Selecting Mutual Funds from the Stocks They Hold: a Machine Learning Approach*

Bin Li [†]
Wuhan University

Alberto Rossi [‡]
Georgetown University

December 13, 2021

Abstract

We combine individual mutual fund holdings and a large number of stock characteristics (factors) to compute fund-level exposures to factors on the basis of the stocks they hold. Fund performance is non-linearly related to fund factor exposures and their interactions. This feature proves important when we predict fund performance, as machine learning methods such as boosted regression trees (BRTs) significantly outperform standard linear frameworks and the BRT-generated forecasts encompass the ones generated by the predictors of mutual fund performance that have been proposed in the literature so far. Finally, factor exposures explain the vast majority of mutual fund performance.

*We would like to thank Jennie Bai, Turan Bali, Azi Ben-Raphael, Zhi Chen, Serdar Dinc, Francesco D'Acunto, Gavin Feng, Antonio Gargano, Gerard Hoberg, Zhengzi Li, Rose Liao, David Mclean, Stanislav Sokolinski, Andrea Tamoni, Simon Trimborn, Stephen Utkus, Michael Weber, Quan Wen, Xiaofei Zhao, and the seminar participants at Central University of Finance and Economics, China FinTech Research Conference, the 5th Shanghai-Edinburgh-London Fintech Conference, City University of Hong Kong, Georgetown University, Hunan University, Rutgers University, and Wuhan University for valuable comments. Bin Li acknowledges financial support from the National Natural Science Foundation of China (No. 71971164). All errors are our own.

[†]Economics and Management School, Wuhan University, P.R. China. Email: binli.whu@whu.edu.cn.

[‡]McDonough School of Business, Georgetown University, Washington, DC, USA. Fellow of the Luohan Academy. Email: agr60@georgetown.edu.

Introduction

The mutual fund industry is a multi-trillion-dollar industry that has been thriving for decades.[1] Within this industry, active management has been studied for years in an attempt to assess whether fund managers display systematically superior performance—e.g. Kosowski, Timmermann, Wermers, and White (2006) and Fama and French (2010)—and whether successful mutual fund managers can be selected in real time, see Barras, Scaillet, and Wermers (2010), Harvey and Liu (2018) and Groenborg, Lunde, Timmermann, and Wermers (2021). As for the sources of outperformance, most of the literature focuses on managers' ability to select stocks and generate abnormal returns in excess of the ones generated by known factors, such as size, value and momentum. The evidence shows that a nontrivial percentage of mutual fund managers indeed outperforms static benchmarks in that they are able to select stocks that do better compared to the ones that have similar size, book-to-market and momentum characteristics—see Daniel, Grinblatt, Titman, and Wermers (1997), Amihud and Goyenko (2013), Hoberg, Kumar, and Prabhala (2018a) and Avramov, Cheng, and Hameed (2019), among others.

At the same time, a very large literature in asset pricing has shown that value, size and momentum are not the only sources of abnormal returns and many additional factors have been uncovered over the years (Green, Hand, and Zhang, 2017 and Hou, Xue, and Zhang, 2020). The literature also shows that these factors are often transient and deliver abnormal returns for a number of years, but their performance dissipates as these novel strategies become widely known across the investment community (McLean and Pontiff, 2016).

Evaluating managers' skills by focusing on their exposure to a small and fixed set of static factors as in Daniel et al. (1997) could potentially confound managers' stock selectivity and factor timing skills: the econometrician may conclude a manager has stock selectivity skills using a three-factor model when in fact a manager is exposed to a fourth factor that is omitted from the benchmark computations. At the same time, it is impossible to construct DGTW benchmarks based on a larger number of factors following the procedure in Daniel et al. (1997) because of the course of dimensionality. Assuming a universe of 6,000 stocks to invest in, the benchmark returns based on three factors result in 125 DGTW portfolios with 48 stocks each. If we were to increase the number of factors included to 5, the procedure would require the construction of 3,125 portfolios with less than 2 stocks each.

More broadly, relatively little is known regarding how mutual fund managers expose their portfolios to asset pricing factors. Edelen, Ince, and Kadlec (2016) focus on seven factors and find that, on average, mutual funds are negatively exposed to such factors at the annual horizon, but they are positively exposed to them at the quarterly horizon. Calluzzo, Moneta, and Topaloglu (2019) focus on 14 factors and show institutional investors increase their factor-based trading when information about these factors becomes widely available through academic publication. However, several key questions remain unanswered: if mutual funds have time-varying exposures to factors

[1] According to the Investment Company Institute, the US mutual fund industry has total net assets of $21.3 trillion.

and factors have time-varying returns, can investors select mutual funds with superior performance in real time, based on the characteristics of the stocks they hold? Are traditional statistical methods suitable for this task or is the adoption of more sophisticated machine learning methods necessary? Finally, and more importantly, once we control for exposure to a large number of factors, do mutual fund managers possess stock picking skills, or is mutual funds' performance completely explainable by their factor exposure? Our paper provide novel insights to these questions.

We start with mutual fund holdings and merge them with a large number of stock characteristics (94 in total) to construct characteristics exposure at the mutual fund level. We first show that, on average, mutual funds are significantly exposed to a large number of stock characteristics. For example, they tend to trade large, liquid, and older companies. They also tend to hold stocks with high operating profits, high return on equities, and high cash-to-debt ratios. On the other hand, mutual funds tend to shy away from illiquid stocks, stocks with high idiosyncratic volatility, and value stocks. When we unpack the cross section, we find the majority of the funds in our sample are negatively exposed to between 20 and 30 characteristics and are positively exposed to between 30 and 40 characteristics.

That mutual funds are positively and negatively exposed to specific stock characteristics is a clear indication that explaining funds' over- and under-performance on the basis of the characteristics of the stocks they hold should be possible. The fact that each fund is exposed both positively and negatively to a wide range of stock characteristics is instead an indication that traditional tools, such as portfolio sorts, may not be suitable for an investor that wants to select high-performing mutual funds in real time. Our tests indeed suggest selecting mutual funds on the basis of univariate sorts delivers disappointing results. Only six variables out of 94 deliver statistically significant long-short portfolios, but the economic magnitudes are small—only 2%-3% per year—and the results disappear when we focus on risk-adjusted returns.

When we attempt to select high-performing mutual funds on the basis of their holdings characteristics, we face a high-dimensional cross-sectional prediction problem. We have up to 94 predictors at every point in time, and a number of funds that vary between a little more than 100 at the beginning of the sample (1980) to several thousands toward the end of the sample (2018). Theory also cannot guide us as to whether the relation between fund returns and each regressor is linear or non-linear or whether there are significant interactions between the predictor variables at hand. We overcome these limitations resorting to a machine learning tool known as boosted regression trees (BRTs) (Friedman, 2001).

We focus primarily on BRTs for a variety of reasons. First, BRTs have exhibited strong predictive performance in various fields, including finance. Second, BRTs can handle large, high-dimensional datasets, because they perform both variable selection and shrinkage in an automated fashion and are robust to outliers. Third, BRTs are not "black boxes" as many others machine learning methods are, and are instead known for their interpretability. However, we show that although slightly worse, our results are similar when we use other machine learning methods, such as lasso, elastic net, random forests, and neural networks with 1 through 5 hidden layers.

We start by showing BRTs deliver outstanding performance in predicting high-performing mutual funds. In our baseline results, we show a long-short portfolio that goes long in the top 10% of funds with the highest predicted performance and short in the 10% of funds with the lowest predicted performance deliver an annual excess return of 6.68%, statistically significant at the 1% level.[2] The risk-adjusted returns of such a portfolio are even higher, at 7.46% and statistically significant at the 1% level, indicating the long-short portfolio generated is not exposed to commonly known sources of risk. Furthermore, we show the results delivered by BRTs are superior to those of standard linear regression models and are superior to the best ex-post univariate sort results.

Our baseline results are robust to alternative specifications that use either net or gross mutual fund returns, as well as equal-weighted and value-weighted portfolios. Our results are also similar when we don't use fund characteristics but use their cross-sectional rank as regressors—either normalized between 0 and 1 or grouped into deciles.

Using Fama-MacBeth regressions, we show our BRT-generated forecasts encompass the ones generated by the predictors of mutual fund performance that have been proposed in the literature so far, such as fund size, expense, past returns, tracking error, and active share (Mamaysky, Spiegel, and Zhang, 2007; Bollen and Busse, 2005; Amihud and Goyenko, 2013; Cremers and Petajisto, 2009; Kacperczyk, Sialm, and Zheng, 2005, 2008). We also show our predictability is present in both recessions and expansions and is evenly distributed across all months.

In the second part of the paper, we focus on providing an economic interpretation of our results. First, we show the funds with the highest and lowest expected returns tend to be younger, have less assets under management, and hold fewer stocks. Funds with the highest and lowest expected returns also tend to have higher expenses and turnover and more concentrated portfolios. Second, we show the BRT procedure does not systematically select specific fund classes, even though it tends to select "Small Cap" funds slightly more often than other fund types.

Next, we analyze which characteristics are most useful in selecting mutual funds. Among the top 10 most important characteristics, seven are related to trading frictions, including dollar trading volume, idiosyncratic volatility, size, bid-ask spread, companies' beta, and share turnover. The remaining three characteristics: 12-month, 6-month, and 36-month momentum are instead momentum related. Overall, however, all regressors contribute to the final model and are important for the performance of BRTs.

The BRT framework also allows us to estimate non-parametrically the relation between each characteristic and fund returns. By focusing on the most important predictors, we show the relation between each characteristic and fund returns is often non-linear and monotonic. We also show it is not stable over time. The complex relation between fund returns and fund characteristics is the main reason machine learning methods such as BRTs outperform traditional statistical methods in our setting.

The predicted returns generated by BRTs can be thought of as dynamic extensions of the

[2]Note that, although mutual funds are generally hard to short, our long-short portfolio results are important to gauge the economic value of our forecasts. Note also that throughout the paper we stress the importance of the long leg of the strategy.

DGTW benchmarks proposed by Daniel et al. (1997). DGTW benchmarks control statically for value, size, and momentum, whereas our predicted returns effectively construct benchmark returns for each fund and each period that are based only on the funds' exposure to 94 characteristics. Also, the characteristics included in BRTs vary over time, and the functional form relating each characteristic and predicted fund returns is not necessarily linear. According to this interpretation, we can measure managers' skills by testing whether they systematically outperform their predicted returns. Those managers who systematically outperform their predicted returns are the ones whose outperformance is generated by exposure to factors other than the 94 we consider in our analysis. When we perform this exercise, we find that no manager in our sample displays any degree of outperformance, while 0.4% of the managers display significant under-performance with respect to the predicted returns. We interpret this result as an indication managers performance is driven by exposure to factors, rather than stock selection skills.

In the last part of the paper, we perform a battery of robustness tests regarding the implementation of BRTs. First, we show that our findings are similar when we use other machine learning methods such as lasso, elastic net, random forests, and neural networks with one through five hidden layers. Second, our BRT-generated forecasts may be successful because they select funds with certain styles when such styles are profitable. We therefore re-estimate our results conditioning on fund style and confirm our baseline findings. A third important aspect relates to how information becomes available to real-time investors and how much information to include in our BRT estimates. We show our results are robust to delaying the availability of holdings information up to one year, but that highest performance is achieved when we use the delayed up to three months—or one quarter. A fourth aspect relates to how much information is included in our BRT estimates. Our results show short rolling windows lead to the best out-of-sample performance, suggesting the relation between fund exposure to characteristics and fund performance is transient and time varying. Finally, we show our results are robust if we implement our framework at the quarterly frequency, rather than the monthly frequency, if we vary the number of boosting iterations—one of the key hyper-parameters of BRTs—and whether we vary the number of portfolios we form in the construction of our investment strategies.

An important question is whether the predictable outperformance of certain managers and the underperformance of others is compensation for risk or the result of mispricing. On this debate, the literature seems to be converging on the idea that the majority of the characteristics we analyze in this paper reflect mispricing rather than compensation for risk, because the performance of many of the characteristics we analyze decline over time (McLean and Pontiff, 2016; Green, Hand, and Zhang, 2017). Our results seem to be consistent with this view, because they show fund exposure to certain characteristics can predict their performance, but the characteristics that explain the cross section of fund returns change over time and are present both in periods of low and high risk—economic expansions and recessions.

1 Related Literature

Our paper contributes to four different strands of the mutual fund literature. First, it contributes to the literature that explores whether it is possible to identify high-performing mutual funds using fund characteristics, such as mutual fund size (Berk and Green, 2004; Chen, Hong, Huang, and Kubik, 2004), fund flows (Zheng, 1999; Lou, 2012), fund fees and costs (Elton, Gruber, Das, and Hlavka, 1993; Pástor, Stambaugh, and Taylor, 2017), and past performance (Barras, Scaillet, and Wermers, 2010; Harvey and Liu, 2018; and Groenborg et al., 2021) and what forces limit mutual fund managers from generating consistent outperformance (Pástor, Stambaugh, and Taylor, 2015; Berk and Van Binsbergen, 2015; Hoberg, Kumar, and Prabhala, 2018b). Extant literature has also established that some level of predictability can be detected using past returns, mainly using univariate sorts, such as past-one-year returns, past alphas, and R-squares—see, for example, Hendricks, Patel, and Zeckhauser (1993); Carhart (1997); Mamaysky, Spiegel, and Zhang (2007); Busse and Irvine (2006); Amihud and Goyenko (2013)—or (and) funds holding characteristics, such as tracking error (Cremers and Petajisto, 2009), active share (Cremers and Petajisto, 2009), return gap (Kacperczyk, Sialm, and Zheng, 2008), industry concentration (Kacperczyk, Sialm, and Zheng, 2005), risk shifting (Huang, Sialm, and Zhang, 2011), and active weights (Doshi, Elkamhi, and Simutin, 2015), among others. Other papers have used the characteristics of underlying stocks with some success, exploiting the underlying stocks' momentum (Grinblatt, Titman, and Wermers, 1995), size and book-to-market ratios (Chan, Chen, and Lakonishok, 2002), analysts' recommendations (Kacperczyk and Seru, 2007), accruals (Ali, Chen, Yao, and Yu, 2008), alpha (Elton, Gruber, and Blake, 2011), tangibility (Gupta-Mukherjee, 2014), accruals quality (Nallareddy and Ogneva, 2017), abnormal returns after earnings announcements (Jiang and Zheng, 2018), and mispricing factors (Avramov, Cheng, and Hameed, 2019). Our paper is closest to this latter group of papers in that it selects mutual funds using the characteristics of the stocks they hold. It adds to this literature by being comprehensive and including a large number of characteristics (94) at the same time. It also sheds light on which stock characteristics matter the most for mutual fund performance prediction. We also find that accounting for nonlinearities in fund-level characteristics is important in the selection of mutual funds.

Our paper also contributes to the literature that studies time variation in mutual fund returns. Kacperczyk, Nieuwerburgh, and Veldkamp (2014) find fund performance changes over the course of the business cycle and certain fund managers perform significantly better in recessions than in expansions. Also, Jones and Mo (2020) find instead that many predictors of mutual fund performance perform poorly out of sample, mainly because of changes in the level of arbitrage activity in the market. We contribute to this literature by showing funds' exposure to stock characteristics explains their performance, and hence link time-varying factor/characteristics returns to the time-varying performance of mutual funds. We also show that when we work with mutual fund holdings characteristics, the predictability of mutual fund performance is rather stable.

Third, our paper contributes to the emerging machine learning literature in finance. Various firm characteristics and methods have been proposed to explain the cross section of stock returns—see Green, Hand, and Zhang (2017) and Hou, Xue, and Zhang (2020)—in some cases applying machine learning tools—see Feng, Giglio, and Xiu (2020), Freyberger, Neuhierl, and Weber (2020), Gu, Kelly, and Xiu (2020a), Gu, Kelly, and Xiu (2020b) and Chinco, Neuhierl, and Weber (2020). In the context of investment management Wu, Chen, Yang, and Tindall (2021) and DeMiguel, Gil-Bazo, Nogales, and Santos (2021) construct portfolios of hedge funds and mutual funds, respectively, exploiting fund characteristics and machine learning methods. We contribute to this literature by using machine learning methods to predict the cross section of mutual fund returns on the basis of the stocks they hold. Intuitively, because mutual fund managers aggregate stocks into portfolios, they inevitably expose their portfolios to the factors/characteristics of their individual holdings. We therefore show that constructing mutual funds' characteristics from the stocks they hold can help explain the cross section of mutual fund returns.

Fourth, our paper contributes to the literature that studies the relation between institutional investors and stock market anomalies. Lewellen (2011) finds that, on aggregate, institutional investors show little evidence to trade on any of the main characteristics known to predict stock returns, such as book-to-market, momentum, or accruals. Akbas, Armstrong, Sorescu, and Subrahmanyam (2015) find mutual fund flows exacerbate cross-sectional mispricing, whereas hedge fund flows attenuate aggregate mispricing. Edelen, Ince, and Kadlec (2016) examine institutional demand prior to well-known stock return anomalies and find institutions have a strong tendency to buy stocks classified as overvalued. Calluzzo, Moneta, and Topaloglu (2019) show that anomaly-based trading increases when information about the anomalies is available through academic publication and the release of necessary accounting data. We contribute to this literature by showing that there is a large cross-sectional heterogeneity in funds' exposure to various stock characteristics. Second, we show mutual funds' performance can be explained by the characteristics of the stocks they hold. Third, even though the relation between fund returns and characteristics is complex and often non-linear, it is stable enough to allow for the selection of high-performing mutual funds in real time. As a result, our approach has the potential to be widely applicable by institutional investors who invest in funds rather than individual securities, such as fund of funds (FOF).

2 Data and Sample Selection

We use data from a variety of publicly available sources. The first is the CRSP Survivorship-Bias-Free Mutual fund database (MFDB). This database contains monthly returns and characteristics, such as expenses, total net assets, and portfolio turnover, among others, for over 64,000 open-ended funds.

Following the literature, we select equity mutual funds using the Lipper, Strategic Insight, and Wiesenberger classifications and obtain virtually identical samples when we select equity mutual funds using the recently introduced CRSP Style Code.[3] We exclude index funds and target-date

[3]US equity funds are those with policy code "CS", Lipper codes "EIEI, G, LCCE, LCGE, LCVE, MCCE, MCGE,

funds using the index fund flag in MFDB and string searches on fund names. We exclude index funds by eliminating funds whose names contain the strings "exchange-traded," "ETF," "Index," "indx," and so on, and eliminate target-date funds by removing funds whose names contain the strings "target" and specific years, namely 2005, 2010, 2015, and so on. Following Amihud and Goyenko (2013), we also remove funds with (lagged) total assets below \$15 millions and we require a minimum of 80% and a maximum of 105% holdings in common stocks. To address the incubation bias (Evans, 2010), we eliminate observations before the funds started reporting to CRSP. Finally, we eliminate funds with missing names.

Our second source of data is the Thomson Reuters Mutual Fund Holdings database (TFN/CDA S12), from which we obtain quarterly mutual-fund-holdings information.[4] Because the majority of our results are computed at the monthly frequency, whereas mutual fund holdings are computed at the quarterly frequency, we map quarterly holdings data into monthly data using the latest available holdings data for each month.[5] We do, however, compute our results at the quarterly frequency in Section 8.3 and find the results are robust. We use Mutual Fund Links (MFLINKs) to aggregate all the share classes associated with a given mutual fund and merge MFDB and TFN/CDA's holding datasets.

