# Machine learning and the cross-section of emerging market stock returns

Latest version here

Matthias X. Hanauer
TUM School of Management
Technical University of Munich

Tobias Kalsbach
TUM School of Management
Technical University of Munich

This draft: January 2023

**Abstract**

This paper compares various machine learning models to predict the cross-section of emerging market stock returns. We document that allowing for non-linearities and interactions leads to economically and statistically superior out-of-sample returns compared to traditional linear models. Although we find that both linear and machine learning models show higher predictability for stocks associated with higher limits to arbitrage, we also show that this effect is less pronounced for non-linear models. Furthermore, significant net returns can be achieved when accounting for transaction costs, short-selling constraints, and limiting our investment universe to big stocks only.

**Keywords**: Machine Learning, Return Prediction, Cross-Section of Stock Returns, Emerging Markets, Random Forest, Gradient Boosting, Neural Networks

**JEL Classifications:** C14, C52, C58, G11, G12, G14, G15, G17

We thank David Blitz, Nicole Branger, Clint Howard, Christoph Kaserer, Tim Kroencke, Markus Leippold, Tizian Otto, Michael Weber, Steffen Windmüller, Adam Zaremba, and seminar participants at the Munich Finance Day 2022, TUM School of Management, and Robeco for their helpful comments and suggestions. Furthermore, we are grateful for the provision of code to compute the efficient discrete generalized estimator (EDGE) of the bid-ask spread estimator by Emanuele Guidotti. Any remaining errors are our own. Disclosures: Hanauer is also employed by Robeco. The views expressed in this paper are those of the authors and not necessarily shared by Robeco.
Email: matthias.hanauer@tum.de, tobias.kalsbach@tum.de

# 1 Introduction

Machine learning algorithms have been available for a long time. However, due to increased computing power and data availability, decreased data storage costs, and algorithmic innovations in recent years (cf., Rasekhschaffe and Jones, 2019), machine learning methods see increasing popularity in research fields such as economics, finance, and accounting.[1]

This paper compares various machine learning models to predict the cross-section of emerging market stock returns. More specifically, we analyze the predictive power of nine algorithms: ordinary least squares regression and elastic net as examples for traditional linear models; tree-based models such gradient boosted regression trees and random forest; and neural networks with one to five layers. Furthermore, we investigate the performance of an ensemble comprising the five different neural networks and an ensemble of methods that allow for non-linearities and interactions, i.e., the two tree-based models and the ensemble of neural networks. In the remainder of the paper, we often use the term 'machine learning' only for the two tree-based methods, the neural networks, and the two ensembles. Our data set contains stocks from 32 emerging market countries and the 36 firm-level characteristics from Kelly et al. (2019) and Windmüller (2022) falling into categories such as value, past returns, investment, profitability, intangibles, and trading frictions. The data sample covers the sample period from July 1995 to December 2021, while our 20-year out-of-sample period is from January 2002 to December 2021.

Our main findings can be summarized as follows. First, we document that the different

[1]For instance, machine learning methods are applied to predict stock returns in Moritz and Zimmermann (2016), Rasekhschaffe and Jones (2019), Freyberger et al. (2020), Gu et al. (2020), Tobek and Hronec (2020), Chen et al. (2021), Drobetz and Otto (2021), Leippold et al. (2022), Azevedo et al. (2022), Cakici et al. (2022a), and Rubesam (2022), stock market betas in Drobetz et al. (2021), country stock returns in Cakici and Zaremba (2022), industry stock returns in Rapach et al. (2019), option returns in Bali et al. (2022a), corporate bond returns in Kaufmann et al. (2021) and Bali et al. (2022b), the equity premium in Rossi (2018), Treasury bond returns in Bianchi et al. (2021b) and Bianchi et al. (2021a), commodity returns in Struck and Cheng (2020), short-term bitcoin returns in Jaquart et al. (2021), cryptocurrency returns in Cakici et al. (2022b), (changes) in future company profitability in Anand et al. (2019), Van Binsbergen et al. (2020) and Chen et al. (2022), peer-implied market capitalizations in Hanauer et al. (2022), mutual fund selection in Kaniel et al. (2022), hedge fund selection in Wu et al. (2021), mortgage risk in Sadhwani et al. (2021), or corporate directors in Erel et al. (2021).

prediction algorithms pick up similar characteristics. However, we observe that tree-based methods and neural networks also identify non-linearities and interactions of characteristics. In contrast, linear methods are restricted to linear relationships and do not allow for interactions among characteristics.

Second, return forecasts based on machine learning models lead to economically and statistically superior out-of-sample long-short returns compared to traditional linear models. Furthermore, the Fama and French (2018) six-factor model can only partly explain these long-short returns, and their alphas remain highly significant. These findings are robust to several methodological choices and for emerging market subregions. Finally, we document that machine learning forecasts beat linear models consistently over our sample period, and we cannot observe a decline in predictability over time.

Third, developed market long-short returns based on machine learning forecasts derived in the same way as their emerging market counterparts cannot explain emerging market out-of-sample returns. However, models estimated solely on developed markets data also predict emerging market stock returns. These findings indicate that similar relationships between firm characteristics and future stock returns exist for developed and emerging markets but that the pricing of these characteristics is not fully integrated between developed and emerging markets.

Fourth, the high returns of the machine learning strategies in emerging markets do not primarily stem from higher-risk months and do not revert quickly, suggesting that an under-reaction explanation is more likely than a risk-based explanation. Furthermore, both linear and machine learning models show higher predictability for stocks associated with higher limits to arbitrage. However, we also document that this effect is less pronounced for machine learning forecasts than for linear regression forecasts, indicating that the superiority of machine learning models in emerging markets does not stem from limits to arbitrage.

Finally, accounting for transaction costs, short-selling constraints, and limiting our investment universe to big stocks only, we document that machine learning-based return forecasts can lead to a significant net outperformance over the market and net alphas, at least when efficient trading rules are applied.

This paper contributes to the literature in at least three aspects. First, we contribute to the rapidly expanding literature on predicting the cross-section of stock returns with machine learning methods. Rasekhschaffe and Jones (2019), Freyberger et al. (2020), and Gu et al. (2020) document that more complex machine learning models are superior to linear models for the U.S. Tobek and Hronec (2020) and Drobetz and Otto (2021) find similar evidence for developed markets and Europe, respectively. However, none of the studies mentioned above investigates emerging markets. Emerging markets are important as they account for around 58% of the global gross domestic product (GDP), which is forecasted to rise to 61% by 2026.[2] Furthermore, under the hypothesis that developed markets are integrated, the same risk factors should apply to these markets. Therefore, similar results within developed markets are not surprising, and emerging markets provide an attractive alternative for out-of-sample tests in terms of independent and new samples.

Two contemporaneously written papers, Azevedo et al. (2022) and Cakici et al. (2022a), also include emerging markets in their analysis next to developed markets. While Azevedo et al. (2022) also find that most machine learning models outperform a linear combination of anomalies, their results do not discriminate between emerging and other markets. Therefore, their results are mainly driven by developed markets. In contrast to our study, Cakici et al. (2022a) do not find superior forecasts for machine learning models compared to linear models. A potential reason for this difference might be that they train their models for each country separately while we train our models on a pooled sample of countries. However, more data might be necessary for more complex models to robustly identify non-linearities and interactions in the data.[3] We provide some supportive evidence for this claim by documenting

[2]See, IMF, World Economic Outlook database, April 2022, `https://www.imf.org/en/Publications/WEO/weo-database/2022/April`.

[3]While a linear model asks for a single parameter for each predictor, in the case of non-linear models, the number of parameters to estimate rapidly expands even with a moderate number of predictors (cf., Gu et al., 2020; Hanauer et al., 2022). As such, pooling data across countries will arguably improve the observations-to-parameters ratio.

that models trained on emerging market subregions underperform models that are trained on the pooled sample of emerging market subregions and that the performance loss is more pronounced for machine learning models and smaller subregions. Finally, Leippold et al. (2022) show that machine learning models dominate linear models for Chinese A-shares. In contrast, our sample purposely excludes Chinese A-shares to represent an international investor's investable emerging market universe: for the majority of our sample period, the China A-share market was only accessible to local investors and only gradually opened up to international investors (cf., Jansen et al., 2021).

Second, we add to the literature on the drivers of emerging market stock returns. Bekaert and Harvey (1995) and Harvey (1995) were among the first to investigate emerging market country returns and their market integration. First studies on the cross-section of emerging market stocks, such as Rouwenhorst (1999), van der Hart et al. (2003), van der Hart et al. (2005), Griffin et al. (2010), Cakici et al. (2013), and Hanauer and Linhart (2015) mainly focus on size, value, and momentum. Later studies such as Zaremba and Czapkiewicz (2017) and Hanauer and Lauterbach (2019) also investigate firm characteristics belonging to categories such as profitability, investment, intangibles, and trading frictions. Our study includes characteristics from all these groups, but machine learning models can also take non-linearities and interactions into account next to linear relationships.

