>Avg</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>	<i>alpha</i>	<i>t-stat</i>
EW	4.37	8.07	4.56	8.69	4.52	8.08	3.91	8.96	4.21	6.29	3.80	7.43	5.20
VW	3.10	6.26	3.34	6.76	3.27	6.54	2.77	6.80	3.05	5.50	2.72	6.45	5.02
		SR	2.00	1.39									

This table shows the risk-adjusted in-sample performance of BRT portfolios. The first three columns report average monthly returns for the long-short portfolios and the associated Sharpe Ratios. The remaining columns report risk-adjusted returns—see Table 2 for details. The results are reported for both equal-weighted and value-weighted portfolios. In Panel A, the optimal hyper-parameters are determined using 1962-1986 as the training period. In Panel B, the optimal hyper-parameters are determined using a leave-one-year-out cross-validation procedure over the period 1987-2019, following [Martin and Nagel \(2021\)](#).

Appendix E. Relative Influence Measures

One criticism of machine learning algorithms is that they are “Black Boxes” that do not provide a lot of intuition to the researcher and the reader. This criticism hardly applies to BRTs that feature very useful and intuitive visualization tools. The first commonly used measure is referred to as the “relative influence” measure. Consider the reduction in the empirical error every time one of the covariates x_l , is used to split the tree. Summing the reductions in empirical errors (or improvements in fit) across the nodes in the tree gives a measure of the variable’s influence (Breiman, Friedman, Stone, and Olshen, 1984):

$$I_l(\mathcal{T}) = \sum_{j=2}^J \Delta e(j)^2 I(x(j) = l),$$

where $\Delta e(j)^2 = T^{-1} \sum_{t=1}^T (e_t(j-1)^2 - e_t(j)^2)$ is the reduction in the squared empirical error at the j^{th} node and $x(j)$ is the regressor chosen at this node, so $I(x(j) = l)$ equals 1 if regressor l is chosen, and 0 otherwise. The sum is computed across all observations, $t = 1, \dots, T$, and over the $J - 1$ internal nodes of the tree.

The rationale for this measure is that at each node, one of the regressors gets selected to partition the sample space into two sub-states. The particular regressor at node j achieves the greatest reduction in the empirical risk of the model fitted up to node $j - 1$. The importance of each regressor, x_l , is the sum of the reductions in the empirical errors computed over all internal nodes for which it was chosen as the splitting variable. If a regressor never gets chosen to conduct the splits, its influence is zero. Conversely, the more frequently a regressor is used for splitting, and the bigger its effect on reducing the model’s empirical risk, the larger its influence.

This measure of influence can be generalized by averaging over the number of boosting iterations, B , which generally provides a more reliable measure of influence:

$$\bar{I}_l = \frac{1}{B} \sum_{b=1}^B I_l(\mathcal{T}_b).$$

This is best interpreted as a measure of relative influence that can be compared across regressors. We therefore report the following measure of relative influence, \overline{RI}_l , which sums to 1:

$$\overline{RI}_l = \bar{I}_l / \sum_{l=1}^L \bar{I}_l.$$

The figure below shows the relative influence of the top 25 signals in the baseline BRT model

estimated in the paper. We first compute the signals' relative influence in each year of the test period, 1987-2019, and average their values across all test years. Note that the relative importance measure across all signals sums to one every year. We then rank and plot the signals according to their average relative influence. The Y-axis reports the 25 most important signals, while the x-axis presents each signal's average relative influence measure.