We obtain mutual fund-level characteristics as the weighted average of the characteristics of the stocks they hold, where the weight is the relative weight of each stock in the mutual fund portfolio. In an effort to be comprehensive in the initial number of stock characteristics, we follow Green, Hand, and Zhang (2017) and Gu, Kelly, and Xiu (2020b) and build a cross section of 94 firm-level characteristics, constructed using information from CRSP, Compustat, and IBES, see Table A.1 for details. Although we refer the reader to Green, Hand, and Zhang (2017) for details, we report below some of the instances where we depart from Green, Hand, and Zhang (2017) in our data construction. In particular, we include stocks with prices below \$5 and we include financial firms. We also include stocks with share codes other than 10 and 11 and exchange codes other than 1, 2, and 3 to maximize the number of stocks we can match to the mutual-fund-holdings data. Note that when we impose the same restrictions as in Green, Hand, and Zhang (2017), we obtain very similar results.

Different firm characteristics are updated at different frequencies. Of the 94 characteristics, 61 are updated annually, 13 are updated quarterly, and 20 are updated monthly. Because our baseline results are computed at the monthly frequency, we convert all these characteristics to the monthly frequency using their latest available value. Furthermore, to avoid the forward-looking biases, we follow Gu, Kelly, and Xiu (2020b) and use the following convention. When computing results for month $t+1$, we use monthly characteristics measured as of month t, quarterly characteristics

MCVE, MLCE, MLGE, MLVE, SCCE, SCGE, SCVE, CA, EI, G, GI, MC, MR, SG," Strategic Insight codes "AGG, GMC, GRI, GRO, ING, SCG," and Wiesenberger codes "G, GCI, IEQ, LTG, MCG, SCG."

[4]The primary source for the mutual-fund-holdings data are SEC N-30D filings. CRSP MFDB also contains quarterly mutual-fund-holdings information, but the data only start from 2001. Because of this limitation, we use the Thomson Reuters Mutual Fund Holdings database.

[5]During the early years, funds were required to disclose holdings only semi-annually. We thus fill the quarterly observations using the latest available holdings.

measured by the end of month $t-4$, and annual characteristics measured by the end of month $t-6$. We winsorize all characteristics around the 1st and 99th percentiles.

In the final step, we merge firm characteristics and mutual-fund-holdings data and obtain fund-level characteristics as the weighted average of the characteristics of the stocks mutual funds hold in their portfolio. A very small fraction of stocks in the holdings dataset cannot be matched with the characteristics dataset. Our final dataset consists of 2,980 unique mutual funds and 407,050 fund-month observations, spanning from January 1980 to December 2018, and the average number of months in our sample is 136.

3 Do Mutual Fund Managers Trade on Anomalies?

Before proposing our approach to select mutual funds, we start with two preliminary tests. In the first, we assess whether—on average—mutual fund managers trade on the stock characteristics—and associated anomalies—we exploit in this paper. Second, we test whether simple univariate sorts on the portfolio characteristics of mutual fund managers allow us to identify and select in real time mutual fund managers with superior performance.

The first test is meant to test whether mutual fund managers trade on stock characteristics and whether selecting successful managers based on the characteristics of the stocks they own is possible. The second is meant to test whether the relation between portfolio characteristics and mutual fund performance is simple or is characterized by interactions and non-linearities that warrant the use of powerful machine learning algorithms in order for it to be uncovered.

3.1 Anomalies and Mutual Fund Trading

The finance literature has identified several stock market anomalies, whereby portfolios sorted on certain stock characteristics deliver abnormally high portfolio returns. On the other hand, whether sophisticated investors profit from these anomalies has not been studied extensively, with the exception of a handful of papers studying whether mutual fund managers trade on momentum, see Grinblatt, Titman, and Wermers (1995) or accrual anomalies, see Ali et al. (2008). In fact, the majority of the literature explicitly constructs fund managers' skill measures *controlling* for their stock holdings information, see for example, Daniel et al. (1997). In this section, we conduct a comprehensive analysis of the mutual funds in our sample and study whether—as a group—they trade on any of the 94 anomalies we have in our sample.

We follow Ali et al. (2008) and construct the anomalies investing measure (AIM) for each of the 94 anomalies, where AIM is calculated as the weighted average of firm characteristics' decile ranks of individual stocks in the CRSP universe. For each fund i, characteristic k, and month t, we compute the AIM measure as follows:

$$AIM_{i,k,t} = \sum_{j=1}^{N} w_{i,j,t} * Rank(Firm\ Characteristics_{j,k,t}),$$

where $AIM_{i,k,t}$ is the AIM for fund i and characteristic k at the end of month t, the decile ranking of stock j based on characteristic k in month t is $Rank(Firm\ Characteristics_{j,k,t})$, N is the number of stocks held by mutual fund i at the end of month t, and $w_{i,j,t}$ is the weight of stock j held by mutual fund i at the end of month t. A high AIM value indicates the fund primarily holds stocks with high values of characteristic k. If mutual funds are not exposed to characteristic k, the average AIM computed across all funds would be around $(1 + 10)/2 = 5.5$.

In the first five columns of Panel A of Table 1 we report the mean, standard deviation and 25th, 50th, and 75th percentiles of the AIM measures across all funds and time periods. In the remaining two columns, we report the percentage of funds that are significantly negatively and positively exposed to each characteristic. Given the persistence in exposure characteristic by each fund, we compute Newey-West standard errors with 12 lags.

The 94 characteristics have been ranked according to mean exposure. At the very top, we have characteristics with cross-sectional averages above 6.5. Among these characteristics, we find *dolvol*—trading volume—and *mve_m*—size. The average AIM for these characteristics is very high at 9.19 and 9.18, respectively, confirming the stylized fact that mutual funds tend to hold and trade stocks of large and very liquid companies. We find similar values for other characteristics related to firm size, such as firm age, *age*, and industry-adjusted size—*mve_ia*. Other significant characteristics are the ones related to companies' profits and returns on investments—*roaq*, *roic* and *roeq*, and *operprof*—that capture returns on equities, returns on assets and operating profits. Finally, active mutual funds also tend to purchase companies with high *cashdebt*, cash-over-debt ratios, and *cashpr*, cash productivity, and companies with high realized returns over the past 12 and 36 months—high *mom36m* and *mom12m*. Within this category, we find the majority of funds are individually exposed to these characteristics as well. For example, 97.5% of the funds are positively exposed to the size characteristic while only 0.35% of funds are negatively exposed to it.

A second group of characteristics have average AIM values between 6.5 and 4.5. These variables include characteristics such as earnings-price ratios, dividend-yield variables, employee growth rate, and industry-sale concentration, among others. Although, on average, the mutual fund industry is not exposed to any of these variables, a significant number of mutual funds are exposed positively and negatively to them. For example, 61.5% of the funds buy stocks that have higher-than-average earnings-price ratios, whereas 2.85% hold stocks with lower-than-average earnings-price ratios. The results for cash holdings, *cash*, are more balanced. The average AIM—first column—is exactly equal to 5.5; 31.42% of the funds have negative exposure to *cash* whereas 27.99% have positive exposure to it. Finally, funds tend to prefer stocks with low firm tangibility—*tang*. The average AIM is 4.58, 71.90% of the funds are negatively exposed to this characteristic and only 2.54% of the funds are positively exposed to it.

The final set of characteristics have average AIM below 4.5, indicating mutual funds tend to shy away from stocks with these characteristics. Some of these characteristics proxy for illiquidity, such as *ill* and *zerotrade*, and others proxy for volatility, such as *idiovol* and *roavol*. Finally, funds prefer growth stocks to value stocks. The average AIM measure equals 4.07, with 68.35% of the funds holding growth stocks and only 4.22% of the funds holding value stocks.

Panel B of Table 1 reports the number of characteristics each fund is significantly exposed to. The majority of funds, 57.42%, are negatively exposed to between 20 and 30 characteristics over our sample. Among the remaining funds, 37.90% are negatively exposed to between 10 and 20 characteristics, 0.98% are negatively exposed to less than 10 characteristics, and 3.71% are exposed to between 30 and 40 characteristics.

Positive exposure is instead more widespread: 34.83% of the funds are positively exposed to between 20 and 30 characteristics, 33.92% of the funds are exposed to between 30 and 40 characteristics, and 22.52% of the funds are exposed to between 40 and 50 characteristics.

The results reported in this section suggest mutual funds are significantly exposed to a large number of the characteristics associated with abnormal stock returns. The fact that, as a group, mutual fund managers trade on certain characteristics, however, does not imply they successfully achieve abnormal returns doing so. It also does not imply selecting high-return mutual funds on the basis of the stocks they hold is possible, because the relation between fund characteristics and returns may be highly non-linear. In the next section, we test whether it is possible to select funds with high returns based on the stock characteristics they hold using simple univariate sorts.

3.2 Univariate Sorts

In this section, we construct mutual fund portfolio characteristics by computing the weighted average of the characteristics of the stocks they own. We then sort funds every month on the basis of each characteristic. For month t, we sort all mutual funds on the basis of each characteristic, $x_{t-1,i}$, out of the 94 we have available, and form decile portfolios. The high (low) decile portfolio consists of the mutual funds with the highest (lowest) values of $x_{t-1,i}$. We then construct a long-short portfolio that longs the high-decile portfolio and shorts the low-decile portfolio and hold it for one month.

Note this exercise is likely to return long-short portfolios with positive abnormal returns if funds trade on characteristics and if a direct and simple relation exists between funds' holdings of certain characteristics and their risk-adjusted performance.

We report equal- and value-weighted portfolio excess returns and risk-adjusted performance for the long-short portfolio associated with each characteristic. The risk-adjusted performance of each portfolio is computed as:

$$R_{i,t} = \alpha_i + b_i\ R_{m,t} + s_i\ SMB_t + h_i\ HML_t + m_i\ MOM_t + \epsilon_{i,t},$$

where $R_{i,t}$ is portfolio i's excess return in period t in excess of the risk-free rate, $R_{m,t}$ is the market excess return, SMB_t, HML_t, and MOM_t are the factor portfolios' returns related to size, book-to-market, and momentum, respectively. Each portfolio's risk-adjusted performance is measured by α_i.

Table 2 reports results for each of the 94 characteristics, sorted by the significance of the equal-weighted excess returns, where t-statistics are computed using Newey and West (1987) standard errors.

Few characteristics deliver statistically significant excess and risk-adjusted returns. For equal-weighted returns, we find that a strategy that goes long (short) in funds that hold stocks with high (low) sales-to-receivables, *salesrec*, deliver an annual excess return of 2.20, with a t-statistic of 2.63. Excess returns of this long-short strategy are also significant for the percentage change of sales-to-inventory, *pchsaleinv*, R&D to market capitalization, *rd_mve*, industry-adjusted change in asset turnover, *chatoia*, earnings-announcement-returns, *ear*, and sales to inventory, *saleinv*. All the long-short strategies based on the remaining 88 characteristics do not deliver significant economically or statistically significant outperformance, and the results are robust when we use FFC alphas as measures of risk-adjusted performance, and if we use value-weighted returns.

These results imply that, although selecting mutual funds on the basis of the stocks they hold may be possible, the relation between fund performance and the various characteristics may be complex and non-linear. This is likely the reason why univariate sorts deliver underwhelming results.

4 Selecting Mutual Funds using a Machine Learning Approach

As shown in Section 3, the vast majority of mutual funds are exposed to various stock characteristics, with the majority of mutual funds being negatively exposed to 20-30 characteristics and positively exposed to 20-30 characteristics. Meanwhile, it is not possible to identify successful mutual funds using univariate sorts, the main reason being that mutual funds are exposed to so many possibly opposite characteristics at the same time, and there may be non-linearities between stock-characteristics exposure and mutual fund performance.

To overcome these limitations, we model mutual funds' excess returns using their lagged fund-level characteristics, constructed using the characteristics of the stocks they hold. We then model the non-linearities between the fund excess returns and fund characteristics as well as the interactions between fund characteristics using machine learning tools.

More formally, we cast our problem as a supervised regression task,

$$r_{i,t+1} = f\left(\mathbf{z}_{i,t}|\theta\right) + \epsilon_{i,t+1},$$

where $r_{i,t+1}$ denotes mutual funds i's excess return for month $t+1$, $\mathbf{z}_{i,t}$ denotes a vector of fund i's fund-level aggregated characteristics for month t, and $f(\cdot)$ is a flexible function that maps the fund-level characteristics to funds' excess return. For simplicity, in our baseline specifications we do not allow our functional form $f\left(\cdot\right)$ to vary across i and t, implicitly assuming a stable relation between returns and characteristics across funds and time. As we detail in the implementation section, we do however standardize both the covariates and the returns cross-sectionally, to account for the time-variations in realized returns and other stock characteristics. Finally, $f\left(\cdot\right)$ depends on $\mathbf{z}_{i,t}$, meaning our prediction only relies on fund i's information at period t, and does not consider other funds' information or the history before t.

We face several challenges in estimating $f(\cdot)$. First, the relation between fund-level characteristics and fund returns are complex and possibly nonlinear (Gu, Kelly, and Xiu, 2020b; Bianchi, Büchner, and Tamoni, 2020). Traditional parametric methods impose linearities and tend to overfit the training sample if we were to impose non-linearities parametrically, given that we are working with up to 94 characteristics per fund. On the other hand, traditional non-parametric methods face the so-called "curse of dimensionality," if we were to include more than two or three co-variates at the same time. To handle these challenges, we resort to Boosted Regression Trees (BRTs).

4.1 Boosted Regression Trees

While there exists a wide range of machine learning methods aimed at addressing the challenges reported above, the "no free lunch theorem" shows no machine learning algorithm can be expected ex ante to outperform others on any given task (Wolpert, 1996). We choose BRTs, rather than other nonlinear methods such as neural networks for three main reasons. First, BRTs have exhibited strong predictive performance in various fields. For example, BRTs routinely place at the very top in many Kaggle machine learning competitions (see Machine Learning Challenge Results for examples of competition results). In financial settings, extensive horse races show tree-based methods perform as well as neural networks and outperform other linear and non-linear methods (Gu, Kelly, and Xiu, 2020b; Bianchi, Büchner, and Tamoni, 2020). Second, BRTs can handle large, high-dimensional datasets because they perform both variable selection and shrinkage in an automated fashion. They are also robust to outliers and can handle missing values. Third, although most machine learning methods, including neural networks, focus only on predictive performance and are criticized as "black boxes," one advantage of BRTs is their good interpretability inherited from regression trees (Hastie, Tibshirani, and Friedman, 2009).[6] For example, we can estimate which covariates matter, among the many available, using *relative-influence measures* and obtain non-parametric estimates of the relation between funds' expected returns and each of their characteristics using *partial dependence plots*. We present below a more formal treatment of BRTs. Section 4.2 describes regression trees, Section 4.3 describes boosting. Finally, Section 4.5 describes the implementation of BRT adopted in the paper.[7]

4.2 Regression Trees

Suppose we have P potential predictor ("state") variables and a single dependent variable over T observations, i.e., (x_t, y_{t+1}) for $t = 1, 2, ..., T$, with $x_t = (x_{t1}, x_{t2}, ..., x_{tp})$. 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

[6]See Du, Liu, and Hu (2019) for some explainable machine learning techniques.

[7]Our description draws on Friedman (2001), who provides a more in-depth coverage of the approach.

regression tree can be written as an additive model:

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

where $S_j,\ j = 1, ..., J$ 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 $\hat{c}_j = mean(y_{t+1}|x_t \in S_j)$.

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,

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

that minimize the sum of squared residuals:

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

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

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

The best splitting pair (p, s) in the first iteration can be determined by searching through each of the predictor variables, $p = 1, .., P$. 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 generally employed in high-dimensional datasets where the relation between predictor and predicted variables is potentially non-linear. This feature is important in our context, because which variables may be more or less relevant ex-ante is unclear. Furthermore, it is difficult to know in our context whether there is a linear relation between predictor and predicted variables. On the other hand, 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 method known as boosting.

4.3 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:

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

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

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

where $\mathcal{T}_b\left(x; \{S_{b,j}, c_{b,j}\}_{j=1}^J\right)$ 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

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

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

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

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 jth state.

Forecasts are simple to generate from this approach. The boosted regression tree is first estimated using data from $t = 1, ..., t^*$. 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.

4.4 Relative Influence Measures and Partial Dependence Plots

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 is hardly applicable to BRTs that instead feature very useful and intuitive visualization tools.

4.4.1 Relative Influence Measures. The first measure commonly used is generally referred to as "relative influence" measures. Consider the reduction in the empirical error every time one of the covariates $x_{l,\cdot}$, 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 $\Delta e(j)^2 = T^{-1} \sum_{t=1}^{T} \left(e_t(j-1)^2 - e_t(j)^2 \right)$ 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, $t = 1, ..., T$, 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, $x_{l,\cdot}$, 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:

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

4.4.2 Partial Dependence Plots. The second visualization tool featured by BRTs is *partial dependence plots*, which are defined as follows. Suppose we select a particular covariate, X_p, from the set of P predictor variables $X = (X_1, X_2, ..., X_P)$ and denote the remaining variables X_{-p}, i.e. $X_{-p} = X \backslash \{X_p\}$. We use the following measure of the average marginal effect of X_p on the dependent variable:

$$f_p(X_p) = E_{X_{-p}} f(X_p, X_{-p}).$$

This quantity is called the average partial dependence measure. It fixes the value of X_p and averages out the effect of all other variables. By repeating this process for different values of X_p, we trace out the marginal effect of this covariate on the predicted variable.

An estimate of $f_p(X_p)$ can be computed by averaging over the sample observations:

$$\bar{f}_p(X_p) = \frac{1}{T} \sum_{t=1}^{T} f(X_p, x_{t,-p}),$$

where $x_{t,-p} = \{x_{1,-p}, ..., x_{T,-p}\}$ are the values of X_{-p} occurring in the data.

4.5 Implementation

Now we describe the implementation of the BRT model for selecting mutual funds.[8] Each month, we train a BRT model and use the information available as of month t to forecast the returns of each mutual fund in our sample on month $t + 1$. We then form 10 portfolios based on the funds' expected returns. Finally, we report the performance of each portfolio and provide asset pricing tests on long-short portfolios that purchase funds with the highest expected returns and sell funds with the lowest expected returns.

In particular, at the end of month t, we train a BRT model using lagged fund-level characteristics and excess return data. Each training sample $(\mathbf{z}_{i,t-1}, r_{i,t})$ is fund i's holding characteristics at the end of month $t-1$ and fund i's excess return for month t. Our baseline specification uses a one-month rolling window, but we also explore longer rolling windows and recursive window specifications.[9]

We adopt two refinements to the basic BRT methodology. The first is *shrinkage*, a simple regularization technique that diminishes the risk of over-fitting and improves the out-of-sample prediction by slowing the rate at which the empirical risk is minimized on the training sample. The shrinkage parameter, $0 < \lambda < 1$, determines how much each boosting iteration contributes to the overall fit:

$$f_b(x) = f_{b-1}(x) + \lambda \sum_{j=1}^{J} c_{j,b} I\{x \in S_{j,b}\}.$$

Following common practice, we set λ to the default value of 0.1.