Finally, our paper also contributes to the understanding of the source of return predictability from machine learning forecasts. Avramov et al. (2022) show that return forecasts from deep learning models for the U.S. extract their profitability mainly from difficult-to-arbitrage stocks and during high limits to arbitrage market states. The authors also argue that the performance of machine learning forecasts further deteriorates when microcaps are excluded and when reasonable transaction costs are considered. Similarly, Leung et al. (2021) find that the economic gains of a gradient boosting machine model for developed market stocks tend to be more limited and critically dependent on the ability to take risk and implement trades efficiently. Furthermore, Cakici et al. (2022a) document that machine learning strategies work best for small stocks, as well as in countries with many listed firms and high idiosyncratic risk. In our paper, we follow Hou et al. (2020) and exclude microcaps from our analysis. While we also find that both linear and machine learning models show higher predictability for stocks associated with higher limits to arbitrage, we also document that this effect is less pronounced for machine learning models. Furthermore, we also provide evidence that a positive and significant outperformance and six-factor alpha can be achieved even when accounting for transaction costs, short-selling constraints, and limiting the investment universe to big stocks only.

The remainder of the paper is structured as follows: Section 2 describes the data sources, sample composition, and utilized firm-level characteristics. Section 3 outlines our methodology for predicting returns with machine learning algorithms, portfolio construction, and benchmark models. Section 4 presents evidence of the superiority of more complex machine learning models, while Section 5 strives to understand the source of this superiority better. We provide our conclusions in Section 6.

# 2 Data

## 2.1 Stock market data

Our sample comprises data from emerging stock markets as classified by Morgan Stanley Capital International (MSCI). The accounting data is from Refintiv Worldscope, and the stock market data is from Refintiv Datastream. The sample period starts in July 1990 and ends in December 2021. Countries are part of the sample only in those years in which they are included in the MSCI Emerging Markets Index.[4] Furthermore, the countries are only part of the final sample in those months for which at least 10 stock-month observations are available after screens. The following 32 countries fulfill these criteria: Argentina, Brazil, Chile, China, Colombia, Czechia, Egypt, Greece, Hungary, India, Indonesia, Israel, Jordan,

[4]See https://www.msci.com/market-classification for details.

Korea, Kuwait, Malaysia, Mexico, Morocco, Pakistan, Peru, Philippines, Poland, Portugal, Qatar, Russia, Saudi Arabia, South Africa, Sri Lanka, Taiwan, Thailand, Turkey, and the UAE.[5]

Several static and dynamic screens are applied to ensure that our sample comprises exclusively of common stocks and provides the highest data quality. First, stocks are identified using Refinitiv Datastream constituent lists, particularly Refinitiv Worldscope lists, research lists, and - to eliminate survivorship bias - dead lists. Following Ince and Porter (2006), Griffin et al. (2010), Schmidt et al. (2019), and Hanauer (2020), non-common equity stocks are eliminated through generic and country-specific static screens. Furthermore, several dynamic screens are applied to stock returns and prices to exclude erroneous and illiquid observations. Appendix A provides a detailed description of the utilized constituent lists and the associated static and dynamic screens. Furthermore, we require stocks to have market capitalization data for the previous month.

We follow the size group methodology of Fama and French (2008, 2012, 2017) and Hanauer and Lauterbach (2019) and assign stocks into three size groups (micro, small, and big) separately for each country and month. Big stocks are defined as the biggest stocks, which together account for 90% of a country's aggregated market capitalization. Small stocks are defined as those stocks that comprise the next 7% of aggregated market capitalization (so that big and small stocks together account for 97% of the aggregated market size of a country). Microcaps comprise the remaining 3%.[6] Although micro stocks represent only 3% of the total market capitalization of our emerging market universe, they account for 67% of the number of stocks, which is similar to the portion reported in Fama and French (2008) and Hanauer (2020) for the U.S. and developed markets, respectively. To avoid our results being

[5]The Chinese sample includes only stocks that are classified as non "A"-shares to proxy the investment universe for an international investor as for the majority of our sample period, the China A-share market was only accessible to local investors (cf., Jansen et al., 2021).

[6]To distinguish between these size groups, Fama and French (2008) use the $20^{\text{th}}$ and $50^{\text{th}}$ percentiles of end-of-June market cap on NYSE stocks as size breakpoints for the U.S. market, which on average are bigger than AMEX or NASDAQ stocks. However, these breakpoints are applied to all (NYSE, AMEX, and NASDAQ) stocks. For international markets, Fama and French (2012, 2017) propose to calculate breakpoints based on aggregated market capitalization, as we do.

driven by microcaps, we follow Hou et al. (2020) and Hanauer and Lauterbach (2019) and exclude them. Finally, we cap the market capitalization of each stock within each month by its 99% percentile to avoid our results being driven by erroneous data and a few mega-caps.

We calculate returns from the total return index in USD. Following Jacobs (2016) and Hanauer and Lauterbach (2019), we winsorize all returns each month within a country at 0.1% and 99.9% to eliminate potential errors. To calculate the excess returns, we obtain the risk-free rate from Kenneth R. French's homepage.[7]

[Table 1 about here.]

The result is a comprehensive dataset spanning 15.152 unique stocks and more than 1.42 million stock-month observations. Table 1 depicts the descriptive statistics for the final sample.

## 2.2 Firm-level characteristics

The 36 firm-level characteristics in this study are analogous to those in Kelly et al. (2019) and Windmüller (2022) and constructed using data from Refinitiv Datastream and Worldscope. Appendix B outlines the detailed construction of the characteristics. We follow Windmüller (2022) and substitute the daily bid-ask spreads with the daily version of Amihud (2002) illiquidity as a proxy for trading frictions. As shown by Fong et al. (2017), the Amihud (2002) illiquidity measure increases the number of observations in the cross-section and is the best daily cost-per-dollar-volume proxy for international data.

The 36 characteristics are: assets-to-market (`A2ME`), total assets (`AT`), sales-to-assets (`ATO`), book-to-market (`BEME`), market beta (`Beta`), cash-and-short-term-investment-to-assets (`C`), capital turnover (`CTO`), capital intensity (`D2A`), leverage (`Debt2P`), ratio of change in property, plants, and equipment to change in total assets (`DPI2A`), earnings-to-price (`E2P`), fixed costs-to-sales (`FC2Y`), cash flow-to-book (`FreeCF`), idiosyncratic volatility (`Idiovol`),

[7]See `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

investment (INV), market capitalization (LME), turnover (LTurnover), net operating assets (NOA), operating accruals (OA), operating leverage (OL), price relative to its 52-week high (P2P52WH), price-to-cost margin (PCM), profit margin (PM), gross profitability (Prof), Tobin's Q (Q), momentum ($r_{12-2}$), intermediate momentum ($r_{12-7}$), short-term reversal ($r_{2-1}$), long-term reversal ($r_{36-13}$), return on net operating assets (RNA), return on assets (ROA), return on equity (ROE), sales-to-price (S2P), the ratio of sales and general administrative costs to sales (SGA2S), unexplained volume (SUV), and Amihud (2002) illiquidity (Illiqu).

Moreover, in a robustness check, we add the following four characteristics that have been shown to be strong return predictors for emerging markets (Hanauer and Lauterbach, 2019): monthly updated book-to-market ($\text{BEME}_m$, Asness and Frazzini, 2013), composite equity issuance (CEI, Daniel and Titman, 2006), cash flow-to-price (CF2P, Lakonishok et al., 1994), and gross profitability-to-assets (GP2A, Novy-Marx, 2013).

We do not exclude financial firms but set the following characteristics as missing as they are not meaningfully defined for financials: ATO, C, D2A, DPI2A, FC2Y, FreeCF, CF2P, GP2A, OA, PCM, PM, Prof, RNA, SGA2S, and NOA.

Following Freyberger et al. (2020), Gu et al. (2020), and Leippold et al. (2022), we cross-sectionally rank all stock characteristics each month for every country into the [-1,1] interval to limit the effect of outliers. These country-based ranks intend to control for different accounting standards across countries, particularly in the earlier part of the sample period, and therefore account for cross-country differences in characteristics. In case of missing characteristics, we replace them with a 0 to ensure extensive cross-sectional coverage. The annually updated characteristics incorporating balance sheet data from the fiscal year ending in calendar year t-1 are used from end-of-June in year t to end-of-May in year t+1 to predict stock returns from July in year t to end-of-June in year t+1.

# 3 Methodology

## 3.1 Return prediction using machine learning

Rasekhschaffe and Jones (2019) stress that domain knowledge is essential to structure the forecasting problem in a way that increases the signal-to-noise ratio. As we are interested in the cross-section of stock returns and rank stocks in portfolio sorts later in a country-neutral manner, we aim to forecast the outperformance of a stock relative to its country market return. Therefore, we define the abnormal return of a stock $i$, $i = 1, ..., N$ in month $t$, $t = 1, ..., T$ in the country $c$, $c = 1, ..., C$ as

$$r^{abn}_{i,t,c} = r_{i,t,c} - Mkt_{t,c}, \tag{1}$$

where $r_{i,t,c}$ is the return of stock $i$ in month $t$ of country $c$ and $Mkt_{t,c}$ is the value-weighted market return in month $t$ of country $c$.