The second refinement is *subsampling* and is inspired by "bootstrap aggregation" (bagging) (Breiman, 1996). Bagging is a technique that computes forecasts over bootstrap samples of the

[8]The main results are computed using the package LightGBM in R. Results are virtually identical if we use GBM and are comparable, but slightly worse, when we use XGBoost. This result is in line with the findings in Ke, Meng, Finley, Wang, Chen, Ma, Ye, and Liu (2017), who find LightGBM outperforms XGBoost and other boosting implementations in various public datasets.

[9]When we work with longer rolling windows and recursive windows, we standardize both regressors and returns to account for the time variation in their respective distribution.

data and averages them in a second step, therefore reducing the variance of the final predictions. In our context, the procedure is adapted as follows: at each boosting iteration we sample without replacement one half of the training sample and fit the next tree on the subsample obtained.

Finally, we set the number of boosting iterations to 100, but show different choices for this hyper-parameter do not have a significant impact on the results.

5 Selecting Mutual Funds with BRTs

We start by presenting the baseline results of applying BRT to select mutual funds on the basis of the stocks they hold. We then turn to alternative specifications to show our BRT results are robust to alternative approaches.

5.1 Baseline Results

The first four columns of Table 3 report the results for BRT using monthly observations from 1980 to 2018. Each month, our BRT procedure predicts the returns on each fund. We then sort the funds on the basis of their expected returns and form 10 equal-weighted portfolios. The first portfolio contains mutual funds with the lowest expected returns, while the 10th decile portfolio contains funds with the highest expected returns. Finally, we compute the next-month realized returns for each portfolio.

The results indicate a clear monotonic pattern, with the exception of the second-decile portfolio. The first-decile portfolio has an average realized annualized excess return of 4.23%, the fifth portfolio has a return of 7.03%, and the 10th portfolio has a return of 10.91%. As a result, a portfolio that goes long on the 10th portfolio and short the first portfolio has an excess return of 6.68% per year, with a t-statistic of 3.56. We find a monotonic pattern also in terms of Fama-French Carhart four-factor alpha, reported in columns 3 and 4. The alpha for the first portfolio is negative at -4.59% and statistically significant with a t-statistic of -5.09. The alpha associated with the fifth portfolio is slightly negative at -1.10%, with a t-statistic of -2.36. Finally, the top decile portfolio has a positive alpha with a t-statistic of 2.16. These results indicate that the outperformance is not the result of the decile portfolios being exposed to the sources of risk proxied by the Fama-French Carhart factors. Finally, the monotonic increase in performance is such that a portfolio that is long in the top-decile portfolio and short in the bottom-decile portfolio has a Fama-French Carhart four-factor alpha of 7.46% with an associated t-statistic of 3.76.

Columns 5 through 8 of the table report instead the results for a linear regression model that uses the same covariates as the BRT model. The results indicate that, indeed, there is some degree of predictability for the linear model but the evidence, overall, is rather weak. In particular, the first-decile portfolio has an excess return of 6.03%, while the 10th portfolio has an excess return of 9.78%. A long-short portfolio has an excess return of 3.75%, but it is not statistically different from zero: the t-statistic only equals 1.48. The alpha estimates associated with each portfolio are mostly negative and significant for the first three decile portfolios, but are not significant for the top-decile portfolio. Finally, the alpha estimate associated with the Fama-French Carhart four-factor alpha is 4.93 that is barely significant at the 10% level—the t-statistic is 1.96.

Finally, as a reality check, we also report—in columns 9 through 12—the performance for "sales to receivables" (*salesrec*), the best ex-post univariate regressor as reported in the first column of Table 2. The variable would not have been available ex ante to a real-time investor, given it has been selected ex-post. At the same time, it is an important quantity because it represents the best someone could have done when using univariate sorts. The results are much weaker than in the BRT model as the first-decile portfolio has an average return of 6.15%, whereas the 10th-decile portfolio has a return of 8.35%. The associated long-short portfolio has an excess return of only 2.20%, which is nevertheless statistically significant. The risk-adjusted performance of the various decile portfolios is certainly non-monotonic, with portfolios 1, 2, 3, and 5 being negative and significant and none of the portfolios being positive and significant. Finally, a portfolio that goes short the first decile and long the tenth 10th has an economically small risk-adjusted alpha of 1.85% that is nevertheless statistically significant at the 1% level—the t-statistic is 2.65.

5.2 Additional Specifications

In this section, we dig deeper and provide results based on alternative specifications, compared with the ones reported in Section 5.1. In the first column of Table 4, we use the same BRT specification as in Table 3. In the second column we repeat the analysis, but we form value-weighted rather than equal-weighted decile portfolios, the main difference being that more wealth is placed in larger funds compared with smaller ones. In the third and fourth columns, we instead repeat the analysis in the first two columns but use gross returns—returns before fees—rather than net returns. In all cases, we find the results are virtually identical to the baseline ones. In all cases, the long-short portfolios have statistically significant excess returns of approximately 6.5% that are also statistically different from zero. The estimated alphas are even larger, at around 7.5%, and statistically different from zero.

At the bottom of each column, we present the results of monotonicity tests—see Patton and Timmermann (2010). The returns associated with the 10 sorted portfolios are non-monotonic under to the null hypothesis. They are instead monotonic under the alternative. The estimated t-statistics are all approximately equal to 3, indicating average returns are monotonically increasing among the 10 portfolios.

The fact that equal- and value-weighted results are equally strong indicates the procedure is not systematically selecting small or large funds as high or low returns. The fact that the results for gross returns are similar to those for net returns indicates instead the procedure is not systematically selecting funds with high or low fees.

Columns 5 through column 8 of Table 4 repeat the analysis but change the regressors. Rather than focusing on the raw regressor values, we rank the stocks in the CRSP dataset according to each of the 94 characteristics, and we standardize each rank so that it ranges between 0 and 1. We then construct fund-level characteristic using firm-characteristics ranks. The idea of this approach is to reduce the impact of outliers in the stock characteristics and to standardize characteristics cross-sectionally. The results show this alternative approach produces virtually identical results. The long-short portfolios have annualized excess returns of 6.60%, and the Fama-French Carhart alphas have alphas of 7.42%. These quantities are all statistically different from zero. The monotonicity tests confirm a positive monotonic relation across the 10 portfolios for all specifications.

The last four columns of Table 4 are similar to the rank results but focus on deciles over the CRSP universe. Each month, we categorize stocks into deciles on the basis of the 94 characteristics. We then construct fund-level characteristics using stocks' deciles. Once again, the results are virtually identical to the baseline ones. The monotonicity tests confirm a positive monotonic relation across the 10 portfolios for all specifications.

5.3 Fama-MacBeth Regressions Controlling for other Mutual Fund Predictors

The results reported so far show BRT forecasts contain economically valuable information because they allow us to distinguish between high- and low-performing mutual funds in real time. However, the literature has identified other predictors of fund performance such as fund size, expense, past returns, tracking error, and active share—see Mamaysky, Spiegel, and Zhang (2007), Amihud and Goyenko (2013), Cremers and Petajisto (2009), Kacperczyk, Sialm, and Zheng (2005), and Kacperczyk, Sialm, and Zheng (2008) among others. To test if our BRT forecasts encompass—and are not subsumed by— the ones associated with predictors already established in the literature, we implement the following Fama-MacBeth regressions:

$$Return_{f,t} = \beta_1 \widehat{BRT}_{f,t|t-1} + cM_{f,t-1} + \epsilon_{f,t},\ \ for\ t = 1, \ldots, T,$$

where $Return_{f,t}$ is fund f's excess return at month t (expressed in percentages); $\widehat{BRT}_{f,t|t-1}$ is the BRT prediction for fund f's excess return at month t, conditional on information available as of time $t-1$, and $M_{f,t-1}$ is a vector of regressors available as of time $t-1$ that contains the following covariates: Lag(Excess Returns) is the monthly lagged excess return of the fund; TR^2 is logistic transformation of the R^2 associated with a regression of the fund's monthly excess returns on a Carhart 4-factor model using an estimation period of 24 months—see Amihud and Goyenko (2013) for details; Active share is obtained from Antti Petajisto's website[10]; and Fund Flows, log(TNA), $\log(\mathrm{TNA})^2$, log(Fund Age), Expense, and Turnover represent fund flows, logged fund size, logged fund size squared, logged age of the fund, the expense ratio of the fund, and the turnover of the fund. These quantities are obtained from CRSP and calculated following Avramov, Cheng, and Hameed (2019). All regressors are available from 1980 to 2018 except for Active share that is available from 1980 to 2015. The regression results are reported in Table 5.

Consistent with the hypothesis that our predicted returns contain information that is not sub-

[10]The authors post the Active share data on two websites. The data on https://activeshare.nd.edu/data/ starts in 1990 and ends in 2015, and the data on https://www.petajisto.net/data.html spans from 1980 to 2009. We merge the two datasets to obtain the data ranging from 1980 to 2015. Most of the above active share data are quarterly. We convert them to the monthly frequency using the latest available active share observations.

sumed by other predictors reported in the literature, we find the β_1 coefficient is positive, statistically significant, and very stable across specifications (1) through (4). Furthermore, we find other common predictors of fund returns, such as funds' R^2, lagged excess returns, active share, are barely significant, suggesting their information content is dominated by that contained by our BRT forecasts. Among the other covariates, we find the coefficients on logged fund size and expense ratios have a negative and significant coefficient, confirming the findings in Chen et al. (2004), Elton et al. (1993), and Carhart (1997), respectively.

Overall, the results reported in this section show the information contained in the BRT forecasts encompasses and is not subsumed by the information contained in the predictors of mutual fund performance that have been proposed so far in the literature.

5.4 Time-Varying Predictability

The results computed so far average across all years and months in our dataset. Because mutual-fund-holdings information is available quarterly (December, March, June, and September), we could expect time-varying predictability across different months. The reason is that, for example, the December holdings information may be more informative for January returns rather than February and March returns, because the holdings information is likely to be less stale for January, than for February and March. On the other hand, if the information contained in the holdings data is only slowly incorporated into asset prices, we would expect that holdings information would predict mutual fund returns across all months, irrespective of how stale the mutual-fund-holdings information is.

The results for this test are reported in Table 6. We do not find distinct and recognizable patterns in the months that deliver higher long-short portfolios. The month in which the long-short portfolio performs the best is November, followed by February, April, September, May, and August. All the other months display insignificant returns predictability. We interpret the lack of significant patterns of predictability across different months as evidence that the information incorporated into our BRT framework is persistent in nature rather than concentrated in specific periods associated with the release of new holdings information.

Also, mutual funds performance is well-known to vary across different stages of the business cycle (Kacperczyk, Nieuwerburgh, and Veldkamp, 2014). In Table 7, we estimate whether the predictability of mutual fund performance using BRTs is time varying or persistent over time. We use two business-cycle measures. The first is the NBER recession indicator. The second is the Baker and Wurgler (2006) sentiment index. We highlight two facts. First, the long-short portfolio has a higher performance in economic recessions than in expansions. The equal-weighted long-short portfolio is 13.52% in the first case and 6.01% in the second. Statistically, however, the long-short portfolios have returns that are different from zero in both cases. Second, the results for high- and low-sentiment suggest the performance of BRTs is similar across the two states of the economy.

Overall, the results in this section suggest BRTs are not picking up time-varying sources of risk and that the outperformance they deliver is rather stable over time.

5.5 Controlling for Information Arrival

The results reported so far use the latest holdings information to predict fund performance. For example, we predict January fund returns using December fund holdings. This timing convention has the potential to overstate the predictability of BRTs, because the Securities and Exchange Commission (SEC) requires mutual funds to disclose their holdings information within 60 days—or two months—after each quarter ends. Although anecdotal evidence suggests the majority of the funds report their holdings information within 30 days, we cannot make sure that this is the case in all instances. Because of these reporting delays, the literature commonly delays holdings information by three months (Kacperczyk, Sialm, and Zheng, 2008).

Rather than choosing a specific number of months for the reporting delay, we take a more comprehensive approach and delay the fund-level characteristics and the holdings information from 1 to 12 months. If the information contained in our regressors can persistently predict mutual fund returns, we should expect little deterioration in performance as we delay the information arrival from 1 to 12 months. Note that while we are being rather comprehensive, we believe that the most realistic specifications are the ones that delay holdings by a quarter. Under this quarter-delay specification, we predict April's returns based on December's fund-level characteristics.

The results are reported in Panel A of Table 8. The first column reports our baseline results as a benchmark. As we increase the delay, we see the performance of the BRT methods stays rather strong. If anything, we see a slight improvement in performance at the three-month frequency, which has a long-short portfolio with an excess return of 7.19% and an annualized alpha of 7.64%, both significant at the 1% level. As we increase the holdings delay, we find that the performance is still strong but deteriorates slightly. For example, the results in the last column—associated with a 12-month delay—report an excess return of 6.22% and an alpha of 7.03%, both statistically different from zero.

In Panel B of Table 8, we repeat the exercise but delay only portfolio holdings information—and not stock prices and firm characteristics. The results are virtually identical to those in Panel A and indicate delaying holdings information has only a small impact on portfolio performance.

The results reported in this section highlight two facts. First, fund-level characteristics have a persistent effect on mutual funds' performance. Second, BRTs are able to extract the information contained in the predictor variables and deliver good risk-adjusted performance.

5.6 Results for Different Rolling-Window Size

The baseline results use rolling window estimates with training windows of one month. Using such a short rolling window means we are using local information to predict successful funds. If the relation between stock characteristics and stock returns is stable, we would expect that increasing the rolling-window size and eventually using recursive-window specifications should improve the performance of BRTs. On the other hand, if the relation between mutual fund characteristics and performance changes over time, we would expect that lengthening the window should result in a lower performance for BRTs.

When we move away from a one-month rolling window to multiple-month and recursive windows, appropriately standardizing both regressors and regressands is important because excess returns are time varying, whereas mutual fund characteristics are rather persistent. We standardize both regressors and returns in such a way that they have a cross-sectional mean of zero and a cross-sectional standard deviation of one.

The results reported in Table 9 allow the training window to vary from one month—our baseline result—to 6, 12, 36, and 60 months. We also include a recursive specification. Throughout, the results suggest lengthening the training window leads to a deterioration of the BRT performance. The first four columns of Panel A report the results for equal-weighted net returns. The table suggests the BRT performance declines from 6.68% at the one-month horizon to 5.42% when we use a recursive-window specification. In all cases, however, the long-short portfolios are significantly different from zero, and the same is true for the risk-adjusted alphas. Columns 5 through 8 repeat the exercise for value-weighted returns. Even in this case, we observe virtually identical patterns: the BRT performance deteriorates as the window horizon lengthens, but in all cases, we obtain long-short portfolios with positive returns that are significantly different from zero.

Panel B of Table 9 repeats the analysis but focuses on gross returns rather than net returns. Qualitatively, the results are identical to those reported in Panel A.

Overall, the results reported in this section suggest BRTs exploit the local and potentially time-varying relation between fund characteristics and their performance. It is well-known that certain stocks outperform others in certain periods but not in others, because they are exposed to specific risk factors. Because certain mutual funds are more exposed to certain stock characteristics but not others, their performance is time varying. The results reported here suggest BRTs with short rolling windows are capable of capturing this transient outperformance.

6 Which Funds Are Selected by BRT?

In Section 5, we showed our procedure is able to select funds that perform well on the basis of their holdings. We also showed higher returns were not proxying for higher risk-taking, because we identified monotonic patterns for both excess returns and Fama-French Carhart four-factor alphas. In this section, we analyze in more detail the types of funds our procedure selects, by averaging the fund characteristics in each decile portfolio. Following Kacperczyk, Nieuwerburgh, and Veldkamp (2014), we focus on seven characteristics: Age, $Size$, $Number\ of\ Stocks$, $Expense$, $Turnover$, $Concentration\ Index$, and the $Industry\ Concentration\ Index$ (ICI).

The results, reported in Table 10, indicate several non-monotonic patterns. In particular, we find an inverted U-shaped relation for the first three characteristics, in that the funds with the highest (Decile 10, or D10) and lowest predicted returns (Decile 1, or D1) are younger, have less assets under management, and hold fewer stocks than the funds in the middle of the distribution. When we compare these characteristics for the funds in the top- and bottom- decile portfolios (D1 and D10) with the ones in the middle portfolios (D5 and D6), we find extreme deciles contain funds that are 11 months younger (statistically significant with a t-statistic of 6.02), are smaller by \$200 million (t-statistic of 5.4), and hold 17 fewer stocks (t-statistic of -8.2).

These results are consistent with the procedure being able to select smaller funds that are likely to receive fund flows in the future because of their managers' skills and distinguish them from the small funds that instead have no manager skills (Berk and Green, 2004).

The expense-ratio results indicate instead a U-shaped relation (significant with a t-statistic of 14.7), with the least successful and the most successful funds charging higher expense fees, indicating part of the poor performance of some smaller funds is their willingness to charge high fees even though their managers do not have skill.

In line with expense ratios, we find both the least and the most successful funds have higher turnover. In the first case, this excess trading does not lead to high returns, whereas it does in the second case.

Finally, the last quantity we analyze is the concentration index. Here, we adopt two measures. The first is the Herfindahl index of the decile portfolio, computed across all stocks. The second, the Industry Concentration Index (ICI), is constructed following Kacperczyk, Sialm, and Zheng (2005) and measures mutual funds' portfolio concentration across industries. In both cases, we find that funds that perform well and those that perform poorly are, on average, significantly less diversified.

Overall, the results reported here are consistent with Kacperczyk, Nieuwerburgh, and Veldkamp (2014), who found superior funds tend to be younger, smaller, have higher expense ratios, and are more active. The results suggest our procedure that focuses directly on fund holdings is able to select these funds in real time.

In Table 11, we perform an auxiliary exercise based on fund styles. We start from the universe of funds available in our dataset and compute the number and the percentage of funds in each category. We then compute the percentage of funds that, on average, are classified in the decile portfolio of funds predicted to have the worst performance (D1) and the decile portfolio of funds predicted to have the best performance (D10). Overall, the results indicate the procedure does not systematically select funds in any specific category, in that the percentage of funds in the top and bottom portfolios are similar to the one in the full dataset. However, two categories of funds are over-represented: the US Equity Micro and Small Cap funds. The Micro Cap funds represent only 1.24% of all funds, but they represent 2.07% and 2.31% of the funds in the bottom and top deciles, respectively. Small Cap funds represent 20.40% of the overall funds universe, but they represent 23.87% and 25.67% of the funds in the bottom and top deciles, respectively.

These results suggest our procedure does not systematically select specific fund classes. Even the over-weighting of Small Cap funds is unlikely to be the main driver of the results, given that we obtain very strong monotonic patterns for risk-adjusted alphas that explicitly control for value, size and momentum effects.

7 Which Covariates Matter?

As we discussed in Section 4.4, one major advantage of BRTs over other machine learning methods is the interpretability of their results using *relative influence measures* and *partial dependence plots*. In this section, we study which covariates are important for selecting mutual funds and what economic quantities explain BRT predictions.

7.1 Relative Influence

Relative influence measures estimate the importance of each of the covariates in explaining the dependent variable. In our setting, they convey which of the 94 covariates are more and less important in predicting fund returns. One complication of our setting is that we estimate the model using one-month rolling windows, so, in principle, we would need to report one set of relative influence measures for every period in our analysis. Instead, we follow Gu, Kelly, and Xiu (2020b) and average the relative-influence measures we estimate in every period. We report the ranking and the relative influence of each variable in Figure 1. We note two facts. First, no single characteristic stands out and explains the majority of the variation in mutual fund returns. Instead, we find the 12-month momentum, which is the most important variable, accounts for only 3% of the variation in expected fund returns. Second, all regressors have an impact, because none of them has a relative influence measure equal to zero.

Given the large number of covariates we work with, appreciating what "types" of regressors matter the most by looking at Figure 1 is difficult, except for noticing that three momentum variables (12-month, 6-month, and 36-month momentum) are among the Top 10 variables with the greatest relative influence. To provide a better understanding of what variables matter the most, in Panel A of Table 12, we report the most important stock characteristics, together with their categorization following Hou, Xue, and Zhang (2020), which assigns firm characteristics into six categories: *Momentum*, *Investments*, *Value Versus Growth*, *Intangible Assets*, *Profitability*, and *Trading Frictions*. Overall, the results suggest 70% of the most relevant variables in the Top 10 are related to *Trading Frictions*. Firm characteristics in this category include *Size*, *Bid-ask Spread*, Companies' *Beta*, and *Share Turnover*. The remaining three characteristics—12-month, 6-month, and 36-month momentum—are instead *Momentum* related.

Panel B of Table 12 shows that, although *Trading Frictions* and *Momentum* categories comprise 100% of the Top 10 characteristics, they comprise only 26.59% of the overall company characteristics. The categories with the most characteristics are *Intangible*, with 29.79% of the covariates, followed by *Investments* and *Profitability*, with 18.09% and 17.02% of the covariates, respectively.

7.2 Partial Dependence Plots

Partial dependence plots provide a non-parametric estimate of the relation between each mutual fund characteristic and fund performance, integrating out the effect of all other covariates. Given that we have 468 months in our sample (1980-2018) and 94 covariates, we would need to report a total of 468 $\times$ 94=43,992 plots. Rather than reporting an overwhelming number of figures, we focus on the most important regressors: 1) 12-month momentum; 2) Industry-adjusted size; 3) Dollar trading volume, 4) and Beta. Also, we report partial dependence plots at three different points in time: January 1995, January 2005, and January 2015.

Figure 2 reports the results of the above exercise. The first row plots the results for 12-month momentum and uncovers a monotonic relation between momentum exposure and fund performance. However, this monotonic relation is not constant. In 1995, BRTs estimate a negative relation between fund momentum and performance. It is instead positive in 2005 and 2015.

The second row reports the results for dollar trading volume. Overall, the relation is non-monotonic but relatively stable over time. There is a positive relation between dollar trading volume and fund returns at the low end of the distribution, but this relation is inverted and negative for high values of dollar trading volume.

The third variable, industry-adjusted size, is different at every point in time. It is increasing in 1995, decreasing in 2005, and increasing and non-monotonic for very high levels of the regressor in 2015.

Finally, the partial-dependence plot between the beta regressor and fund returns is rather stable over time. The partial-dependence plot implies a negative relation; that is, higher beta funds have lower returns than lower beta funds.

The results reported in this section highlight that the relation between fund returns and their characteristics is time varying. BRTs are able to capture this time-variation through the rolling-window specifications, but overall, relying on a handful of regressors for the purpose of fund selection is difficult. These results also provide an explanation for the deterioration in performance we register when we move from short-rolling-window specifications to recursive-window specifications.

7.3 Dissecting BRT Predictions Using Fama-MacBeth Regressions

The results in Section 5.3—and associated Table 5—show BRT forecasts are not encompassed by the information contained in the predictors proposed in the literature so far. In this section, we look at the BRT predictions from a different perspective; that is, we assess how much of the information in the BRT forecasts is contained in the predictors of mutual fund performance that have been proposed in the literature so far. We estimate the following Fama-MacBeth regressions:

$$\widehat{BRT}_{f,t|t-1} = c\ M_{f,t-1} + \epsilon_{f,t},\ for\ t = 1, \ldots, T,$$

where $\widehat{BRT}_{f,t|t-1}$ is the BRT prediction for fund f's excess return on month t, conditional on information available as of time $t-1$, and $M_{f,t-1}$ is a vector of regressors available as of time $t-1$ that contains the following covariates: $Lag(Excess\ Returns)$ is the monthly lagged excess return of the fund; and $Fund\ Flows$, $log(TNA)$, $log(TNA)^2$, $log(Fund\ Age)$, $Expense$, and $Turnover$ represent fund flows, logged fund size, logged fund size squared, logged age of the fund, the expense ratio of the fund, and the turnover of the fund, respectively. These quantities are obtained from

CRSP and calculated following Avramov, Cheng, and Hameed (2019). The regression results are reported in Table 13.

The specification reported in Column 1 uses lagged excess returns as the only covariate and shows BRT predictions are closely related to funds' lagged excess returns. The coefficient on lagged excess returns is 0.6 with a t-statistic larger than 33. The R^2 further shows 51.4% of the variation in the BRT forecasts is explained by funds' short-term performance (Bollen and Busse, 2005). The second column adds additional covariates and shows BRT conditional returns are negatively related to fund size and positively related to fund age, but unrelated to trading frequency ($Turnover$), fund flows ($Fund\ Flows$), and expense ratios ($Expense$). These additional covariates, however, do not increase the R^2 of the Fama-MacBeth regressions.[11]

Overall, the results suggest the BRT procedure contains much more information than the one contained in the traditional predictors, because only half of the variation in BRT forecasts is explained by the predictors that have been proposed in the literature so far.

7.4 Interpreting the Predicted Returns as Dynamic DGTW Benchmarks

The idea behind the DGTW benchmarks proposed by Daniel et al. (1997) is to measure fund-manager skills by controlling for three known sources of outperformance: value, size and momentum. The predicted returns we generate in this paper using BRTs can be thought of as extensions of the DGTW benchmarks that control for 94 characteristics instead of three. In addition, our benchmark returns are dynamic, because the relevance of each of the characteristics included in BRTs varies over time and the functional form relating each characteristic and predicted fund returns is not necessarily linear. According to this interpretation, we can measure managers' skills by testing whether they systematically outperform their predicted returns, the idea being that those managers who systematically outperform their predicted returns are the ones whose outperformance is generated by exposures to factors other than the 94 we consider in our analysis.

We perform our analysis as follows. First, we keep in the sample only those funds for which we have complete returns and predicted returns series for at least three years, leaving us with a sample of 2,455 funds. We then construct each fund's under- or over-performance in a given month by subtracting each fund's predicted returns constructed out of sample using BRTs from each fund's realized returns. Finally, we compute each fund's t-statistic for the average under- or over-performance with respect to the predicted returns. We report the distribution of the t-statistics across all funds in Figure 3.

As shown in the figure, we find that no fund in our sample displays any degree of outperformance. If anything, 0.2% of the funds have t-statistics below 2. We interpret these results as an indication that the set of characteristics we consider is large enough to fully explain the returns dynamics of the mutual fund industry in both the time series and the cross-section.

[11] Note the R^2 in the first and second column are not directly comparable because they are based on samples of different sizes.

8 Robustness

Section 5 reported results for baseline specification as well as extensions along different dimensions, such as delay in information arrival and size of the rolling window. This section reports additional robustness tests that test for the stability of our baseline results. In particular, we start by reporting results using other machine learning methods and show they provide slightly worse, but comparable, performance. We then report results at the quarterly frequency, present results that feature different numbers of boosting iterations, and focus on specifications that use different numbers of sorted portfolios. All these results show the robustness of the proposed BRT approach—and other machine learning methods—for selecting mutual funds.

8.1 Selecting Mutual Funds with Other Machine Learning Methods

The goal of our paper is to focus on BRTs because they are well known for their predictive power and interpretability, and not to provide a comprehensive horse-race across all the machine learning methods currently available.[12] In fact, what we could learn from such a horse race is not clear, because the "no free lunch theorem" asserts that no machine learning algorithm can be expected ex-ante to outperform others on any given task (Wolpert, 1996). Learning that a certain machine learning method outperforms others in a given sample and setting does not mean it would also outperform in a different sample of the same setting or in a different setting.

Nevertheless, to assess whether BRTs are unique or we can obtain comparable performance from other machine learning methods, we repeat the analysis reported in Table 3 using other standard machine learning methods, including two shrinkage methods, namely, lasso, elastic net, random forests (a tree-based method), and neural networks with one to five hidden layers.[13] We select these methods because they are reported as the most effective machine learning methods in other asset-pricing studies.[14]

The results, reported in Table 14, show all machine learning methods perform reasonably well in distinguishing between low- and high-performing mutual funds. For example, the long-short portfolios for lasso and elastic net have returns of 5.52% and 5.77%, respectively. Random forests have a long-short portfolio with returns equal to 5.33%. Finally, the neural networks with one to five hidden layers deliver long-short portfolio returns of around 6%, with the three-hidden-layer neural network performing the best, with a long-short portfolio return of 6.58%. These results are reassuring because they validate that the fund-level characteristics we construct, although noisy, are informative. They also show our findings are not unique to BRTs but extend to other machine learning methods.

[12]For comprehensive horse races across machine learning methods, see Fernández-Delgado, Cernadas, Barro, and Amorim (2014), Gu, Kelly, and Xiu (2020b) and Bianchi, Büchner, and Tamoni (2020).

[13]We stop at five hidden layers following Gu, Kelly, and Xiu (2020b). In all cases, we follow Gu, Kelly, and Xiu (2020b) and cross-validate the hyper-parameters recursively.

[14]Both Gu, Kelly, and Xiu (2020b) and Bianchi, Büchner, and Tamoni (2020) report the outperformance of tree-based methods and neural networks over other methods in predicting asset returns.

8.2 Results Conditioning on Fund Style

The results we have reported so far focus on all funds, irrespective of their style. However, the procedure may be successful simply because it selects funds with certain styles when such styles are profitable. To allay this concern, we construct three fund categories and re-estimate the baseline results contained in Table 3 within each category. The first category is “Growth” and includes funds categorized as “Equity-Domestic-Style-Growth” in CRSP, for a total of 1,322 funds over our sample. The second category is “Growth and Income” and includes funds categorized as “Equity-Domestic-Style-Growth,” “Growth and Income,” and “Income” in CRSP, for a total of 1,912 funds. The third category is “Small and Mid Cap” and includes funds categorized as “Equity-Domestic-Cap-based-Micro Cap,” “Small Cap,” and “Mid Cap,” for a total of 1,052 funds. The results are reported in Table 11, where the first three columns report results for equal-weighted returns and the remaining three columns report the ones for value-weighted returns.

Our estimates confirm the baseline results in that the BRT forecasts generate highly significant long-short portfolios for the “Growth,” “Growth and Income,” and “Small and Mid Cap” categories. In all cases, the economic magnitudes are in line with the ones in Table 3 that are based on the universe of mutual funds.

8.3 Estimating the Results at the Quarterly Frequency

The baseline results reported in the paper focus on monthly returns. The holding variables, however, vary only quarterly. Therefore, implementing our BRT framework at the quarterly frequency rather than the monthly frequency is natural. The implementation entails estimating quarterly the expected returns for all the returns in the sample and split them into portfolios that are held for one quarter.

Table A.3 shows the BRT approach’s performance in the quarterly frequency. The structure is virtually identical to that of Table 4: the first four columns focus on raw firm characteristics; columns 5 through 8 focus on firm-characteristics ranks; and column 9 through 12 focus on firm-characteristics deciles.

Starting from the results that use raw characteristics, we highlight three facts. First, both net and gross returns portfolios are monotonically increasing in expected returns, irrespective of whether we use equal-weighting or value-weighting schemes. Second, the long-short portfolios have economically and significant excess returns. For example, for equal-weighted net return specification, the long-short portfolio has an annual excess return of 4.08%, statistically significant at the 5% level. The Fama-French Carhart four-factor alphas are also economically significant but have a statistical significance of only 10%. Interestingly, we find value-weighted portfolios have better returns and alphas than equal-weighted portfolios. Third, the quarterly results are considerably worse than the monthly returns reported in Table 4, suggesting that working at the higher frequency allows us to capture short-lived investment opportunities.

The results for firm-characteristics ranks and firm-characteristics deciles are qualitatively similar to the raw firm-characteristics results and are quantitatively slightly better: they entail long-short portfolios with superior excess returns and risk-adjusted alphas.

8.4 Estimating Results across Different Boosting Iterations

One key parameter for BRT is the number of boosting iterations. Many studies (Hastie, Tibshirani, and Friedman, 2009) show that in many empirical applications, the out-of-sample performance of BRTs stabilizes as the number of boosting iterations increases, and further increasing the number of boosting iterations does not lead to overfitting. In other applications, on the other hand, increasing the number of boosting iterations leads to overfitting. Ultimately, whether BRTs overfit as we increase the number of boosting iterations is an empirical question.

In Table A.4, we report the results for BRTs as we vary the number of boosting iterations from 10 to 10,000 and report the performance for the equal-weighted long-short portfolio that buys the funds with the highest expected returns and sells the funds with the lowest expected returns. As the number of boosting iterations increases, the BRT long-short portfolio performance quickly increases from an outperformance of 4.77% at 10 boosting iterations to 6.92% at 30 boosting iterations. The performance then slightly worsens to 6.68% at 100 boosting iterations, the baseline results reported here, and then slightly deteriorates to 6.51% at 10,000 boosting iterations.

Across all specifications, we find BRTs lead to long-short portfolios with significant excess returns and risk-adjusted alphas, suggesting BRTs do not overfit the training sample as we increase the number of boosting iterations. This result is particularly important because it shows cross-validating the optimal number of boosting iterations—which is generally a computationally intensive task and a common approach to set the values of hyper-parameters in machine learning methods—is not crucial for BRTs.

8.5 Performance with Different Numbers of Portfolios

The baseline results are computed using 10 portfolios, which is an arbitrary number. In principle, if the procedure is able to select high-performing and low-performing funds in the tails of the distributions, we would expect to see that the long-short portfolios generated using 20, 30, 50, or more portfolios would have better performance. If, on the other hand, the predicted returns are very noisy, we may expect that increasing the number of portfolios would decrease the performance of the procedure.

In Table A.5 and Figure A.1, we report the BRT results for different numbers of portfolios. The results indicate the out-of-sample performance of the long-short portfolios increases monotonically as we increase the number of portfolios. It goes from 6.68% when we use 10 portfolios to 11.20% when we use 80 portfolios.

Overall, the results reported here suggest our results are robust to the choice of the portfolio number and, if anything, our baseline results are a lower bound of the potential of BRTs.

9 Conclusion

We select mutual funds in real time by combining individual fund holdings and a large number (94) of stock characteristics to compute fund-level characteristics on the basis of the stocks they hold. We show that, first, the majority of funds are largely exposed—both positively and negatively—to approximately 40-50 characteristics, that fund performance is non-linearly related to fund characteristics, and there are significant degrees of interaction between different fund characteristics and fund performance.

Second, when we predict fund performance, these non-linearities and interactions prove important as machine learning methods such as boosted regression trees (BRTs) significantly outperform standard linear frameworks and the BRT-generated forecasts encompass the ones generated by the predictors of mutual fund performance that have been proposed in the literature so far.

Furthermore, although in our setting, BRTs outperform other machine learning methods, such as lasso, elastic nets, random forests, and neural networks with one to five hidden layers, these other machine learning methods deliver good performance and they all outperform ordinary least squares models.

Finally, although we detect significant predictability using machine learning methods, the fund characteristics that matter the most in predicting fund returns and the functional relation between fund characteristics and fund performance are time varying.

References

Akbas, F., W. J. Armstrong, S. Sorescu, and A. Subrahmanyam. 2015. Smart money, dumb money, and capital market anomalies. *Journal of Financial Economics* 118:355–82.

Ali, A., X. Chen, T. Yao, and T. Yu. 2008. Do Mutual Funds Profit from the Accruals Anomaly? *Journal of Accounting Research* 46:1–26.

Amihud, Y., and R. Goyenko. 2013. Mutual Fund's R2 as Predictor of Performance. *Review of Financial Studies* 26:667–94.

Avramov, D., S. Cheng, and A. Hameed. 2019. Mutual Funds and Mispriced Stocks. *Management Science* forthcoming.

Baker, M., and J. Wurgler. 2006. Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance* 61:1645–80.

Barras, L., O. Scaillet, and R. Wermers. 2010. False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas. *The Journal of Finance* 65:179–216.

Berk, J. B., and R. C. Green. 2004. Mutual Fund Flows and Performance in Rational Markets. *Journal of Political Economy* 112:1269–95.

Berk, J. B., and J. H. Van Binsbergen. 2015. Measuring skill in the mutual fund industry. *Journal of Financial Economics* 118:1–20.

Bianchi, D., M. Büchner, and A. Tamoni. 2020. Bond Risk Premiums with Machine Learning. *The Review of Financial Studies* forthcoming.

Bollen, N. P. B., and J. A. Busse. 2005. Short-Term Persistence in Mutual Fund Performance. *The Review of Financial Studies* 18:569–97. ISSN 0893-9454. Publisher: Oxford Academic.

Breiman, L. 1996. Bagging Predictors. *Machine Learning* 24:123–40.

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen. 1984. *Classification and regression trees.* CRC press.

Busse, J. A., and P. J. Irvine. 2006. Bayesian Alphas and Mutual Fund Persistence. *The Journal of Finance* 61:2251–88.

Calluzzo, P., F. Moneta, and S. Topaloglu. 2019. When Anomalies Are Publicized Broadly, Do Institutions Trade Accordingly? *Management Science* 65:4555–74.

Carhart, M. M. 1997. On Persistence in Mutual Fund Performance. *The Journal of Finance* 52:57–82.

Chan, L. K. C., H.-L. Chen, and J. Lakonishok. 2002. On Mutual Fund Investment Styles. *The Review of Financial Studies* 15:1407–37.

Chen, J., H. Hong, M. Huang, and J. D. Kubik. 2004. Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization. *The American Economic Review* 94:1276–302.

Chinco, A., A. Neuhierl, and M. Weber. 2020. Estimating the anomaly base rate. *Journal of Financial Economics* .

Cremers, K. J. M., and A. Petajisto. 2009. How Active Is Your Fund Manager? A New Measure That Predicts Performance. *The Review of Financial Studies* 22:3329–65.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. *The Journal of Finance* 52:1035–58.

DeMiguel, V., J. Gil-Bazo, F. J. Nogales, and A. Santos. 2021. Can machine learning help to select portfolios of mutual funds? *SSRN working papers* .

Doshi, H., R. Elkamhi, and M. Simutin. 2015. Managerial Activeness and Mutual Fund Performance. *The Review of Asset Pricing Studies* 5:156–84.

Du, M., N. Liu, and X. Hu. 2019. Techniques for interpretable machine learning. *Communications of ACM* 63:68–77.

Edelen, R. M., O. S. Ince, and G. B. Kadlec. 2016. Institutional investors and stock return anomalies. *Journal of Financial Economics* 119:472–88.

Elton, E. J., M. J. Gruber, and C. R. Blake. 2011. Holdings Data, Security Returns, and the Selection of Superior Mutual Funds. *Journal of Financial and Quantitative Analysis* 46:341–67.

Elton, E. J., M. J. Gruber, S. Das, and M. Hlavka. 1993. Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios. *The Review of Financial Studies* 6:1–22.