Following Gu et al. (2020), we employ a general additive prediction model to describe the one-month-ahead abnormal return of a stock $r^{abn}_{i,t+1,c}$, which can be written as

$$r^{abn}_{i,t+1,c} = E_t[r^{abn}_{i,t+1,c}|x_{i,t}] + \epsilon_{i,t+1,c}, \tag{2}$$

where $E_t[r^{abn}_{i,t+1,c}|x_{i,t}]$ is the conditional expected abnormal return of stock $i$ in month $t$ for month $t+1$ given a vector of stock-specific $p$ characteristics known at month $t$, $x_{i,t} \in \mathbb{R}^p$, and $\epsilon_{i,t+1,c}$ is the prediction error term. Our objective is to model the expected abnormal return by using an unknown function $f^*$, $f^* : \mathbb{R}^p \to \mathbb{R}$, which estimates the expected returns independently of any other information besides the vector of $p$ stock-specific characteristics available in month $t$:

$$E_t[r^{abn}_{i,t+1,c}|x_{i,t}] = f^*(x_{i,t}). \tag{3}$$

In the case of supervised machine learning, the unknown function $f^*(x)$ is approximated by some function $f(x, \theta, \rho)$, which is parameterized by a vector of coefficients $\theta$ and a set of hyperparameters $\rho$. While $\theta$ is directly derived from the underlying training data with respect to $\rho$ and a specific loss function $L$, $\rho$ itself depends on the user input but is optimized concerning $L$ based on available data. The exact functional form of $f$ depends on the family and can either be linear or non-linear, as well as parametric or non-parametric.

For this paper, we build on Rasekhschaffe and Jones (2019), Gu et al. (2020), Tobek and Hronec (2020), Drobetz and Otto (2021), and Leippold et al. (2022) to select a representative amount of machine learning models from the finance literature. We analyze the predictive power of nine different algorithms: ordinary least squares ($OLS$) regression, elastic net ($ENet$), gradient-boosted regression trees ($GBRT$), random forest ($RF$), and neural networks with one to five layers ($NN_1$, $NN_2$, $NN_3$, $NN_4$, $NN_5$). We also investigate the performance of an ensemble of the five different neural networks ($NN_{1-5}$) and the average combination of the more advanced machine learning methods ($ENS$): $GBRT$, $RF$, and $NN_{1-5}$. We provide a more detailed description of the models in Appendix C.

Besides the model selection, we also follow the standard approach in the literature (Gu et al., 2020; Leippold et al., 2022) for selecting the hyperparameter range, the training of the models, and the performance evaluation. One of the most crucial things when estimating the different machine learning models is to avoid data leakage. This happens when information from exceeding the training dataset is used to create the model. Therefore, we divide our data into three disjoint periods, which always maintain the temporal ordering: the training sample, the validation sample, and the testing sample. We first estimate the models for a range of hyperparameters based on the training data. Next, we determine the respective loss of each hyperparameter set and model in the validation sample. The optimal hyperparameter set minimizes the individual model's respective loss function. Afterward, we retrain the model with the optimal hyperparameter set on the combined training and validation data. Next, the models are used to predict the monthly returns for the test dataset. We exemplary describe this procedure for the first two years in our sample: we first estimate the models for a range of hyperparameters based on the training data from July 1990 to December 1995. Afterward, we determine the best hyperparameters through the validation sample from January 1996 to December 2001. Finally, the model is retrained with the optimal hyperparameter using the data from July 1990 to December 2001 and evaluated in the testing sample using data from January 2002 to December 2002. To test our models from January 2003 to December 2003, we extend the training sample by one year (July 1990 to December 1996) and roll the validation sample forward by one year (January 1997 to December 2002). This procedure ensures that no future information is leaked from a previous period. Since machine learning models are computationally intensive, we retrain them only once a the end of every year but do the prediction every month using the latest model and data. Appendix C.5 summarizes the hyperparameter tuning schemes for each model.

## 3.2 Machine learning portfolios

We rely mainly on portfolio sorts to evaluate the predictive performance of the different machine learning models. For a given machine learning model, we adopt the following approach: At the end of each month $t$, we predict the next month's abnormal return ($\hat{r}^{abn}_{i,t,c}$). To avoid that small stocks or certain countries dominating our results, we estimate the quintile breakpoints for each country individually using the big-stock subsample based on $\hat{r}^{abn}_{i,t,c}$ as recommended in Hou et al. (2020) and applied in Hanauer and Lauterbach (2019). Furthermore, the machine-learning-based signals should not only contain information on the return predictability in equal-weighted sorts, which smaller stocks might drive but also in value-weighted sorts, which on the other hand, are dominated by larger stocks. Finally, we construct a zero-net investment portfolio (long-short) that goes long in the highest quintile portfolio and short in the lowest quintile portfolio. For all the portfolios, reassignment and rebalancing occur at the end of each month.

### 3.3 Benchmark factor models

To benchmark the results of the different machine learning portfolio sorts, we consider the Fama and French (2018) six-factor model, i.e., the Fama and French (2015) five-factor model with a cash-based profitability factor and augmented with the Carhart (1997) momentum factor. The corresponding six factors are market ($RMRF$), size ($SMB$), value ($HML$), profitability ($RMW$), investment ($CMA$), and momentum ($WML$). Appendix D provides a detailed description of the factor construction.

## 4 Empirical results

This section reviews the evidence on applying the different machine learning models within emerging markets. We start by analyzing the out-of-sample $R^2_{OOS}$ of individual stock returns. Afterward, we evaluate the importance of the different characteristics, the sensitivity of the predicted return to different characteristics, and the sensitivity to the interaction effects of various characteristics. In the next step, we utilize portfolio sorts to assess the economic gains of the different machine learning models. Finally, we investigate the impact of various methodological changes and the robustness of our findings within emerging market subregions.

### 4.1 Prediction performance

Table 2 displays the predictive power for our set of machine learning models as measured by the out-of-sample $R^2_{OOS}$. In Panel B, we include the Newey and West (1987) adjusted Diebold and Mariano (1995) test statistics, which enables us to compare the out-of-sample stock-level prediction performance between each machine learning model. We measure the pooled out-of-sample $R^2_{OOS}$ in Panel A as:

$$R^2_{OOS} = 1 - \frac{\sum_t^T \sum_i^N (r^{abn}_{i,t,c} - \hat{r}^{abn}_{i,t,c})^2}{\sum_t^T \sum_i^N (r^{abn}_{i,t,c})^2}. \tag{4}$$

[Table 2 about here.]

The first row in Panel A of Table 2 reports the $R^2_{OOS}$ of the full sample. The $OLS$ yields a benchmark $R^2_{OOS}$ of 0.29%, which all other models improve besides the $ENet$ ($R^2_{OOS}$ of 0.18%). As the $ENet$ shrinks certain coefficients towards zero but doesn't consider interactions or non-linearities, it seems that this regularization does not increase the predictability. The $RF$ and $GBRT$ are superior to the $OLS$, producing fits of 0.40% and 0.52%, respectively. Only the $NN_1$ underperforms the $GBRT$ but outperforms all other linear and non-parametric models and yields a $R^2_{OOS}$ of 0.49%. The $NN_2$ till $NN_5$ show $R^2_{OOS}$ between 0.53% and 0.55%, with the $NN_4$ performing best. Creating an ensemble of neural networks ($NN_{1-5}$) and an ensemble of the non-linear machine learning models ($ENS$) produces fits for both models of 0.60%.

A closer look at the second and third rows in Panel A of Table 2 reveals an interesting pattern: in all the cases, the predictive performance is better for small firms than for large firms. The ensemble of neural networks ($NN_{1-5}$) and the ensemble of non-linear machine learning models ($ENS$) yield a $R^2_{OOS}$ of 0.34% and 0.38% for large firms and 0.75% and 0.73% for small firms, respectively.

Whereas Panel A measures the individual predictive performance of the different machine learning models, Panel B assesses the statistical significance of differences among the models using the Newey and West (1987) adjusted Diebold and Mariano (1995) test statistics ($DM_{kj}$) comparing a column model ($k$) versus a row model ($j$). We compute the Newey-West adjusted Diebold-Mariano test statistics as:

$$\begin{aligned}
MSFE_t^m &= \frac{1}{N_t}\sum_{i=1}^{N_t}(r_{i,t,c}^{abn} - \hat{r}_{i,t,c,m}^{abn})^2 \\
d_{kj,t} &= MSFE_t^k - MSFE_t^j \\
\bar{d}_{kj} &= \frac{1}{T}\sum_{T}^{t=1} d_{kj,t} \\
DM_{kj} &= \frac{\bar{d}_{kj}}{\hat{\sigma}_{d_{kj},NW(4)}},
\end{aligned} \tag{5}$$

where $\hat{\sigma}_{d_{kj},NW(4)}$ is the Newey and West (1987) standard error of $d_{kj,t}$ with four lags. The Diebold-Mariano test statistic is normally distributed with a mean of 0 and a standard deviation of 1 ($\mathcal{N}(0, 1)$) with the null hypothesis that there exists no difference between the models, which allows us to map the magnitudes of the test statistic to $p$-values. Bold numbers indicate a statistical significance for each test at the 1% level ($DM \geq 2.60$). An asterisk indicates significance at the 1% level for 10-way comparisons via the conservative Bonferroni adjustment, which increases the critical value to 3.33.

We conclude that besides the $ENet$, every machine learning model is superior to the $OLS$, and every model surpasses the Bonferroni adjusted critical value of 3.33. Comparing the $RF$ to all the other non-linear models yields a similar result. Both the $GBRT$ as well as all other neural networks besides the $NN_1$ are superior to the $RF$. In the case of the $GBRT$, only the two ensembles can deliver statistically significant better predictions with a $DM$ statistic of 3.49 and 7.59, respectively. The different neural networks with one to five layers do not differ much in their prediction performance. In the case of the $NN_1$, the neural networks with four and five layers are superior. The two best-performing machine learning models are the two ensembles. Whereas the ensemble of neural networks ($NN_{1-5}$) can significantly outperform all other machine learning models, the other ensemble of the trees and neural networks yields statistically significant outperformance measures when comparing it to the $OLS$, $ENet$, $RF$, $GBRT$, $NN_1$, $NN_3$, and $NN_5$.

## 4.2 Characteristics importance and marginal relationships

We examine whether specific characteristics are more important than others to predict the next month's abnormal returns and the model-implied marginal impact of individual characteristics on expected abnormal returns.

We define the importance of the characteristics for each model as the average reduction in $R^2_{oos}$ by setting each value of the particular characteristic to zero and keeping the remaining model estimates fixed. Figure 1 visualizes the sum over the cross-sectional ranked characteristics for the different machine learning models.[8] A darker color indicates higher importance of the characteristic for the individual model, while a lighter color indicates lower importance for the $R^2_{oos}$.

[Figure 1 about here.]

The most influential characteristics are similar among the different machine learning models. Among the top 15, we find the following characteristics: turnover ($LTurnover$), idiosyncratic volatility ($Idiovol$), price relative to its 52-week high ($P2P52WH$), Amihud (2002) illiquidity ($Illiqu$), total assets ($AT$), market capitalization ($LME$), and market beta ($Beta$) from the trading frictions category; momentum ($r_{12-2}$), short-term reversal ($r_{2-1}$), and intermediate momentum ($r_{12-7}$) from the past returns category; and assets-to-market ($A2ME$), Tobin's Q ($Q$), book-to-market ($BEME$), and leverage ($Debt2P$) from the value category. In contrast, characteristics of the profitability and intangibles categories are not present among the top 15, except for return on asset ($ROA$).

Figure 2 visualizes the marginal impact of individual characteristics on expected abnormal returns for the $OLS$, $ENet$, $RF$, $GBRT$, and $NN_{1-5}$. We predict the returns for each model and characteristic by iterating over the (-1,1) interval and holding all other characteristics fixed at the value of zero. We do this for each time period and model individually and calculate the average predicted return among the different machine learning models.

[8] We additionally show the most influential characteristics per model and the corresponding normalized importance in Figure E.1 in the appendix.

We exemplary select short-term reversal ($r_{2-1}$), idiosyncratic volatility ($Idiovol$), turnover ($LTurnover$), and operating leverage ($OL$) to visualize how the different machine learning models associate the underlying characteristic with the expected abnormal returns.

[Figure 2 about here.]

Inspecting the relationships in Figure 2, we see that all methods identify the well-known negative relationship of expected returns with short-term reversal ($r_{2-1}$, top-left) or idiosyncratic volatility ($Idiovol$, top-right). While the two linear models are, per definition, restricted to linear relationships, we see that tree-based methods and neural networks identify more pronounced short-term reversal patterns in the extremes.[9] Similarly, these methods also find a rather flat relationship for low and medium levels of idiosyncratic volatility ($Idiovol$) but an increasingly negative relationship for high idiosyncratic volatility, echoing the empirical results in Ang et al. (2006). The differences are even more pronounced for turnover ($LTurnover$, bottom-left). While both $OLS$ and $ENet$ find a positive slope, the two tree-based models, $RF$ and $GBRT$, and the neural network ensemble, $NN_{1-5}$, identify an inverted U-shape pattern: extreme positive and negative values of $LTurnover$ are related to a lower expected return than the middle region in the interval, echoing the pattern documented in Freyberger et al. (2020). Such differences in marginal relationships can, in part, explain the divergence in the performance of linear and non-linear methods. However, we also find that all methods agree on a nearly zero relationship of operating leverage ($OL$, bottom-right) with expected returns.

A significant advantage of the tree-based models and the different neural networks is that they can model complex interactions between the different characteristics. In Figure 3, we exemplary visualize the $NN_{1-5}$'s sensitivity of the expected monthly percentage returns to the effects of the interactions for Amihud (2002) illiquidity ($Illiqu$) and idiosyncratic volatility ($Idiovol$) with short-term reversal ($r_{2-1}$) and market capitalization ($LME$) by

[9] This finding is consistent with the empirical pattern for short-term reversal deciles that can be found on Kenneth R. French's homepage.

varying both pairs of characteristics while holding the predictors fixed. On the one hand, we choose *Illiqu* and *Idiovol* as they are prominent hard-to-value proxies (cf., Kumar, 2009) and on the other hand, $r_{2-1}$ and $LME$ as they are two main control characteristics in the asset pricing literature.

[Figure 3 about here.]

The upper-left figure shows that for very illiquid stocks (purple line), the difference between high and low previous month returns is the most considerable. In contrast, the lines are mainly parallel in the case of the other values of *Illiqu*. This interaction resembles the empirical findings in Medhat and Schmeling (2022) for the interaction between short-term reversal with turnover. In the upper-right figure, we plot the interactions between the Amihud (2002) illiquidity and a stock's market capitalization. In the case of liquid firms (blue and orange line), the expected return increases by increasing market capitalization, while for illiquid firms (red and purple line), the relationship is reversed, indicating a decrease in expected returns for larger firms. The bottom-left figure shows that the short-term reversal effect is most substantial and S-shaped among risky stocks (purple line). Among less risky stocks (blue and orange line), reversal is concave, yielding significantly lower returns when the prior month's return is considerable. Finally, the bottom-right figure shows that no strong interaction effects exist between *Idiovol* and $LME$.

## 4.3 Portfolio performance

After providing evidence on the predictive ability of the different machine learning methods for individual stock returns, we will continue with a general overview of the profitability of machine learning signal-based portfolios.

Table 3 presents the results on equal- and value-weighted country-neutral quintile portfolio sorts using big-stock breakpoints. In Panel A and Panel D, we provide results on the predicted monthly returns for the long-short quintile (Pred), the average monthly return for the long-short quintile (Avg), Newey and West (1987) adjusted $t$-statistics with four lags ($t$-stat), the monthly standard deviations (SD), and Sharpe ratios (SR). Panel B and Panel E reports the alpha ($\alpha$), corresponding Newey and West (1987) adjusted t-statistics with four lags ($t$-stat$_\alpha$), and $R^2$ with respect to the Fama and French (2018) six-factor model:

$$\begin{aligned} r_{p,t,ML} - r_{f,t} = \alpha + \beta_1 RMRF_t + \beta_2 SMB_t + \beta_3 HML_t \\ + \beta_4 RMW_t + \beta_5 CMA_t + \beta_6 WML_t + \epsilon_t. \end{aligned} \tag{6}$$

We additionally provide detailed results on every quintile in Appendix F.1.[10]

Panel C and Panel F describe the maximum drawdowns (Max DD), the most negative monthly return (Max 1M Loss), and the average monthly percentage change in holdings (TO) of different machine learning-based long-short portfolios. We define maximum drawdowns as

$$\text{Max DD} = \max_{0 \leq t_1 \leq t_2 \leq T} (Y_{t_1} - Y_{t_2}), \tag{7}$$

where $Y_t$ is the cumulative log return from date 0 through $t$. The strategy's average monthly turnover is defined as

$$\text{TO} = \frac{1}{T} \sum_{t=1}^{T} \left( \sum_{i=1}^{N_t} \left| w_{i,t+1} - \frac{w_{i,t}(1 + r_{i,t+1})}{1 + \sum_{j=1}^{N_t} w_{j,t} r_{j,t+1}} \right| \right), \tag{8}$$

where $w_{i,t}$ is the weight of stock $i$ in the portfolio at time $t$.

[Table 3 about here.]

We start by analyzing the equal-weighted long-minus-short quintile returns in Panel A of Table 3. All machine learning models yield positive and highly significant long-short

[10] We also include the performance of a strategy that uses the equal-weighted (1/N) average of all standardized characteristics ($\mu_{sign(c)}$) in this table. Thereby, characteristics are sorted in such a way that higher values correspond to higher expected returns. The performance of this simple linear combination is slightly worse (similar) than the performance of the other two linear strategies for equal-weighted (value-weighted) portfolios.

returns, and the order is similar to the one of the monthly out-of-sample stock-level prediction performance in Table 2. The linear methods *OLS* and *ENet* yield a monthly return of 1.38% ($t$-stat 7.82) and 1.20% $t$-stat 6.83), respectively. However, the tree-based methods *RF* and *GBRT* exhibit even higher long-short returns of 1.60% ($t$-stat 9.33) and 1.80% $t$-stat 11.57), which themselves are outperformed by the neural networks with returns between 1.84% ($NN_3$) and 1.91% ($NN_2$) and $t$-statistics between 13.82 ($NN_4$) and 15.75 ($NN_2$). The ensemble of the different neural networks ($NN_{1-5}$) yields a similar performance as $NN_5$, and the ensemble of the tree-based methods and neural networks has a performance similar to the $GBRT$.