Evans, R. B. 2010. Mutual Fund Incubation. *The Journal of Finance* 65:1581–611.

Fama, E. F., and K. R. French. 2010. Luck versus Skill in the Cross-Section of Mutual Fund Returns. *The Journal of Finance* 65:1915–47.

Feng, G., S. Giglio, and D. Xiu. 2020. Taming the Factor Zoo: A Test of New Factors. *Journal of Finance* 75:1327–70.

Fernández-Delgado, M., E. Cernadas, S. Barro, and D. Amorim. 2014. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research* 15:3133–81.

Freyberger, J., A. Neuhierl, and M. Weber. 2020. Dissecting Characteristics Nonparametrically. *The Review of Financial Studies* 33:2326–77.

Friedman, J. H. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29:1189–232.

Green, J., J. R. M. Hand, and X. F. Zhang. 2017. The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *The Review of Financial Studies* 30:4389–436.

Grinblatt, M., S. Titman, and R. Wermers. 1995. Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior. *The American Economic Review* 85:1088–105.

Groenborg, N., A. Lunde, A. Timmermann, and R. Wermers. 2021. Picking Funds with Confidence. *Journal of Financial Economics* 139:1–28.

Gu, S., B. Kelly, and D. Xiu. 2020a. Autoencoder asset pricing models. *Journal of Econometrics* forthcoming.

Gu, S., B. T. Kelly, and D. Xiu. 2020b. Empirical Asset Pricing via Machine Learning. *Review of Financial Studies* 33:2223–73.

Gupta-Mukherjee, S. 2014. Investing in the "New Economy": Mutual Fund Performance and the Nature of the Firm. *Journal of Financial and Quantitative Analysis* 49:165–91.

Harvey, C. R., and Y. Liu. 2018. Detecting Repeatable Performance. *The Review of Financial Studies* 31:2499–552.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hendricks, D., J. Patel, and R. Zeckhauser. 1993. Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974–1988. *The Journal of Finance* 48:93–130.

Hoberg, G., N. Kumar, and N. Prabhala. 2018a. Mutual Fund Competition, Managerial Skill, and Alpha Persistence. *The Review of Financial Studies* 31:1896–929.

———. 2018b. Mutual fund competition, managerial skill, and alpha persistence. *The Review of Financial Studies* 31:1896–929.

Hou, K., C. Xue, and L. Zhang. 2020. Replicating Anomalies. *The Review of Financial Studies* 33:2019–133.

Huang, J., C. Sialm, and H. Zhang. 2011. Risk Shifting and Mutual Fund Performance. *The Review of Financial Studies* 24:2575–616.

Jiang, H., and L. Zheng. 2018. Active Fundamental Performance. *The Review of Financial Studies* 31:4688–719.

Jones, C. S., and H. Mo. 2020. Out-of-Sample Performance of Mutual Fund Predictors. *The Review of Financial Studies* 34:149–93.

Kacperczyk, M., S. V. Nieuwerburgh, and L. Veldkamp. 2014. Time-Varying Fund Manager Skill. *The Journal of Finance* 69:1455–84.

Kacperczyk, M., and A. Seru. 2007. Fund Manager Use of Public Information: New Evidence on Managerial Skills. *The Journal of Finance* 62:485–528.

Kacperczyk, M., C. Sialm, and L. Zheng. 2005. On the Industry Concentration of Actively Managed Equity Mutual Funds. *The Journal of Finance* 60:1983–2011.

———. 2008. Unobserved Actions of Mutual Funds. *The Review of Financial Studies* 21:2379–416.

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.

Kosowski, R., A. Timmermann, R. Wermers, and H. White. 2006. Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis. *The Journal of Finance* 61:2551–95.

Lewellen, J. 2011. Institutional investors and the limits of arbitrage. *Journal of Financial Economics* 102:62–80.

Lou, D. 2012. A Flow-Based Explanation for Return Predictability. *The Review of Financial Studies* 25:3457–89.

Mamaysky, H., M. Spiegel, and H. Zhang. 2007. Improved Forecasting of Mutual Fund Alphas and Betas. *Review of Finance* 11:359–400.

McLean, R. D., and J. Pontiff. 2016. Does Academic Research Destroy Stock Return Predictability? *Journal of Finance* 71:5–32.

Nallareddy, S., and M. Ogneva. 2017. Accrual quality, skill, and the cross-section of mutual fund returns. *Review of Accounting Studies* 22:503–42.

Newey, W. K., and K. D. West. 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:703–8.

Pástor, Ľ., R. F. Stambaugh, and L. A. Taylor. 2015. Scale and skill in active management. *Journal of Financial Economics* 116:23–45.

———. 2017. Do Funds Make More When They Trade More? *The Journal of Finance* 72:1483–528.

Patton, A. J., and A. Timmermann. 2010. Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts. *Journal of Financial Economics* 98:605–25.

Wolpert, D. H. 1996. The Lack of A Priori Distinctions Between Learning Algorithms. *Neural Computation* 8:1341–90.

Wu, W., J. Chen, Z. Yang, and M. L. Tindall. 2021. A cross-sectional machine learning approach for hedge fund return prediction and selection. *Management Science* 67:4577–601.

Zheng, L. 1999. Is Money Smart? A Study of Mutual Fund Investors' Fund Selection Ability. *The Journal of Finance* 54:901–33.

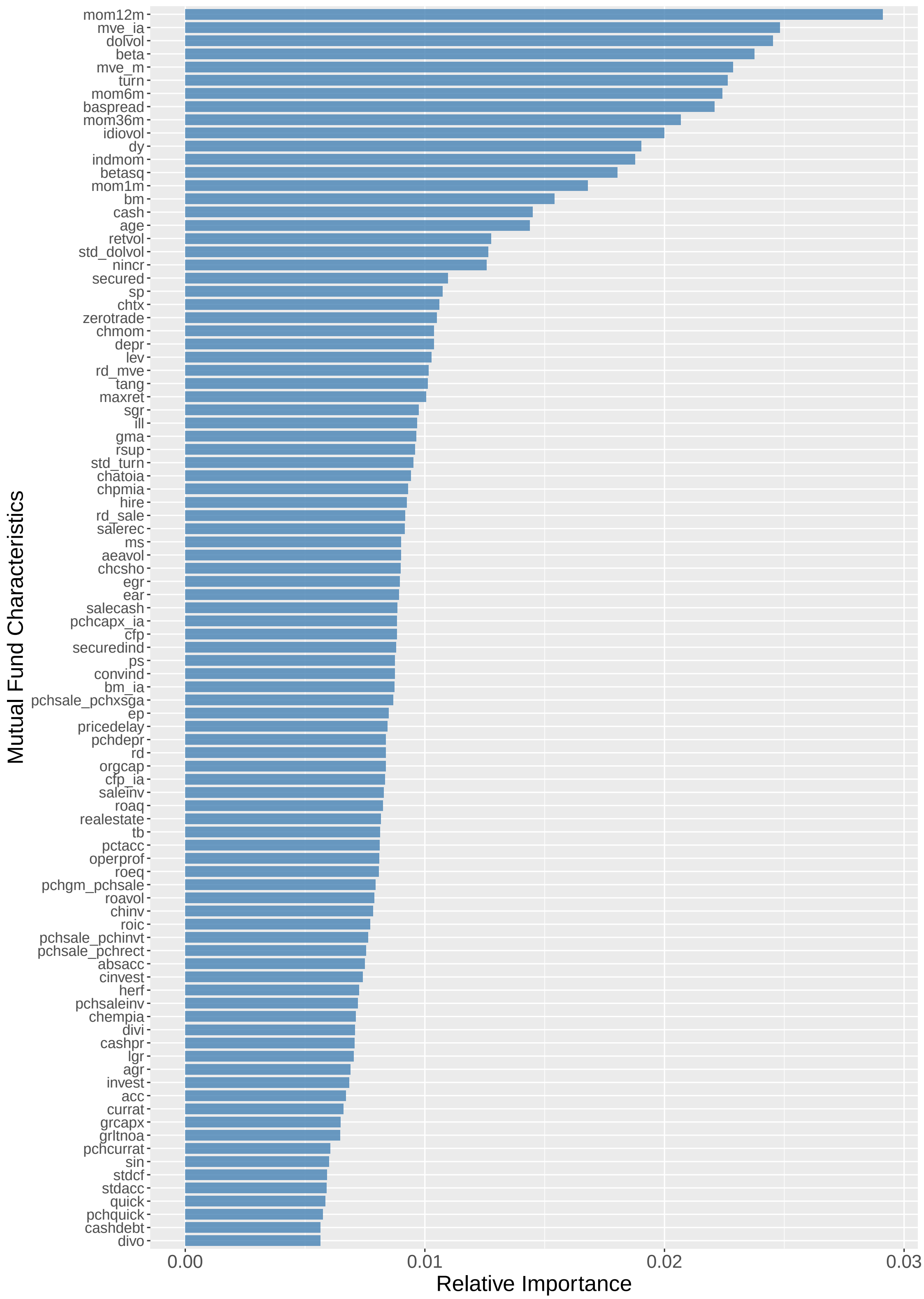
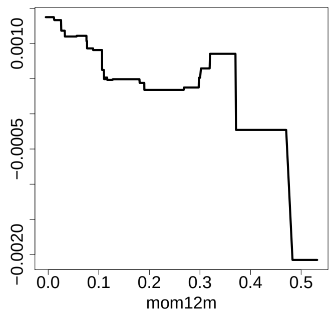
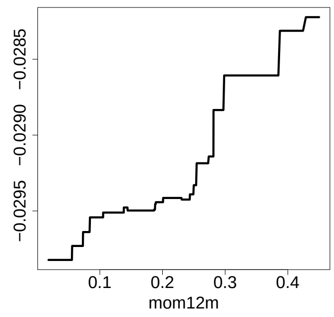


Figure 1: **Covariates' relative influence in the BRTs model**

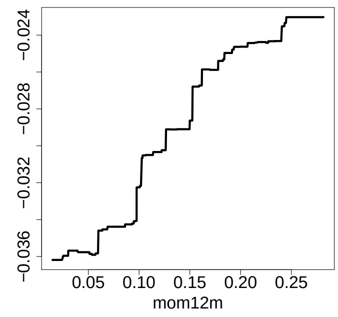
This figure shows the covariates' relative influence measures in the BRTs model. The y axis shows mutual funds' 94 holding characteristics, and the x axis presents each covariate's relative influence measure. The relative importance measure across all covariates sums to 1.



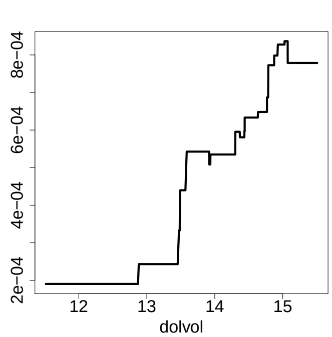
(a) Momentum (1995)



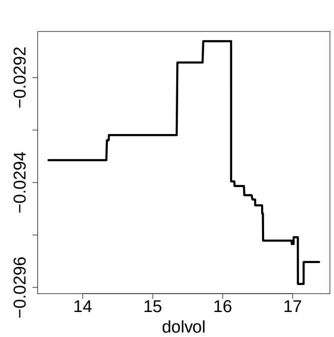
(b) Momentum (2005)



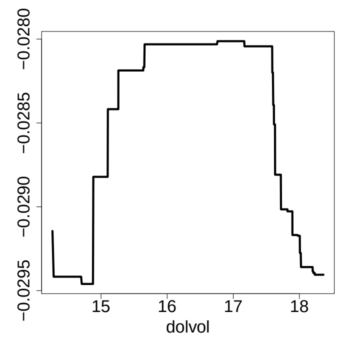
(c) Momentum (2015)



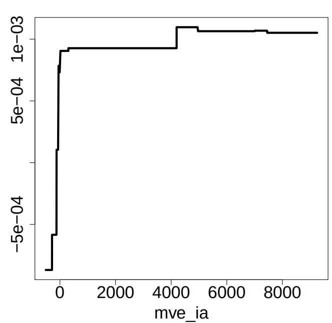
(d) Dollar trading volume (1995)



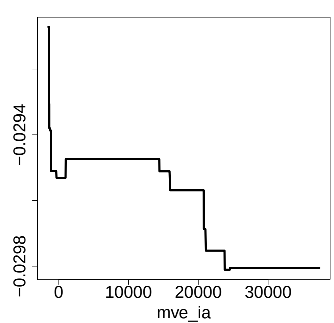
(e) Dollar trading volume (2005)



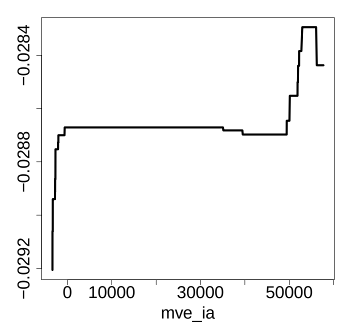
(f) Dollar trading volume (2015)



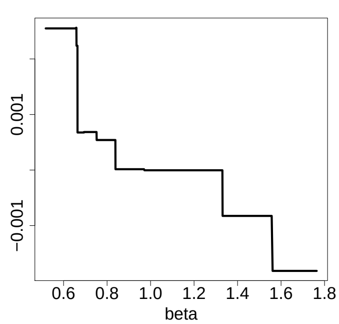
(g) Industry-adjusted Size (1995)



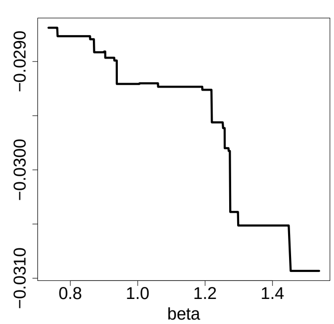
(h) Industry-adjusted Size (2005)



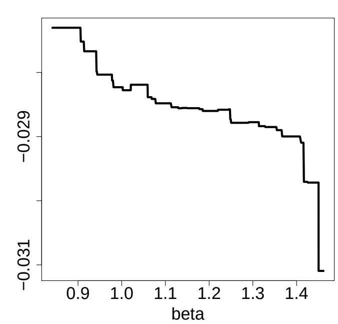
(i) Industry-adjusted Size (2015)



(j) Beta (1995)



(k) Beta (2005)



(l) Beta (2015)

Figure 2: **Top 4 characteristics partial dependent plots: 1995, 2005 and 2015.** This figure shows the partial dependence plots on January 1995, 2005 and 2015 for the four most important fund characteristics: *mom*12*m*, 12-month momentum; *dolvol*, Dollar trading volume; *mve_ia*, Industry-adjusted size; and *beta*. The x-axis contains the support of each fund characteristic, and the y-axis reports the relation between fund return and each fund characteristic, integrating out the effect of all other regressors.

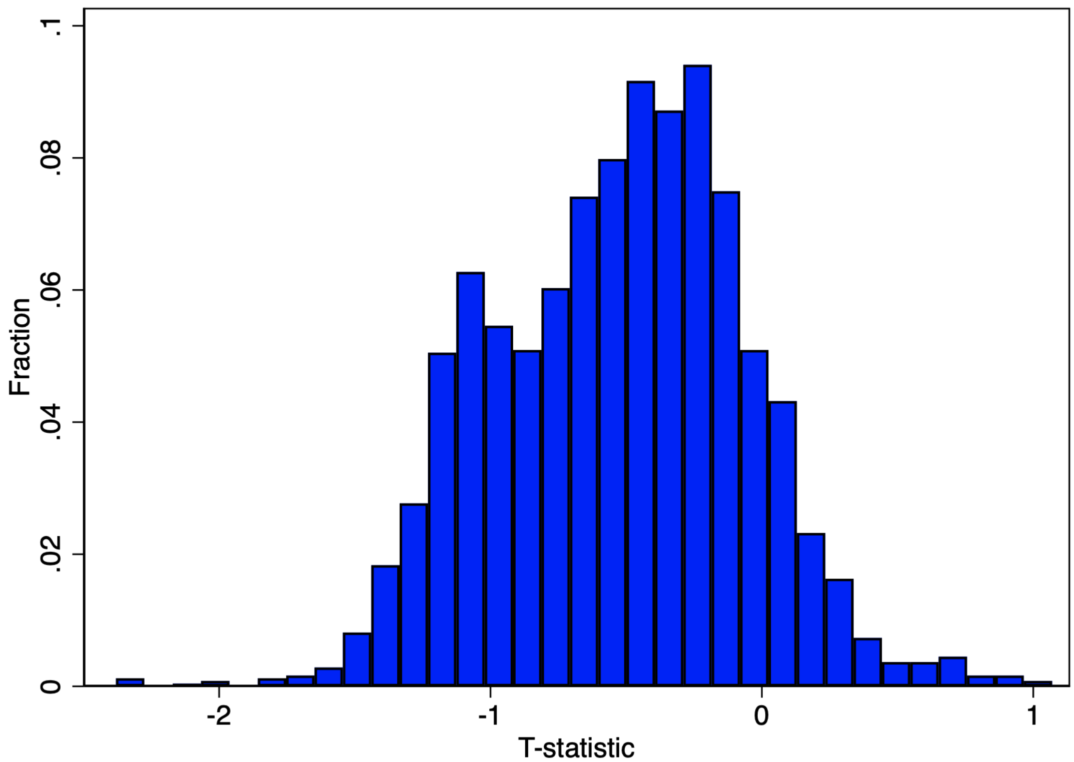


Figure 3: **Interpreting the Predicted Returns as Dynamic DGTW Benchmarks.** This figure presents the cross-sectional distribution of the t-statistics for the average under- or over-performance of each fund with respect to the predicted returns generated using BRTs. The results are computed as follows. We first keep only those funds for which we have complete returns and predicted return series for at least three years. We then construct each fund's under- or over-performance in a given month by subtracting each fund's predicted returns constructed out-of-sample using BRTs from each fund's realized returns. Finally, we compute each fund's t-statistic for the average under- or over-performance with respect to the predicted returns.

Table 1: Mutual funds holding characteristics.

Panel A: Mutual Funds Holdings Characteristics

Char.	mean	std	25%	median	75%	(%<5.5)	(%>5.5)
mve_m	9.19	0.97	8.59	9.64	9.91	0.35	97.46
dolvol	9.18	0.95	8.69	9.59	9.88	0.47	98.56
mve_ia	8.15	1.74	6.75	8.83	9.62	4.96	77.21
age	7.46	1.26	6.63	7.72	8.50	3.01	79.78
roeq	7.28	0.70	6.77	7.36	7.80	0.43	96.21
roic	7.26	0.67	6.81	7.29	7.71	0.20	93.33
roaq	7.24	0.66	6.80	7.27	7.70	0.20	96.49
ps	7.18	0.43	6.97	7.20	7.44	0.27	97.27
ms	7.11	0.76	6.59	7.15	7.66	0.43	90.63
turn	6.90	0.85	6.31	6.85	7.44	1.21	76.54
cashdebt	6.88	0.59	6.45	6.89	7.29	0.16	90.87
mom36m	6.72	0.78	6.24	6.74	7.23	0.74	56.17
cashpr	6.69	1.06	5.87	6.78	7.58	5.04	62.33
mom12m	6.58	0.72	6.13	6.52	7.00	0.39	67.72
operprof	6.52	0.55	6.16	6.56	6.89	0.86	81.97
ep	6.40	0.79	5.85	6.45	6.99	2.85	61.48
realestate	6.38	0.60	5.95	6.41	6.83	1.21	70.84
tb	6.35	0.46	6.05	6.35	6.63	0.20	81.89
gma	6.34	0.80	5.77	6.33	6.97	5.62	64.09
mom6m	6.29	0.50	6.00	6.24	6.53	0.23	78.81
egr	6.23	0.61	5.78	6.21	6.67	1.17	58.94
hire	6.21	0.80	5.60	6.15	6.81	5.54	52.97
aeavol	6.18	0.38	5.91	6.15	6.43	0.51	90.98
agr	6.13	0.72	5.58	6.07	6.65	2.89	48.24
chtx	6.11	0.62	5.63	6.04	6.51	1.83	55.19
dy	6.06	1.20	5.20	6.08	6.96	21.74	48.28
cfp	6.02	0.75	5.44	6.08	6.61	11.75	47.66
herf	5.97	0.62	5.56	5.96	6.37	3.79	27.83
invest	5.97	0.58	5.56	5.93	6.38	3.01	46.76
grltnoa	5.93	0.50	5.58	5.90	6.28	2.46	46.64
indmom	5.93	0.41	5.73	5.92	6.10	0.20	25.25
sgr	5.93	0.80	5.32	5.84	6.49	11.12	42.94
cfp_ia	5.91	0.31	5.74	5.90	6.07	0.20	43.25
lgr	5.91	0.50	5.53	5.87	6.25	4.14	46.45
mom1m	5.90	0.17	5.82	5.89	5.97	0.04	92.04
saleinv	5.88	0.40	5.63	5.92	6.15	3.67	52.34
salecash	5.87	0.63	5.54	5.90	6.30	5.70	43.75
salerec	5.87	0.41	5.62	5.87	6.11	3.04	50.04
chinv	5.86	0.39	5.59	5.87	6.13	1.48	44.85
grcapx	5.85	0.50	5.50	5.79	6.17	2.15	38.84
pchgm_pchsale	5.84	0.28	5.68	5.85	6.01	0.47	48.09
nincr	5.83	0.30	5.62	5.77	5.99	0.62	64.91
ear	5.82	0.23	5.68	5.79	5.93	0.04	79.12
beta	5.82	0.90	5.14	5.72	6.47	12.06	27.17
rsup	5.78	0.48	5.47	5.76	6.08	4.29	38.13
betasq	5.77	0.91	5.08	5.67	6.42	14.60	25.33
chempia	5.75	0.55	5.38	5.74	6.12	4.29	29.31
convind	5.73	0.26	5.55	5.72	5.88	5.39	32.79

pchsale_pchxsga	5.71	0.30	5.55	5.69	5.86	0.90	27.36
std_turn	5.70	1.12	4.80	5.63	6.47	34.39	43.40
securedind	5.66	0.67	5.18	5.69	6.13	15.11	33.14
pchquick	5.62	0.20	5.53	5.61	5.70	0.39	12.18
pchcurrat	5.59	0.19	5.51	5.59	5.67	0.74	9.76
cinvest	5.59	0.15	5.52	5.58	5.65	0.43	17.64
pchcapx_ia	5.56	0.32	5.36	5.55	5.75	5.43	16.24
pchsaleinv	5.54	0.22	5.43	5.54	5.65	2.34	6.32
acc	5.52	0.38	5.27	5.52	5.77	11.67	16.04
pchsale_pchinvt	5.52	0.22	5.40	5.51	5.63	3.08	4.88
cash	5.50	0.81	4.90	5.44	6.06	31.42	27.99
chmom	5.48	0.24	5.37	5.49	5.60	3.71	5.07
rd	5.47	0.16	5.38	5.46	5.55	15.11	3.51
pctacc	5.47	0.44	5.19	5.47	5.74	17.99	14.64
chpmia	5.46	0.24	5.37	5.47	5.57	1.83	1.13
pchsale_pchrect	5.45	0.17	5.38	5.45	5.52	2.34	1.48
chatoia	5.41	0.26	5.30	5.41	5.53	5.35	2.62
depr	5.34	0.61	4.93	5.30	5.76	27.63	17.92
pchdepr	5.29	0.19	5.20	5.29	5.37	25.64	0.20
rd_sale	5.28	0.73	4.81	5.32	5.74	32.63	13.58
pricedelay	5.21	0.36	5.00	5.14	5.36	57.96	2.62
orgcap	5.20	0.66	4.83	5.21	5.60	29.12	9.88
chcsho	5.18	1.07	4.39	5.13	5.89	36.38	15.03
bm_ia	5.15	0.43	4.92	5.18	5.40	30.84	1.72
quick	5.14	0.75	4.59	5.06	5.60	47.89	13.04
currat	5.14	0.74	4.58	5.09	5.63	46.88	13.82
divi	5.13	0.07	5.09	5.13	5.17	97.42	0.00
divo	5.09	0.08	5.03	5.07	5.12	97.58	0.00
sin	5.08	0.10	5.02	5.06	5.11	96.25	0.04
absacc	5.03	0.49	4.67	4.97	5.35	53.32	5.00
lev	4.98	1.11	4.06	5.05	5.85	43.56	19.28
sp	4.91	1.00	4.19	4.91	5.62	42.94	13.35
maxret	4.81	0.87	4.17	4.70	5.41	69.20	12.88
baspread	4.74	0.98	4.03	4.55	5.41	67.49	11.79
rd_mve	4.74	0.48	4.47	4.72	4.97	70.53	1.13
retvol	4.65	0.95	3.96	4.52	5.31	71.23	10.07
tang	4.58	0.71	4.07	4.47	4.98	71.90	2.54
idiovol	4.49	1.12	3.66	4.33	5.25	53.83	7.92
roavol	4.41	0.78	3.84	4.28	4.84	75.14	3.12
bm	4.07	1.08	3.22	3.98	4.85	68.35	4.22
stdacc	4.02	0.60	3.59	3.96	4.38	90.83	0.08
secured	4.01	0.90	3.31	3.73	4.64	80.29	2.42
stdcf	4.01	0.64	3.55	3.93	4.40	88.29	0.20
zerotrade	3.89	0.79	3.40	3.93	4.42	95.24	0.59
std_dolvol	2.66	0.91	1.97	2.42	3.37	98.71	0.27
ill	1.85	0.99	1.11	1.40	2.41	97.85	0.55

Panel B: Percentage of Funds with Significant AIM Scores

	(0, 10]	(10,20]	(20,30]	(30,40]	(40, 50]	(50,60]	(60,70]
# (AIM < 5.5)	25	971	1471	95			
% (AIM < 5.5)	0.98%	37.90%	57.42%	3.71%			
# (AIM > 5.5)	4	125	892	869	577	92	3
% (AIM > 5.5)	0.16%	4.88%	34.82%	33.92%	22.52%	3.59%	0.12%

This table summarizes mutual funds Anomalies Investing Measure (AIM) scores. Results are computed by converting all stocks' characteristics into deciles and calculating each fund's average AIM score, i.e., the average decile for each fund's characteristic using stock characteristics and holdings data. For each characteristic, Panel A shows mean, standard deviation, 25th, 50th, and 75th percentiles of the AIM scores across all funds. The last two columns compute the percentage of funds whose holding characteristics are statistically greater and smaller than 5.5 at the 1% level. Panel B first computes, for each fund, the number of characteristics for which the AIM score is statistically different from zero. It then reports the distribution of the number of significant characteristics across all funds. In particular, we discretize the number of significant characteristics into seven groups and report the number and percentage of funds in each group. Statistical significance is computed using Newey and West (1987) standard errors with 12 lags.

Table 2: Univariate analysis of mutual fund level characteristics and performance.

	Equal Weighted				Value Weighted			
Char.	Ret	t-stat	FFC α	t-stat	Ret	t-stat	FFC α	t-stat
salerec	2.20	2.63	1.85	2.65	2.72	2.13	2.31	2.26
pchsaleinv	2.04	2.21	1.11	1.03	3.06	2.58	2.23	1.73
rd_mve	2.94	2.09	1.27	1.33	3.28	2.17	1.86	1.63
chatoia	2.49	2.06	1.58	1.53	3.67	2.37	2.74	1.95
ear	2.42	2.05	0.67	0.60	2.49	2.08	0.79	0.74
saleinv	1.66	1.96	1.56	1.81	1.00	1.05	0.64	0.70
ps	1.48	1.61	1.87	1.91	1.84	1.55	2.34	1.98
orgcap	1.66	1.59	0.17	0.28	2.18	2.03	1.03	1.43
rsup	2.43	1.54	0.86	0.66	2.09	1.24	0.37	0.26
ill	3.78	1.50	2.18	1.27	1.68	0.87	1.05	1.31
nincr	2.10	1.32	0.45	0.57	2.02	1.35	0.53	0.55
rd	0.91	1.29	0.30	0.38	0.88	0.89	0.25	0.26
cinvest	0.65	1.26	0.34	0.62	1.02	1.85	0.92	1.45
chpmia	1.51	1.20	2.73	2.79	1.93	1.54	2.85	2.88
pchsale_pchinvt	0.90	1.12	1.21	1.23	0.51	0.52	0.90	0.84
tb	0.95	1.02	1.58	1.84	1.29	1.07	2.08	2.04
roaq	1.38	0.98	2.07	1.79	1.45	0.91	2.27	1.68
std_dolvol	1.78	0.97	0.26	0.30	1.58	0.87	0.40	0.47
sp	2.55	0.97	0.64	0.57	2.70	1.02	0.85	0.67
salecash	2.16	0.95	1.47	1.19	2.27	1.00	1.54	1.19
divi	1.21	0.94	-0.08	-0.09	1.18	0.95	-0.06	-0.07
mom1m	2.14	0.85	0.29	0.14	3.72	1.88	1.96	1.23
securedind	1.27	0.84	-0.35	-0.51	0.63	0.45	-0.74	-1.01
chtx	1.79	0.83	0.05	0.05	1.09	0.50	-0.47	-0.44
secured	1.76	0.81	0.26	0.23	1.15	0.51	0.05	0.05
stdacc	0.85	0.80	-0.34	-0.54	1.02	0.99	-0.31	-0.44
cfp	2.42	0.76	1.71	0.78	0.89	0.32	0.58	0.37
pchdepr	1.23	0.73	0.77	0.67	1.13	0.63	0.70	0.56
aeavol	1.19	0.70	-0.09	-0.11	0.59	0.32	-0.45	-0.46
sin	0.49	0.63	1.00	1.53	1.05	0.87	1.11	1.18
mom12m	1.72	0.59	-2.66	-1.32	3.25	1.35	-1.28	-0.99
mom6m	1.48	0.52	-2.92	-1.26	3.87	1.67	-0.74	-0.44
zerotrade	1.23	0.52	2.25	1.66	1.47	0.59	2.82	2.04
divo	0.67	0.51	-0.48	-0.62	0.88	0.63	-0.30	-0.36
stdcf	0.78	0.51	-1.10	-1.23	0.98	0.63	-0.87	-0.82
cashdebt	0.50	0.49	1.77	2.04	0.49	0.39	2.19	1.95
std_turn	1.24	0.48	-0.86	-0.59	0.27	0.10	-1.93	-1.25
bm	1.24	0.48	-0.33	-0.31	1.22	0.46	-0.35	-0.30
gma	1.01	0.44	1.25	1.01	0.56	0.24	1.60	1.26
depr	0.83	0.37	-0.09	-0.07	0.86	0.35	0.38	0.27
absacc	0.86	0.36	-0.45	-0.36	0.64	0.26	-0.07	-0.05
ep	0.88	0.35	1.80	1.27	1.51	0.57	2.03	1.40
lev	0.80	0.33	0.34	0.27	1.65	0.72	0.74	0.57
pchsale_pchxsga	0.85	0.33	0.10	0.05	-1.65	-0.86	-1.53	-1.29
realestate	0.48	0.32	1.16	0.98	1.08	0.63	1.60	1.20
rd_sale	0.63	0.27	0.32	0.25	0.79	0.34	0.76	0.57
roic	0.33	0.26	2.03	1.85	0.12	0.08	1.89	1.50
cash	0.71	0.26	0.85	0.52	0.47	0.16	0.94	0.59
maxret	0.66	0.24	-1.62	-1.09	-0.10	-0.03	-2.10	-1.23
idiovol	0.55	0.20	-1.95	-1.34	0.09	0.03	-2.19	-1.37
retvol	0.46	0.16	-1.87	-1.19	-0.47	-0.15	-2.84	-1.71
tang	0.38	0.16	-0.46	-0.32	0.34	0.13	-0.12	-0.07
roavol	0.37	0.14	-1.36	-1.01	0.41	0.16	-0.92	-0.70
indmom	0.19	0.07	-3.40	-1.56	2.40	1.29	-1.22	-0.89
pchgm_pchsale	0.06	0.07	-0.10	-0.12	-0.48	-0.42	-0.63	-0.63
chempia	0.05	0.03	-0.35	-0.29	-0.13	-0.06	-0.64	-0.50
turn	0.03	0.01	-2.27	-1.51	0.06	0.02	-2.37	-1.62
beta	-0.04	-0.01	-2.42	-1.51	-0.78	-0.25	-3.19	-1.99

	Equal Weighted				Value Weighted			
Char.	Ret	t-stat	FFC α	t-stat	Ret	t-stat	FFC α	t-stat
baspread	-0.05	-0.02	-2.22	-1.35	-0.78	-0.25	-3.13	-1.85
pricedelay	-0.12	-0.10	0.39	0.40	0.72	0.53	1.21	1.11
dy	-0.30	-0.11	0.96	0.77	-0.36	-0.13	0.43	0.33
roeq	-0.28	-0.13	0.17	0.09	1.69	1.11	2.07	1.60
quick	-0.31	-0.16	-1.19	-1.07	-0.56	-0.28	-1.20	-0.91
sgr	-0.45	-0.17	-1.43	-1.11	-0.93	-0.37	-1.36	-1.04
betasq	-0.56	-0.19	-2.93	-1.83	-1.02	-0.33	-3.48	-2.11
currat	-0.35	-0.20	-1.39	-1.36	-1.06	-0.56	-1.94	-1.48
grcapx	-0.53	-0.23	-0.81	-0.68	-0.60	-0.25	-0.53	-0.44
hire	-0.76	-0.31	-1.21	-0.99	-0.65	-0.28	-0.86	-0.76
agr	-0.86	-0.32	-1.39	-1.03	-1.63	-0.63	-1.96	-1.45
pchquick	-0.60	-0.33	-1.12	-1.00	-0.30	-0.15	-0.57	-0.47
chcsho	-1.07	-0.41	-1.55	-1.07	-1.68	-0.65	-1.72	-1.14
age	-1.17	-0.47	0.29	0.22	-1.26	-0.48	-0.04	-0.03
mom36m	-1.12	-0.47	-1.11	-0.89	-1.02	-0.39	-0.93	-0.70
pchcurrat	-0.99	-0.53	-1.52	-1.38	-0.97	-0.47	-1.44	-1.15
operprof	-1.34	-0.63	-0.71	-0.39	-0.17	-0.11	0.65	0.56
ms	-1.12	-0.64	0.96	1.29	-0.20	-0.11	1.71	1.92
bm_ia	-1.22	-0.67	-0.85	-0.60	-1.33	-0.75	-1.20	-0.84
cashpr	-1.36	-0.72	-0.21	-0.23	-1.08	-0.51	0.35	0.32
chmom	-2.10	-0.74	-2.92	-1.13	0.68	0.29	-0.55	-0.26
herf	-1.64	-0.76	-1.54	-0.90	-0.09	-0.06	-0.02	-0.02
convind	-1.10	-0.78	-1.01	-1.20	0.77	0.50	0.59	0.58
pchsale_pchrect	-0.61	-0.80	0.44	0.64	-1.05	-1.09	-0.16	-0.18
invest	-2.25	-0.80	-2.48	-1.36	-0.52	-0.23	-0.87	-0.74
lgr	-3.24	-1.09	-3.13	-1.58	-1.35	-0.57	-1.34	-0.98
egr	-2.96	-1.16	-3.23	-1.75	-0.96	-0.51	-1.21	-1.15
chinv	-2.71	-1.18	-2.91	-1.68	-0.50	-0.32	-0.78	-0.81
grltnoa	-3.09	-1.31	-3.05	-1.75	-1.87	-1.09	-2.00	-2.00
mve_m	-3.34	-1.33	-0.94	-0.54	-1.54	-0.77	0.40	0.43
dolvol	-3.79	-1.38	-1.66	-0.90	-1.79	-0.79	-0.43	-0.43
pchcapx_ia	-1.78	-1.49	-1.56	-1.41	-1.26	-1.21	-1.02	-1.03
pctacc	-1.77	-1.56	-0.57	-0.59	-3.12	-2.45	-1.88	-1.72
mve_ia	-3.67	-1.56	-0.99	-0.62	-1.64	-0.91	0.67	0.90
acc	-3.02	-1.62	-3.04	-1.79	-3.49	-1.96	-3.31	-2.07
cfp_ia	-3.82	-1.74	-3.00	-1.67	-2.08	-1.29	-1.68	-1.28

This table present the results for univariate sorts on each of the 94 mutual fund level characteristics we consider. Fund level characteristics are calculated as the value-weighted average of the characteristics held in each fund's portfolio. For month $t + 1$, we sort mutual funds according to a fund level characteristic at month t into 10 portfolios and compute equal-weighted as well as value-weighted portfolio returns. We then go long in the $10th$ portfolio and short in the first portfolio and hold this long-short portfolio for one month. We report equal-weighted portfolio excess returns and Fama-French Charhart four-factor alphas, together with their t-statistics, computed using Newey and West (1987)'s standard errors. To improve the readability of the table, we sort fund characteristics in descending order according to the t-statistics of their equal-weighted excess returns.

Table 3: Returns of Mutual Fund Portfolios Sorted Using BRT Predictions

	BRT				LR				BEST			
	Ret	t-stat	Alpha	t-stat	Ret	t-stat	Alpha	t-stat	Ret	t-stat	Alpha	t-stat
Low	4.23	1.43	-4.59	-5.09	6.03	2.01	-2.64	-1.48	6.15	2.24	-1.58	-3.13
2	7.59	2.66	-1.34	-0.77	5.74	2.13	-2.41	-3.10	6.74	2.66	-1.08	-2.04
3	6.12	2.32	-2.18	-4.53	6.15	2.32	-1.84	-3.65	7.13	2.72	-1.01	-2.15
4	7.05	2.67	-1.13	-2.14	6.70	2.54	-1.26	-3.01	9.14	3.29	1.09	0.67
5	7.03	2.64	-1.10	-2.36	6.73	2.56	-0.80	-1.57	6.57	2.44	-1.37	-3.55
6	7.57	2.88	-0.36	-0.72	7.11	2.79	-0.51	-0.96	7.41	2.78	-0.71	-1.38
7	7.79	2.92	-0.29	-0.49	7.70	2.89	0.09	0.14	6.82	2.55	-1.06	-1.98
8	8.50	3.12	0.30	0.36	8.27	3.08	0.68	0.77	7.59	2.81	-0.30	-0.53
9	9.46	3.41	1.24	1.32	8.78	3.12	1.20	1.21	8.03	2.91	0.04	0.05
High	10.91	3.59	2.88	2.16	9.78	3.24	2.29	1.81	8.35	3.07	0.27	0.36
High-Low	6.68	3.56	7.46	3.76	3.75	1.48	4.93	1.96	2.20	2.63	1.85	2.65

This table presents annualized excess returns and Fama-French Carhart's four-factor alphas for portfolios generated using BRTs, Linear Regression (LR) and the best univariate sort (BEST). We select BEST as the characteristic with the highest t-statistics for equal-weighted portfolio's excess return, that is, sales to receivables (*salerec*). For each month t, BRT and LR predict the excess returns of each mutual fund, based on its characteristics at month $t-1$. We then sort mutual funds according to their predicted returns, form equal-weighted decile portfolios, and hold them for one month. "Low" ("High") represents the decile portfolio containing mutual funds that are expected to perform the worst (best). "High-Low" denotes the long short portfolio that goes long in the portfolio of funds with the highest expected returns and short in the portfolio with the lowest expected returns. We report both annualized excess net returns and their Fama-French Carhart's four-factor alphas, together with their t-statistics computed using Newey and West (1987)'s standard errors.