The risk-adjusted performance displayed in Panel B leads to the same order as for the raw long-short returns. However, the increase in the six-factor alpha for the machine learning models compared to the linear models is even more pronounced as the six-factor model has less explanatory power for them. Furthermore, Panel C reveals that the neural network portfolios exhibit a smaller maximum drawdown and a smaller maximum one-month loss compared to the linear and tree-based models. The maximum drawdown (worst one-month return) in the case of the ensemble of neural networks is 17.82% (10.75%), whereas this number is 26.35% (13.97%) for the $OLS$. The superior performance of the machine learning models comes at the cost of a somewhat higher turnover. However, compared to the performance gains, this turnover increase from 89.27% for $OLS$ to values between 89.61% for $RF$ and 102.02 for $NN_2$ is relatively small.

Turning to the results for value-weighted portfolios in Panels D to E of Table 3 reveals identical qualitative conclusions, but the return spreads, $t$-statistics, and Sharpe ratios are substantially lower. Although the return forecasts derived from linear models already lead to economically and statistically significant long-short mean returns and six-factor alphas, the tree-based methods and neural networks do even better. Again, the neural network with two layers exhibits the highest $t$-statistics and Sharpe ratios while suffering from the mildest drawdowns. Comparing the ensemble of machine learning methods ($ENS$) with the linear $OLS$ regressions shows performance gains of roughly 50% for the raw quintile returns and even higher for the risk-adjusted performance. In sum, allowing for non-linearities and interactions also leads to economically superior out-of-sample returns compared to traditional linear models, as summarized in Figure 4.

[Figure 4 about here.]

Figure 5 illustrates the results of Table 2 by plotting the equal-weighted and value-weighted cumulative performance of selected long-short strategies. We additionally include the cumulative performance for the long and short sides for select strategies in Appendix E.2.

[Figure 5 about here.]

By using a value-weighted portfolio strategy, $RF$ initially dominates the other methods, while the outperformance of $GBRT$ and $NN_{1-5}$ mainly stems from the period after 2009. As the $ENS$ comprises all three methods, we observe a rather consistent outperformance versus $OLS$ that is not driven by a particular period. In the case of equal-weighted portfolios, there are only small differences between the portfolio returns of $GBRT$, $NN_{1-5}$, and $ENS$ till 2021. As for the value-weighted portfolios, the machine learning methods outperform the linear approaches consistently over time. The model with the lowest cumulative return is the $ENet$, whereby the underperformance versus the $OLS$ is mainly driven by the first years of the sample period. Besides a sharp drawdown in 2009, there are no other notable downturns for all approaches. The drawdown in 2009 probably stems from the models' exposure to momentum that exhibited a momentum crash at that time (Daniel and Moskowitz, 2016). The recent global shock due to the COVID-19 pandemic in early 2020 did not lead to a significant portfolio-level downturn.

## 4.4 Robustness

To check the robustness of the results presented above, we will investigate (i) the impact of various methodological changes and (ii) the robustness within emerging market subregions.

In Table 4, we summarize the robustness tests for methodological changes. We include several performance indicators for our equal-weighted and value-weighted machine learning portfolio strategies. We will focus ourself on comparing the benchmark $OLS$ model to the $ENS$. We select the $ENS$ to be not driven by a look-ahead bias in regard to the model selection and its portfolio performance. Besides the individual long-short return and the six-factor model of the two machine learning models, we include the results of the following two regressions in the last two rows of each panel:

$$\begin{aligned} r_{LS,t,ENS} &= \alpha + \beta_{OLS} r_{LS,t,OLS} + \epsilon_t, \\ r_{LS,t,OLS} &= \alpha + \beta_{ENS} r_{LS,t,ENS} + \epsilon_t. \end{aligned} \tag{9}$$

A positive and significant alpha indicates that the returns of the strategy on the right-hand side cannot fully explain the portfolio returns on the left-hand side.

[Table 4 about here.]

The first two rows in Panel A show again our baseline result for $OLS$ and $ENS$ as already shown in Table 3. Furthermore, the last two rows of Panel A show that both for equal- and value-weighted portfolios, the $OLS$ long-short portfolios cannot span the $ENS$ long-short portfolio, but the $ENS$ spans the $OLS$.

In Panel B, we construct our long-short trading strategy using decile instead of quintile sorts. By focusing on the more extreme predicted abnormal returns and due to the monotonic increase among the portfolios, the equal-weighted and value-weighted long-short returns of the $OLS$ increase to 1.84% ($t$-stat 10.09) and 1.18% ($t$-stat 5.91), whereas the Fama and French (2018) six-factor alpha increases to 1.41% ($t$-stat 11.30) and 0.55% ($t$-stat 4.66), respectively. The $ENS$ shows an increase in the return to 2.50% ($t$-stat 13.93) and 1.66% ($t$-stat 8.12) as well as in the risk-adjusted return to 2.02% ($t$-stat 18.31) and 1.10% ($t$-stat 10.30). Therefore, both $OLS$ and $ENS$ show stronger results when using decile sorts. Still, the increase in returns of the $ENS$ is higher than the $OLS$ resulting in a larger $\alpha$ when regressing the $ENS$ on the $OLS$ compared to Panel A.

The robustness test in Panel C includes the additional predictive characteristics described in Hanauer and Lauterbach (2019). Especially the $OLS$ is profiting from this extended feature set. The average equal-weighted and value-weighted long-short return increase by 9% and 4%, while only the equal-weighted return of $ENS$ increases by 5%. In the case of the value-weighted risk-adjusted return, the $OLS$ alpha increases by 28% and the $ENS$ by 6%.

Reducing the number of characteristics by applying a lasso regression, i.e., feature selection, before training the machine learning models reduces the equal-weighted and value-weighted long-short returns as well as the equal-weighted risk-adjusted returns of both machine learning models but increases the value-weighted alpha of the ensemble as presented in Panel D.

In Panel E, we utilize machine learning models, which were never trained on emerging market stock returns; instead, the models are trained on developed markets (as defined by MSCI). Although the models were solely trained on developed markets, we surprisingly do not observe a big performance loss but actually very similar returns. Furthermore, models that allow for non-linearities and interactions ($ENS$) still significantly outperform linear models ($OLS$). This indicates that machine learning models can create significant results even if they are evaluated on data from a totally different region.

For the robustness test in Panel F, we exclude the high-turnover characteristics `Idiovol`, `LTurnover`, $\mathtt{r}_{2-1}$, `SUV`, `Illiqu` from the feature set. Whereas the risk-adjusted returns of the $OLS$ decrease by 8% and 15%, the ensemble is even more affected as the alphas are reduced by 19% and 31%. This indicates that these high-turnover features are relatively more important for more complex methods. But even after excluding the characteristics, the long-short portfolios based on the $OLS$ cannot span the long-short portfolios constructed based on the $ENS$.

In Panel G, we do not train our models on a pooled sample of all countries but separately for each of the following subregions: Central and Latin America (Americas); Asia; and Europe, Middle East, and Africa (EMEA). On the one hand, this allows the models to capture potential region-specific effects. On the other hand, each model is now trained on less data, which might be a drawback, especially for identifying non-linearities and interactions. We document that subregional training leads to inferior return forecasts than training models on pooled data from all subregions. This finding indicates that region-specific effects play a minor role compared to more data for out-of-sample returns. Furthermore, we find that the performance decay is more pronounced for the machine learning ensemble ($ENS$), i.e., indicating that more data is better for robustly identifying non-linearities and interactions.[11] Nevertheless, the $OLS$ long-short portfolios cannot span the $ENS$ long-short portfolio, but the $ENS$ spans the $OLS$.

Finally, we assess if the superior performance of the machine learning return forecasts is robust across emerging market regions in Table 5. Therefore, we divide the countries of our full sample into three regions: Central and Latin America (Americas); Asia; and Europe, Middle East, and Africa (EMEA).

[Table 5 about here.]

Overall, the results are robust for the different sub-regions Americas, Asia, and EMEA. Both $OLS$ and $ENS$ yield positive and significant long-short returns and alphas for both weighting schemes, but $ENS$ exhibit higher returns and $t$-statistics. Furthermore, significant and positive alphas remain in the spanning regression of $ENS$ on $OLS$ for all sub-regions, but no positive spanning alphas remain when regressing $OLS$ on $ENS$. Comparing the results across sub-regions, we find the strongest results for Asia and EMEA and a bit weaker but still highly significant results for Americas.

[11] Table F.2 in the Appendix shows that the performance decline is most pronounced for the smaller regions Americas and EMEA while smaller for Asia.