Table 4: Returns of Mutual Fund Portfolios Sorted Using BRT Predictions (Additional Specifications).

	Raw firm characteristics				Firm-characteristic ranks				Firm-characteristic deciles			
	Net Return		Gross Return		Net Return		Gross Return		Net Return		Gross Return	
	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW
Low	4.23	4.26	5.20	5.08	4.13	3.90	5.31	5.23	4.10	3.67	5.28	4.80
2	7.59	5.62	8.56	6.52	6.04	6.17	6.88	6.85	5.79	5.76	6.75	6.93
3	6.12	6.14	7.42	7.35	6.30	6.10	7.50	7.06	6.41	6.53	7.57	7.49
4	7.05	6.65	7.77	7.24	6.94	6.38	8.01	7.59	6.80	6.69	7.88	7.49
5	7.03	6.78	8.42	8.04	7.10	7.07	7.82	7.66	7.39	7.15	8.23	8.09
6	7.57	7.16	8.58	8.10	7.29	7.50	8.54	8.31	7.27	7.33	8.30	7.66
7	7.79	7.72	8.83	8.74	8.18	7.89	9.44	9.08	8.20	7.98	9.17	8.77
8	8.50	8.26	9.55	9.51	8.72	8.75	9.63	9.80	8.68	8.85	9.89	9.91
9	9.46	9.77	10.70	10.60	9.43	9.15	10.58	10.03	9.36	9.07	10.43	10.11
High	10.91	10.57	11.90	11.40	10.58	9.97	11.72	11.08	10.70	10.09	11.93	11.28
High-Low	6.68	6.32	6.71	6.35	6.45	6.07	6.42	5.84	6.60	6.43	6.65	6.48
t-stat	3.56	3.03	3.63	3.09	3.41	2.89	3.44	2.87	3.54	3.09	3.50	3.02
Alpha	7.46	6.90	7.56	6.98	7.25	6.56	7.27	6.56	7.42	7.03	7.44	7.12
t-stat	3.76	3.21	3.86	3.29	3.62	3.03	3.69	3.08	3.80	3.30	3.72	3.23
MR t-stat	3.32	2.97	3.36	2.98	3.19	2.85	3.19	2.75	3.23	3.05	3.26	2.96

This table presents annualized excess returns and Fama-French Carhart four-factor alphas for the portfolios selected according to BRT predictions using three sets of stock characteristics, i.e., raw firm characteristics, firm-characteristic ranks, and firm-characteristic deciles. "Raw firm characteristics" are the raw firm characteristics as computed in Green, Hand, and Zhang (2017); "firm-characteristic ranks" convert original firm characteristics to their percentage rank in the CRSP stock universe—see Avramov, Cheng, and Hameed (2019);"firm-characteristic deciles" convert the original firm characteristics to their deciles over the CRSP universe of stocks—see Ali et al. (2008). Mutual fund level characteristics are computed as value-weighted averages of the 94 characteristics of the stocks held. For each month t, BRTs predict mutual funds' returns (in terms of net return or gross return) based on the funds' holding characteristics in month $t-1$. We then sort mutual funds according to their predicted returns, form equal-weighted and value weighted decile portfolios, and hold the portfolios for one month. "Net return" refers to the returns net of fees, while "gross return" refers to the returns before fees. "Low" ("High") represents the decile portfolio that contains the mutual funds that are expected to perform the worst (best). "High-Low" denotes the long-short portfolio that goes long in the top decile portfolio and short in the bottom decile portfolio. We report both annualized excess returns and Fama-French Carhart four-factor alphas, together with their t-statistics computed using Newey and West (1987) standard errors. "MR t-stat" reports t-statistics testing for monotonicity in the expected returns of portfolios moving from Low to High.

Table 5: Fama-MacBeth Regression of Fund Returns on BRT predictions and other controls

	Dependent variable: Excess Return				
	(1)	(2)	(3)	(4)	(5)
$\widehat{BRT}$	0.093***	0.093***	0.098***	0.102***	
	(3.45)	(3.42)	(3.02)	(3.08)	
Lag(Excess Return)			-0.016	-0.033**	0.039**
			(-0.99)	(-2.03)	(2.19)
TR2				-0.039	-0.034
				(-0.84)	(-0.71)
Active Share				0.092	0.074
				(0.31)	(0.24)
Fund Flows		0.153	-0.822	-1.626	-1.598
		(0.38)	(-0.85)	(-1.41)	(-1.40)
log(TNA)		-0.109***	-0.118***	-0.162**	-0.144**
		(-2.87)	(-2.72)	(-2.38)	(-2.25)
$\log(\text{TNA})^2$		0.008**	0.009**	0.013**	0.012**
		(2.46)	-2.355	-2.255	-2.083
log(Fund Age)		0.014	0.017	-0.006	-0.003
		(0.89)	(1.05)	(-0.27)	(-0.12)
Expense		-0.148***	-0.152***	-0.168***	-0.168***
		(-4.03)	(-4.22)	(-4.58)	(-4.48)
Turnover		0.03	0.033	0.027	0.028
		(1.31)	(1.43)	(0.99)	(0.95)
R^2	0.144	0.196	0.208	0.274	0.248

This table presents the results of Fama-MacBeth regressions of the following form:

$$Return_{f,t} = \beta_1 \widehat{BRT}_{f,t|t-1} + \beta_2 Return_{f,t-1} + cM_{f,t-1} + \epsilon_{f,t},\ \ for\ t = 1, \ldots, T,$$

where $Return_{f,t}$ is fund f's excess return at month t (expressed in percentages); $\widehat{BRT}_{f,t|t-1}$ is BRTs' prediction for fund f's excess return at month t, conditional on information available as of time $t-1$; and $M_{f,t-1}$ is a vector of regressors available as of time $t-1$ that contain the following covariates: Lag(Excess Return) is lagged monthly excess return; TR^2 is logistic transformation of the R^2 associated with a regression of the fund's monthly excess returns on a Carhart four-factor model using an estimation period of 24 months—see Amihud and Goyenko (2013) for details; We obtain Active share from Antti Petajisto's websites (`https://activeshare.nd.edu/data/` and `https://www.petajisto.net/data.html`); Fund Flows, log(TNA), $\log(\text{TNA})^2$, log(Fund Age), Expense and Turnover represent fund flows, logged fund size, logged fund size squared, logged age of the fund, the expense ratio of the fund and the turnover of the fund. These quantities are obtained from CRSP and calculated following Avramov, Cheng, and Hameed (2019). All regressors are available from 1980 to 2018 except for Active share that is available from 1980 to 2015. R^2 shows the average R^2 over each month. *, **, and ***: Significance at the 10%, 5%, and 1% level, respectively.

Table 6: Seasonality Tests for the Performance of the Mutual Funds Selected by BRTs.

Rank	Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec
Low	0.43	0.08	0.87	0.46	0.92	0.41	0.07	-0.41	-1.06	0.40	0.66	1.41
2	0.30	0.45	1.00	0.81	0.88	0.24	0.19	-0.31	-0.82	0.42	1.04	3.37
3	0.38	0.44	1.09	0.95	0.92	0.13	0.16	-0.21	-0.77	0.45	1.18	1.44
4	0.37	0.55	1.13	1.04	1.19	0.20	0.27	-0.17	-0.72	0.42	1.38	1.41
5	0.40	0.57	1.28	1.15	1.01	-0.03	0.29	-0.10	-0.66	0.36	1.44	1.35
6	0.50	0.64	1.29	1.18	1.34	-0.06	0.20	-0.07	-0.61	0.31	1.55	1.33
7	0.48	0.86	1.31	1.18	1.17	-0.12	0.06	0.07	-0.60	0.16	1.80	1.45
8	0.38	1.08	1.36	1.25	1.27	-0.07	0.11	0.08	-0.49	0.21	1.90	1.46
9	0.52	1.25	1.35	1.36	1.47	-0.28	0.03	0.20	-0.44	0.18	2.18	1.67
High	0.55	1.41	1.43	1.61	1.78	-0.03	-0.23	0.28	-0.11	0.16	2.52	1.57
High-Low	0.11	1.32	0.56	1.15	0.86	-0.44	-0.30	0.69	0.95	-0.24	1.86	0.16
t-stat	0.28	4.61	1.58	6.07	3.52	-0.93	-0.74	2.42	4.01	-0.63	2.14	0.21
Alpha	0.14	0.83	0.18	0.77	1.19	0.49	-0.02	0.49	0.76	-0.68	2.19	-0.77
t-stat	0.35	2.73	0.52	3.30	4.76	0.71	-0.04	1.87	2.80	-2.64	2.07	-2.47

This table shows the annualized monthly excess returns and Fama-French Carhart four-factor alphas for mutual fund portfolios selected by BRTs—across the different calendar months of the year. For each month t, BRTs predict mutual funds' excess returns based on the funds' holding characteristics at month $t-1$. We then sort mutual funds according to their predicted returns, form equal-weighted decile portfolios, and hold the portfolios for one month. "Low" ("High") represents the decile portfolio containing the mutual funds that are expected to perform the worst (best). "High-Low" denotes the long-short portfolio that goes long in the top decile portfolio and short in the bottom decile portfolio. We report both annualized excess return and Fama-French Carhart four-factor alphas, together with their t-statistics computed using Newey and West (1987) standard errors.

Table 7: Selecting mutual funds with BRTs over business cycles.

	Panel A. Business Cycles			
	Recession		Expansion	
	EW	VW	EW	VW
High-Low	13.52	14.36	6.01	5.17
	(2.81)	(3.26)	(2.93)	(2.26)
Alpha	8.83	9.69	7.32	6.11
	(1.39)	(1.68)	(3.15)	(2.43)

	Panel B. Investor Sentiment			
	High		Low	
	EW	VW	EW	VW
High-Low	6.48	6.65	7.37	5.55
	(2.91)	(2.68)	(2.57)	(1.91)
Alpha	5.99	6.25	7.54	5.72
	(2.15)	(2.17)	(2.91)	(2.08)

This table presents annualized monthly excess returns and Fama-French Carhart four-factor alphas for long-short portfolios of mutual funds selected by BRTs in economic expansions and recessions. At the end of month t, we predict next month's mutual funds excess returns, and form 10 equal- and value-weighted portfolios. We then a construct long-short portfolio that goes long in the tenth portfolio and short in the first portfolio and hold it for one month. We condition the performance of our BRT procedure in recessions and expansions as determined by the National Bureau of Economic Research in Panel A and according to the Baker and Wurgler (2006)'s investor sentiment in Panel B. In this second case, we follow Baker and Wurgler (2006) and label as "High" ("Low") those periods when the sentiment index is higher (lower) than the median values. In each case, we report long-short portfolios' excess returns, their Fama-French Carhart four-factor alphas and their t-statistics, computed using Newey and West (1987)'s standard errors. All results are calculated based on net returns. Results based on gross returns are similar.

Table 8: Selecting mutual funds controlling for information arrival.

Delayed Months	Panel A: Fund-level characteristics delayed by 1 to 12 months.												
	0	1	2	3	4	5	6	7	8	9	10	11	12
Ret.	6.68	6.55	6.96	7.19	6.52	6.68	6.71	6.23	6.38	5.81	6.29	6.15	6.22
t-stat	3.56	3.57	3.76	3.88	3.50	3.73	3.82	3.49	3.38	3.16	3.48	3.43	3.51
Alpha	7.46	7.09	7.57	7.64	7.11	7.07	7.14	6.61	7.05	6.54	6.85	6.81	7.03
t-stat	3.76	3.58	3.79	3.86	3.48	3.65	3.70	3.45	3.48	3.31	3.60	3.41	3.71
Delayed Months	Panel B: Fund holding delayed by 1 to 12 months.												
	0	1	2	3	4	5	6	7	8	9	10	11	12
Ret.	6.68	6.69	6.70	6.44	6.62	6.60	6.42	6.15	6.27	6.07	6.18	6.22	6.24
t-stat	3.56	3.48	3.62	3.45	3.69	3.57	3.49	3.38	3.43	3.41	3.36	3.36	3.40
Alpha	7.46	7.35	7.33	7.07	7.29	7.25	7.20	6.86	6.98	6.81	6.82	6.93	7.04
t-stat	3.76	3.57	3.70	3.46	3.75	3.60	3.63	3.47	3.50	3.53	3.48	3.40	3.52

This table presents annualized monthly excess returns and Fama-French Carhart four-factor alphas for long-short portfolios selected using BRTs, after controlling for information arrival. Panel A shows the results of delaying the fund-level characteristics by one to twelve months, and Panel B shows the results of delaying the fund holding information by one to twelve months. For each month t, BRT estimates mutual funds' performance based on the funds' holding characteristics in month $t-1$ to $t-13$, respectively. We then sort mutual funds according to their predicted excess returns, form equal-weighted long short portfolios, and hold them for one month. All reported results are based on equal-weighted excess net returns, that is, returns after fees. Results based on value-weighted and gross returns are similar.

Table 9: Selecting mutual funds with rolling windows of different size.

Panel A: Net Return (% per year)

Window Size	Equal-Weighted				Value-Weighted			
	Ret.	t-stat	Alpha	t-stat	Ret.	t-stat	Alpha	t-stat
1	6.68	3.56	7.46	3.76	6.32	3.03	6.90	3.21
6	6.18	3.12	5.47	3.15	5.86	2.87	5.14	2.75
12	5.06	2.77	3.85	2.61	4.98	2.64	3.58	2.19
36	5.83	3.10	4.37	3.01	5.31	2.95	3.63	2.52
60	5.70	3.03	4.16	3.10	5.39	2.81	3.56	2.38
Recursive	5.42	3.18	3.34	2.73	5.18	3.21	2.93	2.48

Panel B: Gross Return (% per year)

Window Size	Equal-Weighted				Value-Weighted			
	Ret.	t-stat	Alpha	t-stat	Ret.	t-stat	Alpha	t-stat
1	6.71	3.63	7.56	3.87	6.35	3.09	6.98	3.29
6	6.14	3.16	5.47	3.19	6.00	3.02	5.22	2.88
12	4.87	2.75	3.72	2.57	5.31	2.98	3.85	2.52
36	5.49	2.79	4.36	2.90	5.47	2.83	4.04	2.54
60	5.73	3.05	4.23	3.10	5.22	2.78	3.57	2.48
Recursive	4.98	2.91	2.99	2.45	4.93	2.96	2.65	2.15

This table presents annualized monthly excess returns and Fama-French Carhart four-factor alphas for the long short portfolios generated by BRTs trained using rolling windows of different size. We estimate the BRT model using the characteristics and excess return data from month $t-w$ to $t-1$, and predict the excess returns for month t. We then sort the mutual funds by their return predictions, form equal-weighted and value weighted decile portfolios, and hold the portfolios for one month. We report both annualized excess return and their Carhart (1997)'s adjusted alpha, together with their t-statistics, computed using Newey and West (1987)'s standard errors. All results are calculated based on net returns.

Table 10: Fund characteristics across BRT selected fund deciles.

	Age (months)	TNA ($ Million)	No. of Stocks	Expense	Turnover	Conc. Index	ICI
D1	145	691.09	86	0.0116	0.8725	0.0274	0.0647
D2	152	816.24	101	0.0109	0.8203	0.0237	0.0493
D3	155	873.22	104	0.0108	0.7969	0.0233	0.0446
D4	155	904.85	105	0.0106	0.7834	0.0231	0.0425
D5	159	960.33	106	0.0106	0.7847	0.0231	0.0414
D6	157	934.21	106	0.0106	0.7937	0.0230	0.0422
D7	158	941.11	107	0.0106	0.8079	0.0228	0.0431
D8	156	892.37	107	0.0108	0.8200	0.0233	0.0464
D9	153	880.93	105	0.0109	0.8520	0.0234	0.0514
D10	149	799.42	92	0.0116	0.9170	0.0268	0.0666
D(10+1)-D(5+6)	-11	-202.01	-17	0.0010	0.1055	0.0040	0.0238
t-stat	-6.02	-5.43	-8.22	14.70	7.24	14.37	13.64

This table reports the average values of fund characteristics for the decile portfolios selected by BRTs. For each month t, BRTs predict mutual funds' excess returns based on the funds' holding characteristics in month $t-1$. We then sort mutual funds according to their predicted returns, form equal-weighted decile portfolios, and hold them for one month. D1 (D10) represents the decile portfolio that contains the mutual funds that are expected to perform the worst (best). Age (months), Total Net Assets ($ Million), No. of Stocks, Expense, and Turnover are calculated as average values for the decile portfolios in each month, and then averaged over all months. The Concentration Index, also known as the Herfindahl Index, is calculated as the sum of squared portfolio weights of individual stock holdings. The Industry Concentration Index (ICI) represents portfolio concentration across industries, see Kacperczyk, Sialm, and Zheng (2005) for details.

Table 11: Mutual fund styles selected by BRTs.

Investment Objectives	Full Dataset # of Funds	Full Dataset Percentage	D1 Percentage	D10 Percentage
Equity-Domestic-Cap-based-Micro Cap	37	1.24%	2.07%	2.31%
Equity-Domestic-Cap-based-Mid Cap	407	13.66%	11.56%	13.07%
Equity-Domestic-Cap-based-Small Cap	608	20.40%	23.87%	25.67%
Equity-Domestic-Style-Growth and Income	452	15.17%	13.02%	12.00%
Equity-Domestic-Style-Growth	1322	44.36%	45.46%	43.72%
Equity-Domestic-Style-Income	138	4.63%	3.52%	2.98%
Others	16	0.54%	0.50%	0.25%
Sum	2980	100.00%	100.00%	100.00%

This table compares the investment objectives of the funds selected by BRTs to the ones in the full dataset we use. In each case, we calculate mutual funds' investment objectives for each month and report their averaged distribution over the full sample. D1 (D10) denotes the decile portfolio containing the mutual funds expected to perform the worst (best).

Table 12: Summary of the most important fund level characteristics
and their categories.

Panel A: Top 10 Firm Characteristics and Their Categories

Acronym	Firm Characteristics	Category
mom12m	12-month momentum	Momentum
mve_ia	Industry-adjusted size	Trading Frictions
dolvol	Dollar trading volume	Trading Frictions
beta	Beta	Trading Frictions
mve_m	Size	Trading Frictions
turn	Share turnover	Trading Frictions
mom6m	6-month momentum	Momentum
baspread	Bid-ask spread	Trading Frictions
mom36m	36-month momentum	Momentum
idiovol	Idiosyncratic return volatility	Trading Frictions

Panel B: Relative Proportions of Firm Characteristics

Category	Proportion in the 94 Characteristics	Proportion in Top 10
Trading Frictions	14.89%	70.00%
Momentum	11.70%	30.00%
Intangible	29.79%	0.00%
Investments	18.09%	0.00%
Profitability	17.02%	0.00%
Value Versus Growth	8.51%	0.00%

This table reports the most important fund level characteristics (top 10) in predicting fund returns. Panel A ranks firm characteristics according to their relative importance. Panel B compares the proportion of characteristics in each category on the full set of 94 characteristics and the top 10 most important characteristics. We assign each firm characteristic into one of the six categories according to Hou, Xue, and Zhang (2020): *Trading Frictions*, *Momentum*, *Intangible*, *Investments*, *Profitability* and *Value versus Growth*.

Table 13: Dissecting BRT Predictions using Fama-MacBeth Regressions

	Dependent variable: *BRT Prediction*	
	(1)	(2)
Lag(Excess Return)	0.607***	0.597***
	(33.21)	(34.17)
Fund Flows		−0.316
		(−0.70)
log(TNA)		−0.062***
		(−3.31)
log(TNA)^2		0.004***
		(2.63)
log(Fund Age)		0.039**
		(2.34)
Expense		−0.010
		(−0.21)
Turnover		0.006
		(0.72)
R^2	0.514	0.508

This table presents the results of Fama-MacBeth regressions of the following form:

$$\widehat{BRT}_{f,t|t-1} = c\ M_{f,t-1} + \epsilon_{f,t},\ \ for\ t = 1, \ldots, T,$$

where $\widehat{BRT}_{f,t|t-1}$ is BRT's prediction for fund f's excess return on month t, conditional on information available as of time $t-1$; and $M_{f,t-1}$ is a vector of regressors available as of time $t-1$ and contains the following covariates: $Lag(Excess\ Return)$ is the monthly lagged excess return of the fund; $Fund\ Flows$, $log(TNA)$, $log(TNA)^2$, $log(Fund\ Age)$, $Expense$ and $Turnover$ represent fund flows, logged fund size, logged fund size squared, logged age of the fund, fund's expense ratio and fund's turnover. These quantities are obtained from CRSP and calculated following Avramov, Cheng, and Hameed (2019). All regressors are available from 1980 to 2018. *, **, and ***: Significance at the 10%, 5%, and 1% level, respectively.

Table 14: Selecting mutual funds with other machine learning methods.

Rank	LASSO	ElasticNet	RF	NN1	NN2	NN3	NN4	NN5	BRTs
Low	5.54	5.24	5.53	6.23	4.04	4.07	4.21	4.24	4.23
2	5.34	5.25	5.65	5.63	7.61	7.55	7.55	7.23	7.59
3	6.15	6.50	6.35	6.59	6.51	6.29	6.69	6.67	6.12
4	6.83	6.83	6.65	7.08	6.72	6.75	6.78	6.84	7.05
5	7.14	6.94	6.85	7.15	7.51	7.24	7.08	6.98	7.03
6	7.54	7.66	7.62	7.76	7.57	7.61	7.83	7.81	7.57
7	8.22	8.29	8.12	7.83	8.20	8.26	7.89	8.08	7.79
8	8.81	8.84	8.97	8.24	8.45	8.43	8.35	8.64	8.50
9	9.62	9.70	9.66	9.38	9.15	9.37	9.24	9.39	9.46
High	11.07	11.01	10.85	10.32	10.47	10.65	10.58	10.33	10.91
High-Low	5.52	5.77	5.33	4.08	6.43	6.58	6.37	6.09	6.68
t-stat	2.13	2.22	2.00	1.73	3.50	3.80	3.45	3.34	3.56
Alpha	6.30	6.59	6.51	5.15	7.09	7.43	7.08	6.86	7.46
t-stat	2.46	2.56	2.42	2.24	3.78	4.18	3.67	3.58	3.76

This table presents the annualized excess returns and their Fama-French Carhart's four-factor alphas, net of fees, for equal-weighted portfolios selected by BRTs and other machine learning methods, such as lasso, elastic net, random forests, and neural networks with one to five hidden layers (NN1 to NN5) with hyper-parameters cross-validated recursively. For each month t, we use all machine learning methods to predict the performance of each mutual fund, based on its characteristics in month $t-1$. We then sort mutual funds according to their predicted returns, form equal-weighted decile portfolios, and hold them for one month. Low (High) represents the decile portfolio that contains mutual funds that are expected to perform the worst (best). "High-Low" denotes the long short portfolio that goes long in the portfolio of funds with the highest expected returns and short in the portfolio with the lowest expected returns. We report both annualized excess returns and their Fama-French Carhart's four-factor alphas, together with their t-statistics computed using Newey and West (1987)'s standard errors. Results using value-weighed and gross returns are similar.

Online Appendix:

Selecting Mutual Funds from the Stocks they Hold: A Machine Learning Approach

Bin Li
Wuhan University

Alberto G. Rossi
Georgetown University

Not for Publication

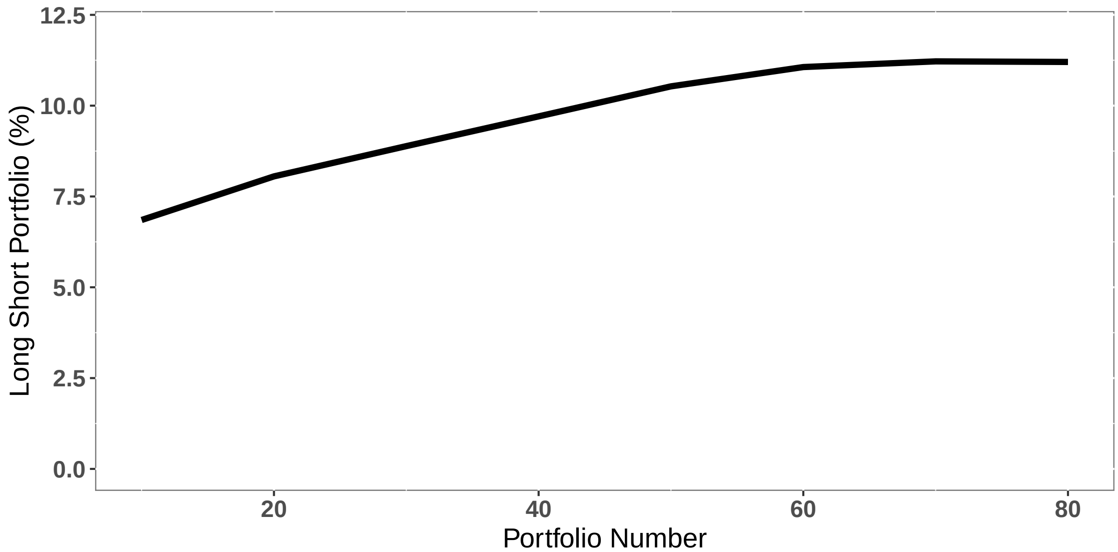


Figure A.1: **Equal-weighted long short portfolio with varying number of portfolios.** This figure presents the excess return for long short portfolios selected using BRTs, when we allow for different number of portfolios. At the beginning of each month t, BRT estimates mutual funds' performance, based on mutual funds' holding characteristics in month $t-1$. We then sort the mutual funds according to their performance predictions, form equal-weighted portfolios and hold them for one month. We allow the number of portfolios to range from 10 to 80 and report annualized monthly excess returns. The sample data ranges from 1980 to 2018.

Table A.1: Firm characteristics and their abbreviations.

No.	Acronym	Firm Characteristics	No.	Acronym	Firm Characteristics
1	absacc	Absolute accruals	48	mom36m	36-month momentum
2	acc	Working capital accruals	49	mom6m	6-month momentum
3	aeavol	Abnormal earnings announcement volume	50	ms	Financial statement score
4	age	# years since first Compustat coverage	51	mve_ia	Industry-adjusted size
5	agr	Asset growth	52	mve_m	Size
6	baspread	Bid-ask spread	53	nincr	Number of earnings increases
7	beta	Beta	54	operprof	Operating profitability
8	betasq	Beta squared	55	orgcap	Organizational capital
9	bm	Book-to-market	56	pchcapx_ia	Industry adj. % Δ in CAP expenditures
10	bm_ia	Industry-adjusted book to market	57	pchcurrat	% change in current ratio
11	cash	Cash holdings	58	pchdepr	% change in depreciation
12	cashdebt	Cash flow to debt	59	pchgm_pchsale	% Δ in gross margin - % change in sales
13	cashpr	Cash productivity	60	pchquick	% change in quick ratio
14	cfp	Cash flow to price ratio	61	pchsale_pchinvt	% change in sales - % change in inventory
15	cfp_ia	Industry-adjusted cash flow to price ratio	62	pchsale_pchrect	% change in sales - % change in A/R
16	chatoia	Industry-adjusted change in asset turnover	63	pchsale_pchxsga	% change in sales - % change in SG&A
17	chcsho	Change in shares outstanding	64	pchsaleinv	% change sales-to-inventory
18	chempia	Industry-adjusted change in employees	65	pctacc	Percent accruals
19	chinv	Change in inventory	66	pricedelay	Price delay
20	chmom	Change in 6-month momentum	67	ps	Financial statements score
21	chpmia	Industry-adjusted change in profit margin	68	quick	Quick ratio
22	chtx	Change in tax expense	69	rd	R&D increase
23	cinvest	Corporate investment	70	rd_mve	R&D to market capitalization
24	convind	Convertible debt indicator	71	rd_sale	R&D to sales
25	currat	Current ratio	72	realestate	Real estate holdings
26	depr	Depreciation / PP&E	73	retvol	Return volatility
27	divi	Dividend initiation	74	roaq	Return on assets
28	divo	Dividend omission	75	roavol	Earnings volatility
29	dolvol	Dollar trading volume	76	roeq	Return on equity
30	dy	Dividend to price	77	roic	Return on invested capital
31	ear	Earnings announcement return	78	rsup	Revenue surprise
32	egr	Growth in common shareholder equity	79	salecash	Sales to cash
33	ep	Earnings to price	80	saleinv	Sales to inventory
34	gma	Gross profitability	81	salerec	Sales to receivables
35	grcapx	Growth in capital expenditures	82	secured	Secured debt
36	grltnoa	Growth in long term net operating assets	83	securedind	Secured debt indicator
37	herf	Industry sales concentration	84	sgr	Sales growth
38	hire	Employee growth rate	85	sin	Sin stocks
39	idiovol	Idiosyncratic return volatility	86	sp	Sales to price
40	ill	Illiquidity	87	std_dolvol	Vol of liquidity (dollar trading volume)
41	indmom	Industry momentum	88	std_turn	Vol of liquidity (share turnover)
42	invest	Capital expenditures and inventory	89	stdacc	Accrual volatility
43	lev	Leverage	90	stdcf	Cash flow volatility
44	lgr	Growth in long-term debt	91	tang	Debt capacity/firm tangibility
45	maxret	Maximum daily return	92	tb	Tax income to book income
46	mom12m	12-month momentum	93	turn	Share turnover
47	mom1m	1-month momentum	94	zerotrade	Zero trading days

This table reports the acronyms and names of the 94 stock characteristics we use in our analysis.

Table A.2: Selecting mutual funds conditioning on fund styles.

Rank	Equal-Weighted			Value-Weighted		
	Growth	Growth and Income	Small- and Mid-Cap	Growth	Growth and Income	Small- and Mid-Cap
Low	3.90	3.77	4.92	4.08	4.02	3.95
2	5.28	5.13	6.13	5.52	5.61	5.47
3	5.75	5.87	11.58	6.42	5.75	5.99
4	6.42	6.20	6.95	6.68	6.29	6.97
5	7.02	7.20	7.74	7.16	7.31	8.53
6	7.78	7.25	9.19	7.99	7.62	9.18
7	7.73	7.73	9.58	8.04	7.65	8.47
8	8.14	7.94	10.56	7.86	7.97	10.57
9	9.00	9.01	10.03	9.17	9.04	10.09
High	9.78	10.11	11.65	9.50	9.79	11.58
High-Low	5.88	6.34	6.73	5.42	5.77	7.63
t-stat	3.50	3.76	3.30	2.93	3.22	3.43
Alpha	6.57	7.06	7.32	6.12	6.22	7.98
t-stat	3.57	3.77	3.64	3.18	3.30	3.68

This table presents the excess return and Carhart (1997) adjusted alphas for the portfolios generated using BRTs, conditioning on fund styles. We first construct three fund categories. The first category is "Growth" and includes funds categorized as "Equity-Domestic-Style-Growth" in CRSP—for a total of 1,322 funds over our sample. The second category is "Growth and Income" and includes funds categorized as "Equity-Domestic-Style-Growth," "Growth and Income," and "Income" in CRSP—for a total of 1,912 funds. The third category is "Small and Mid Cap" and includes funds categorized as "Equity-Domestic-Cap-based-Micro Cap," "Small Cap," and "Mid Cap"—for a total of 1,052 funds. Within each fund category and for each month t, BRTs predict the performance of each mutual fund, based on its characteristics at month $t-1$. We then sort mutual funds according to their predicted returns, form decile portfolios within each category, and hold them for one month. Low (High) represents the decile portfolio that contains mutual funds that are expected to perform the worst (best). "High-Low" denotes the long short portfolio that goes long in the portfolio of funds with the highest expected returns and short in the portfolio with the lowest expected returns. We report both annualized excess returns and their Fama-French Carhart's four-factor alphas, together with their t-statistics computed using Newey and West (1987)'s standard errors. The results in the first three columns are computed using equal-weighted portfolios while the results in the last three columns use value-weighted portfolios. All results are based on net returns. The results on gross returns are similiar.

Table A.3: Selecting mutual funds with BRTs at the quarterly frequency.

	Raw firm characteristics				Firm-characteristics ranks				Firm-characteristics deciles			
	Net Return		Gross Return		Net Return		Gross Return		Net Return		Gross Return	
	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW
Low	5.43	4.94	6.58	6.20	5.61	4.83	6.72	5.80	5.42	4.88	6.44	5.52
2	5.90	5.99	7.17	6.88	5.87	5.72	6.82	6.79	5.95	5.74	7.12	7.24
3	6.49	6.36	6.92	6.67	6.19	6.29	7.34	7.49	6.33	6.42	7.29	7.34
4	6.47	6.62	7.97	7.80	6.59	6.52	7.58	7.32	6.33	6.48	7.44	7.54
5	6.47	6.49	7.90	8.00	6.68	6.79	7.91	7.77	6.84	6.88	7.90	7.71
6	7.16	7.02	7.90	6.90	7.54	7.00	8.31	7.74	7.46	7.52	8.33	8.17
7	7.83	7.56	8.89	8.69	7.55	7.20	8.66	8.54	7.89	7.55	9.17	8.63
8	8.68	7.73	9.68	8.83	8.29	7.99	9.40	8.49	8.10	7.22	9.14	8.20
9	8.76	8.42	9.93	9.70	8.63	8.24	10.03	9.49	8.70	8.33	9.93	9.27
High	9.51	9.31	10.64	10.30	9.78	9.38	10.80	10.81	9.73	9.40	10.84	10.45
High-Low	4.08	4.37	4.05	4.10	4.17	4.55	4.08	5.01	4.32	4.53	4.40	4.94
t-stat	2.18	2.52	2.10	2.37	2.30	2.97	2.12	3.05	2.29	2.74	2.36	2.82
Alpha	2.48	2.93	2.41	2.68	2.65	3.17	2.60	3.63	2.91	3.11	2.98	3.61
t-stat	1.64	2.09	1.54	1.92	1.88	2.63	1.74	2.71	1.92	2.21	2.06	2.58

This table presents performance of the mutual funds selected by BRTs at the quarterly frequency. At the end of quarter t, we predict mutual fund's next quarter returns, form equal- and value- weighted long short portfolios, and hold them for one quarter. We report results for three sets of firm characteristics, i.e., raw, ranks over the CRSP universe, and deciles over the CRSP universe—see the caption of Table 4 for details. The three panels show equal-weighted (EW) and value-weighted (VW) portfolio returns and their Fama-French Carhart's four-factor alpha, together with their corresponding t-statistics computed using Newey and West (1987)'s standard errors. Excess returns are computed both gross and net of fees.

Table A.4: Selecting mutual funds using BRTs with different boosting iterations.

# of iterations	Equal-Weighted				Value-Weighted			
	Ret	t-stat	Alpha	t-stat	Ret	t-stat	Alpha	t-stat
10	4.77	1.77	5.76	2.13	6.06	2.88	6.60	2.95
20	5.11	1.93	6.15	2.32	6.60	3.06	7.15	3.18
30	6.92	3.58	7.70	3.75	6.53	3.16	7.02	3.25
40	6.83	3.51	7.64	3.69	6.69	3.12	7.25	3.24
50	6.87	3.53	7.65	3.72	6.51	3.06	7.11	3.20
100	6.68	3.56	7.46	3.76	6.32	3.03	6.90	3.21
250	6.67	3.62	7.43	3.79	5.99	2.90	6.50	3.02
500	6.59	3.61	7.36	3.81	5.96	2.93	6.58	3.11
1000	6.48	3.60	7.22	3.74	5.88	2.92	6.52	3.10
2500	6.51	3.63	7.24	3.76	5.94	2.95	6.57	3.13
5000	6.51	3.63	7.24	3.76	5.95	2.96	6.58	3.13
7500	6.51	3.63	7.24	3.76	5.95	2.96	6.58	3.13
10000	6.51	3.63	7.24	3.76	5.95	2.96	6.58	3.13

This table presents the performance of long-short portfolios computed using BRTs with different numbers of boosting iterations. We report equal-weighted and value-weighted portfolio returns and their Fama-French Carhart's four-factor alpha, together with their corresponding t-statistics computed using Newey and West (1987)'s standard errors. All results are based on net returns. Results based on gross returns are similar.

Table A.5: Selecting mutual funds with varying number of portfolios.

Portfolio Number	Equal Weighted				Value Weighted			
	Ret	t-stat	Alpha	t-stat	Ret	t-stat	Alpha	t-stat
10	6.68	3.56	7.46	3.76	6.32	3.03	6.90	3.21
20	8.05	3.65	8.98	3.83	7.19	2.97	8.07	3.19
30	8.89	3.68	9.95	3.90	8.37	3.12	9.49	3.45
40	9.70	3.88	10.86	4.08	9.05	3.31	10.37	3.70
50	10.53	3.97	11.55	3.99	9.53	3.32	10.89	3.63
60	11.06	4.06	12.18	4.09	10.13	3.44	11.53	3.79
70	11.22	4.04	12.42	3.99	9.95	3.31	11.32	3.61
80	11.20	4.00	12.41	3.88	9.83	3.32	11.11	3.57

This table presents the performance of equal-weighted long short portfolios selected by BRTs, varying the number of portfolios we form. At the beginning of each month t, BRTs estimates mutual funds' performance, based on mutual funds' holding characteristics in month $t-1$. We then sort mutual funds according to their expected returns, form equal-weighted long short portfolios with varying portfolio numbers, and hold them for one month. We report equal-weighted and value-weighted portfolio returns and their Fama-French Carhart's four-factor alpha, together with their corresponding t-statistics computed using Newey and West (1987)'s standard errors. All results are based on net returns. Results based on gross returns are similar.