# 5 Understanding the sources of return predictability

The results so far provide evidence that return forecasts based on machine learning models lead to economically and statistically superior out-of-sample long-short returns compared to traditional linear models. To further understand the source of return predictability, we first investigate the performance of the two models in higher- versus lower-risk months. Second, we explore how far developed markets' long-short returns can explain emerging markets' long-short returns. Third, we turn to the time-series properties of the long-short machine learning portfolios over the next 36 months after portfolio formation. Fourth, we link the profitability of the machine learning models to several proxies for limits to arbitrage. Finally, we investigate the performance of an investment strategy that considers real-life investment frictions such as short-selling restrictions and transaction costs.

## 5.1 Performance in higher-risk versus lower-risk months

The profitability of return forecasts based on machine learning models may reflect risks beyond what we control so far. Therefore, we investigate the performance of $OLS$ and $ENS$ forecasts in higher- versus lower-risk months. As proxies for risk, we apply whether (i) emerging markets as a whole go up or down, (ii) the rate on long-term U.S. government bonds is going up or down, (iii) the TED spread is below or above its median value, and (iv) the time-varying risk aversion index proposed by Bekaert et al. (2022) (RAbex).[12] Splitting the sample period into up and down markets is done, for instance, by Chan et al. (1998), van der Hart et al. (2005), or Asness et al. (2019). The change in the U.S. government bond rate as a proxy for risk is motivated by the substantial financial instability emerging markets experienced during the so-called 'taper tantrum' in 2013 when U.S. yields surprisingly surged (cf., Estrada et al., 2016). Frazzini and Pedersen (2014) argue that the TED spread is a

[12]The TED spread is defined as the difference between the LIBOR rate and the 3-month U.S. T-bill rate. The 10-year constant maturity U.S. Treasury rate (item DGS10) and the TED spread (item TEDRATE) are from the FRED database of the Federal Reserve Bank of St. Louis, and the time-varying risk aversion index RAbex is from Nancy Xu's website: https://www.nancyxu.net/risk-aversion-index.

measure of funding conditions. Finally, Bianchi et al. (2021b) apply RAbex to investigate the link between time-varying risk aversion and excess bond returns.

[Table 6 about here.]

Table 6 summarizes the top-bottom quintile returns for $OLS$ and $ENS$ for the different subsamples. For both equal- and value-weighted portfolios, we observe that the performance is somewhat higher in down-market months and months with rising bond yields. However, we also document higher returns when the TED spread is low, i.e., when funding conditions are better and in months with below-median risk aversion. Nevertheless, the quintile spreads are positive and significant for all subsamples, prediction models, and weighting schemes. Furthermore, the difference between the subsamples is less pronounced for $ENS$ than for $OLS$, and significant and positive alphas remain in the spanning regression of $ENS$ on $OLS$. At the same time, the converse is not the case. This evidence suggests that the superiority of machine learning models compared to linear models in our sample does not stem solely from higher-risk months, at least for the definitions considered here.

## 5.2 Market integration

The robustness tests in Table 4 reveal an interesting finding: models estimated solely on developed markets data predict emerging market stock returns similarly good as when the models are trained with emerging markets data itself. This result could indicate that developed and emerging market pricing is more integrated, as suggested by the results for value and momentum returns in Cakici et al. (2013) and Hanauer and Linhart (2015). If developed and emerging markets are integrated, the developed market machine learning long-short portfolio returns would be able to explain the market machine learning long-short portfolio returns for emerging markets, i.e., resulting in an insignificant $\alpha$ in the following regression for the global and regional emerging market samples:

$$r_{LS,Region_{EM},t,ENS} = \alpha + \beta_1 r_{LS,Global_{Dev},t,ENS} + \epsilon_t. \tag{10}$$

However, the results in Table 7 reveal that this is not the case. All alphas remain highly significant for both the equal-weighted and value-weighted portfolios. For the equal-weighted factor, Asia has the highest alpha of 1.47% ($t$-stat 8.66), followed by EMEA with 1.24% ($t$-stat 7.39). The value-weighted factor construction yields the highest alpha for Asia with 0.98% ($t$-stat 4.92), followed by EMEA with an alpha of 0.87% ($t$-stat 4.36). Furthermore, the developed market long-short portfolio returns can only explain 32% and 28 % of the variation in emerging market long-short portfolio returns when equal or value-weighting the portfolio returns.

[Table 7 about here.]

Our interpretation of these results is as follows: Although similar relationships between firm characteristics and future stock returns exist for developed and emerging markets, the pricing of these characteristics is still not fully integrated between developed and emerging markets. Furthermore, our results indicate potential diversification benefits of machine learning emerging market strategies for investors already applying such a strategy in developed markets.

## 5.3 Performance for longer holding periods

Is the profitability of the machine learning forecasts the result of temporary or permanent price changes? To answer this question, we analyze the long-run buy-and-hold returns following the methodology in Smajlbegovic (2019). First, we identify stocks used for constructing the long-short machine learning portfolios and calculate their value-weighted raw monthly returns in the month $t + k$, where $k \in \{1, ..., 36\}$. Second, we run a time-series regression for each holding period month $k$ of the machine learning long-short factor on the six-factor model. The corresponding strategy average six-factor alpha at month $k$ is the intercept ($a_k$) in the following regression:

$$r_{t+k,ML} - r_{f,t+k} = \alpha_k + \sum_{i}^{|f|} \beta_{i,k} f_{i,t+k} + \epsilon_{t+k}, \tag{11}$$

where $r_{t+k,ML} - r_{f,t+k}$ is the raw long-short return in month $t+k$ of stocks used for construction of the long-short machine learning factor in month $t$ and $f_{i,t+k}$ indicates the individual factor returns of the six-factor model in month $t+k$: $RMRF_{t+k}$, $SMB_{t+k}$, $HML_{t+k}$, $RMW_{t+k}$, $CMA_{t+k}$, and $WML_{t+k}$. The intercept of the regression ($\alpha_k$) is the alpha of the buy-and-hold strategy k months after portfolio formation, which added up to the cumulative alpha in month $k$ by $ACR_k$:

$$ACR_k = \sum_{t=1}^{k} \alpha_t. \tag{12}$$

Figure 6 presents the value-weighted cumulative six-factor alpha of $OLS$ and $ENS$ over a holding period of 36 months. The figure reveals that both $OLS$ and $ENS$ can predict long-term returns and that their performance does not revert quickly. Together with the fact that standard risk factors cannot explain the performance of the strategies and the consistent performance over calendar time, we conclude that an underreaction explanation is more likely than an overreaction explanation. We further document that the superior performance of $ENS$ compared to $OLS$ is mainly driven by the first six months. Later both lines show a relatively parallel trend. This observation is not surprising as the models are trained on one-month ahead returns and not on longer periods.

[Figure 6 about here.]

## 5.4 Limits to arbitrage

Our results so far suggest that the high returns of the machine learning strategies in emerging markets are not explained by standard risk factors such as the factors from the Fama and French (2018) six-factor model, consistent over time, do not primarily stem from higher-risk months, and do not revert quickly. Therefore, a simple question arises: Why do investors not arbitrage away these abnormal returns? If limits to arbitrage hinder investors from doing that, we would expect that the predictability of the machine learning forecasts is concentrated in stocks with the highest limits to arbitrage.

To test whether the predictability of machine learning methods arises, at least in part, from such frictions, we interact the predicted returns of the machine learning models with different proxies for limits to arbitrage within a Fama and MacBeth (1973) regression. We additionally include both parts of the interaction term as controls as well as country dummies to account for any country effect yielding the following regression framework:

$$r_{i,t+1} - r_{f,t+1} = \alpha + \beta_1 ML_{i,t} + \beta_2 LTA_{i,t} + \beta_3 ML_{i,t} \times LTA_{i,t} + \beta_4 X_i + \epsilon_{i,t+1}, \tag{13}$$

where $LTA_{i,t}$ denotes the cross-sectional and country-neutral standardized variable measuring the limits to arbitrage of stock $i$ while $ML_{i,t}$ is the predicted return based on the underlying machine learning model.

The coefficient $\beta_3$ is most relevant for this analysis as it indicates if the predictability of the different machine learning models is increasing with higher limits to arbitrage. We include three different variables that are closely related to limits to arbitrage and commonly used in the literature: size as a measure of information ambiguity (Zhang, 2006), idiosyncratic volatility as a proxy for arbitrage risk (Pontiff, 2006; Stambaugh et al., 2015), and Amihud (2002) illiquidity as a potential proxy for transaction costs. If limits to arbitrage are important for the persistence of mispricing, we expect that predictability is the strongest for smaller stocks with high idiosyncratic volatility and low liquidity. Therefore, we additionally include the average of these three variables.

The results of this analysis are reported in Table 8. We first examine firm size's role in predicting future returns. Most small firms are less diversified and less fundamental information is available. In the case of fixed information acquisition costs, small firms are less attractive. The results in Columns (1) and (2) underline this hypothesis. The smaller the stock, the higher the return predictability for both methods.

[Table 8 about here.]

We study how arbitrage risk affects the link between machine learning-based prediction and future stock returns in the second specification. According to Pontiff (2006), arbitrageurs prefer to hold fewer stocks with higher idiosyncratic stock return volatility. Columns (3) and (4) provide empirical evidence that stocks with higher $IVOL$ exhibit larger predictable returns than stocks that are less volatile.

Next, we test how stock illiquidity relates to our previous findings. The intuition behind this proxy is based on the tradeability of the stock. The more illiquid the stock, the slower and more costly it should be to trade on the market. However, we are not able to provide empirical evidence that the return predictability of the machine learning models is driven by transaction costs.

In the last interaction setup, we combine all three limits to arbitrage proxies to measure their mutual influence on the effect of future return predictability. Columns (7) and (8) provide evidence that stocks that are associated with more substantial limits to arbitrage characteristics exhibit stronger predictability independent of the underlying machine learning model.

However, we also find that the higher predictability for stocks with higher limits of arbitrage is less pronounced for the machine learning ensemble $ENS$ than for the linear $OLS$ regression, indicating that the superiority of machine learning models in emerging markets does not stem from limits to arbitrage.

## 5.5 Further investment frictions

A common feature of the results presented above is that they are based on theoretical "zero-investment" long-short portfolio returns. However, it is questionable whether these returns can be realized in practice, as short-selling constraints may prevent the implementation of long-short strategies, and transaction costs may erode the strategy's profits. These constraints are particularly relevant for emerging markets (see, e.g., Roon et al., 2001). Therefore, in this subsection, we limit ourselves to long-only portfolios of big stocks (i.e., also remove small stocks) and consider reasonable transaction costs. To estimate transaction cost, we compute for each stock and month the efficient discrete generalized estimator (EDGE) of the bid-ask spread, recently proposed by Ardia et al. (2022). These bid-ask spread estimates vary considerably across time and stocks (cf., Figure E.3 in the appendix) and therefore provide a more sophisticated estimate than the flat 100 basis points per single-trip used in van der Hart et al. (2003) and Hanauer and Lauterbach (2019).[13] The transaction cost per single-trip is half of the estimated bid-ask spread, and we define the transaction cost of portfolio $L$ as:

$$\text{T-Cost}_{L,t} = \left( \sum_{i=1}^{N_{L,t-1\cup t}} \left| w_{i,t} - \frac{w_{i,t-1}(1 + r_{i,t})}{1 + \sum_{j=1}^{N_{L,t}} w_{j,t-1} r_{j,t}} \right| \times \frac{S_{i,t}}{2} \right), \tag{14}$$

where $w_{i,t}$ is the weight of stock $i$ at the end of month t, $r_{i,t}$ is the total return of stock $i$ in month $t$, and $S_{i,t}$ is the estimated bid-ask spread. Furthermore, the net portfolio returns are defined as:

$$r_{L,t,net,ML} = r_{L,t,gross,ML} - \text{T-Cost}_{L,t}. \tag{15}$$

In the final step, we calculate the Fama and French (2018) six-factor model alpha return

[13] Table F.3 in the appendix also provides the results for transaction cost estimates of 100 basis points per single-trip.

as:

$$r_{L,t,net,ML} - r_{f,t} = \alpha_{net} + \sum_{i}^{|f|} \beta_i f_{i,t,net} + \epsilon_t, \tag{16}$$

where $f_{i,t,net}$ is the risk factor return after transaction cost.

Furthermore, we also consider trading cost mitigation rules following Novy-Marx and Velikov (2016) and Blitz et al. (2022), which are common among practitioners. Such buy/hold strategies consist of the stocks that currently belong to the top X% plus the stocks selected in previous months that are still among the top Y% of stocks. In Table 9, we compare the quintile long-only strategy (20%/20%) with the transaction-cost-mitigation strategy buying the top 10% and holding them in our portfolio as long as they belong to the top 30% (10%/30%).

[Table 9 about here.]

Table 9 reports the strategies' average gross excess over the market, their turnover and transaction costs, as well as the resulting net outperfomance. Limiting the investment universe to long-only portfolios of big stocks, we still see positive and significant gross outperformance for the top quintile portfolio (20%/20%) for both $OLS$ and $ENS$ and both weighting schemes. We observe similar gross outperformance when switching to the transaction cost mitigation strategies (10%/30%). However, the turnover and transactions are reduced by roughly 40%. This reduction in transaction costs substantially positively affects the net performance. For the equal-weighted strategies in Panel A, the net outperformance for $OLS$ increase from 0.19% ($t$-stat 2.11) to 0.28% ($t$-stat 3.30). The net outperformance for $ENS$ of 0.46% ($t$-stat 4.88) are also significant for the standard top quintile approach but also increase to 0.59% ($t$-stat 5.63) when applying a more efficient portfolio construction. Value-weighting the returns in Panel B leads to more challenging results. In this setup, the top $OLS$ quintile yields only an insignificant net return of 0.07% ($t$-stat 0.75). Applying the trading-cost mitigation strategy increases the net returns to 0.19% ($t$-stat 2.08) for $OLS$

and even to 0.34% ($t$-stat 3.31) for $ENS$. Similar results can be derived by comparing the Fama and French (2018) net alphas for which only the turnover-reducing strategy for $ENS$ exhibits a significant net alpha of 0.34 ($t$-stat 5.18).[14] Therefore, we conclude that machine learning-based return forecasts can lead to significant net outperformance and net alphas, at least when efficient trading rules are applied.

# 6 Conclusion

This paper compares the out-of-sample predictive power of various machine learning models for a broad sample of 32 emerging market countries and a 20-year out-of-sample period. More specifically, we utilize both linear and more complex algorithms that allow for non-linearities and interactions.

We document that the different prediction algorithms pick up similar characteristics. However, we also observe that tree-based methods and neural networks do identify non-linearities and interactions of characteristics. Furthermore, return forecasts based on machine learning models lead to economically and statistically superior out-of-sample long-short returns compared to traditional linear models. This finding is robust to several methodological choices and for emerging market subregions.

We also find that developed market long-short returns based on machine learning forecasts derived in the same way as their emerging market counterparts cannot explain emerging market out-of-sample returns. However, models estimated solely on developed markets data also predict emerging market stock returns. This finding indicates that similar relationships between firm characteristics and future stock returns exist for developed and emerging markets but that the pricing of these characteristics is not fully integrated between developed and emerging markets.

We also document that the high returns of the machine learning strategies in emerging

[14] When applying the more conservative transaction cost estimates of 100 basis points per single-trip, only the machine learning ensemble in combination with transaction cost mitigation exhibits significant net returns and alphas of 0.23% and 0.25%, respectively.

do not primarily stem from higher-risk months and do not revert quickly, suggesting that an underreaction explanation is more likely than a risk-based explanation. Although both linear and machine learning models show higher predictability for stocks associated with higher limits-to-arbitrage, we also document that this effect is less pronounced for machine learning forecasts than for linear regression forecasts. This finding indicates that the superiority of machine learning models in emerging markets does not stem from limits to arbitrage. Finally, accounting for transaction costs, short-selling constraints, and limiting our investment universe to big stocks only, we document that machine learning-based return forecasts can lead to significant net outperformance over the market and net alphas, at least when efficient trading rules are applied.

**Figure 1:** Characteristic importance by model

This figure shows the ranked characteristic importance for the variables in each model. Characteristic importance is an average over all training samples and importance within each model is normalized to sum to one.

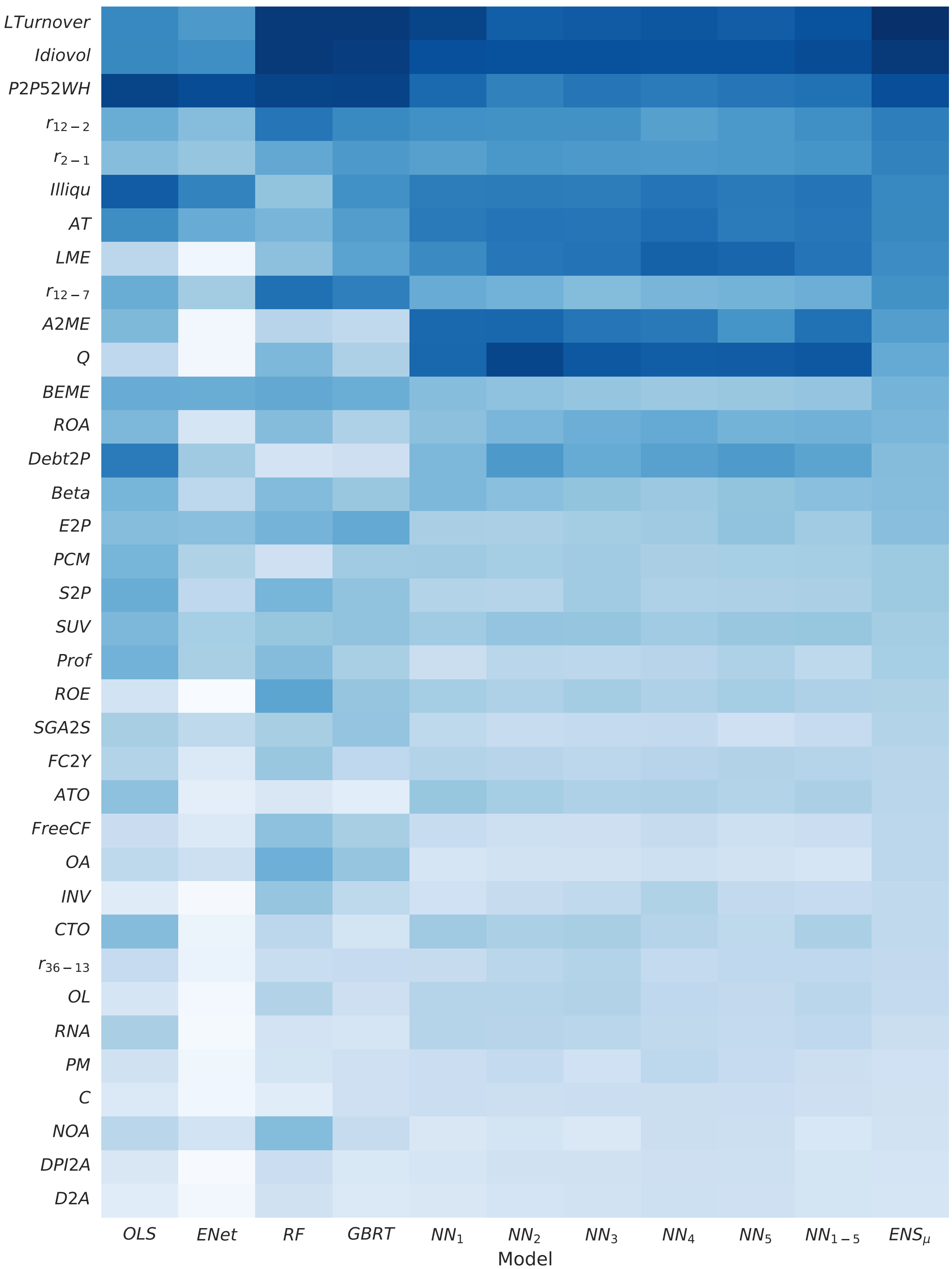


**Figure 2:** Marginal association between expected returns and characteristics

The figure shows the sensitivity of expected returns (vertical axis) to the four following individual characteristics (holding all other covariates fixed at their median values): short-term reversal ($r_{2-1}$, top-left), idiosyncratic volatility ($Idiovol$, top-right), turnover ($LTurnover$, bottom-left), and operating leverage ($OL$, bottom-right).

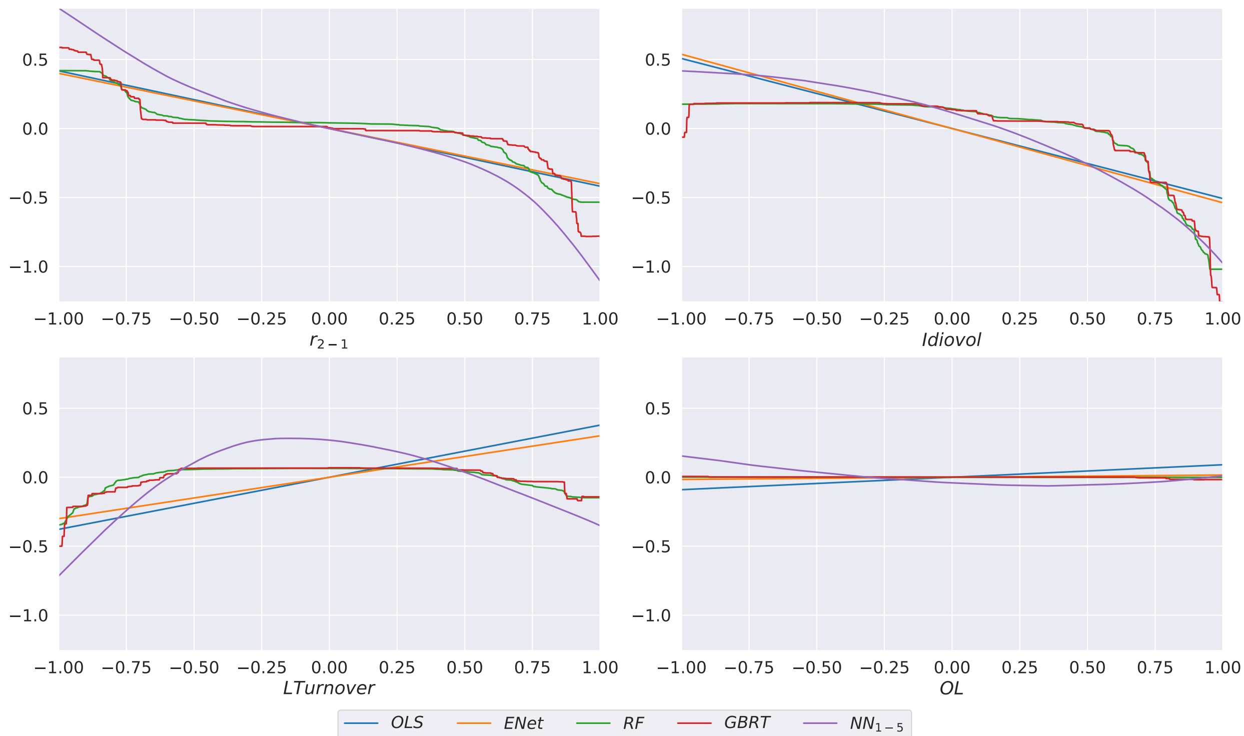


**Figure 3:** Expected returns and characteristic interactions ($NN_{1-5}$)

The figure shows the sensitivity of the expected returns (vertical axis) to interactions effects for four selected combinations in model $NN_{1-5}$ (holding all other characteristics fixed at their median values of 0): Amihud (2002) illiquidity ($Illiqu$) and short-term reversal ($r_{2-1}$) (top-left), Amihud (2002) illiquidity and market capitalization ($LME$) (top-right), idiosyncratic volatility ($Idiovol$) and short-term reversal (bottom-left), and idiosyncratic volatility and market capitalization (bottom-right).

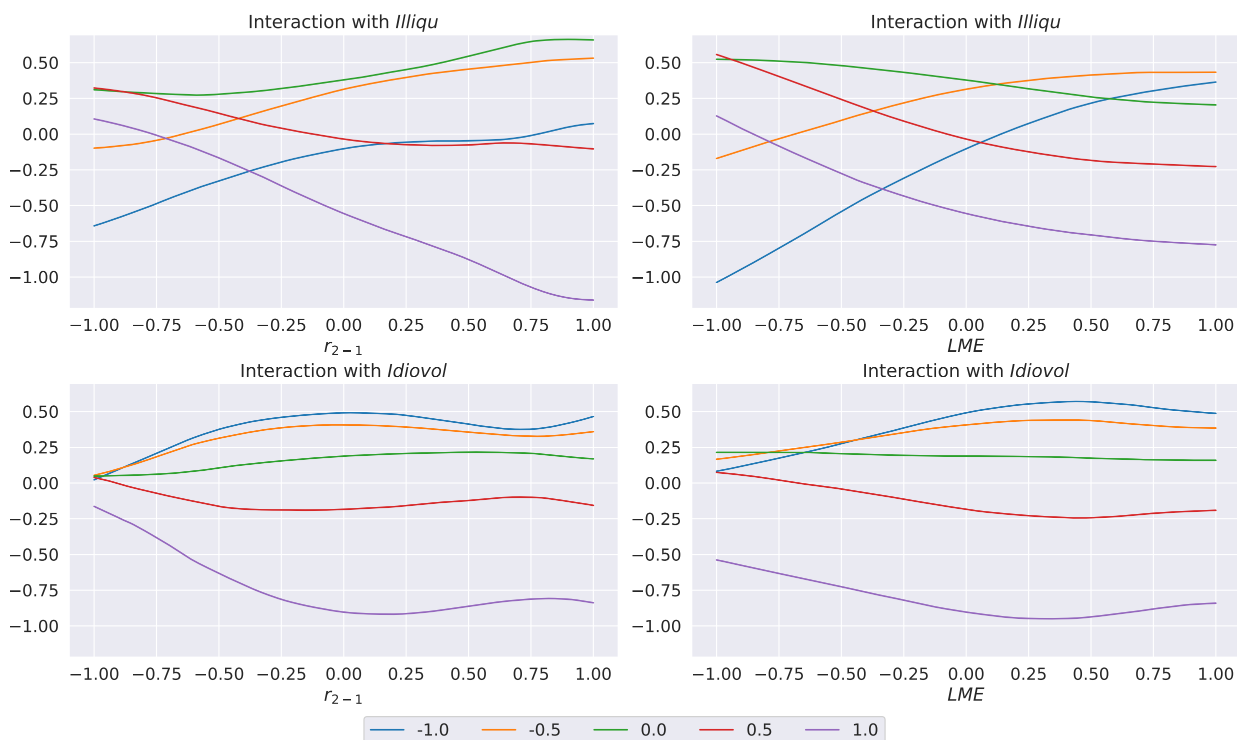


**Figure 4:** Fama and French (2018) six-factor model alphas

This figure shows the Fama and French (2018) six-factor models alphas for various machine learning long-short portfolios. Stocks are sorted into country-neutral and value-weighted quintiles based on their predicted returns for the next month. The sorting breakpoints are based on big stocks only, which are in the top 90% of a country's aggregated market capitalization. The sample period is from January 2002 to December 2021.

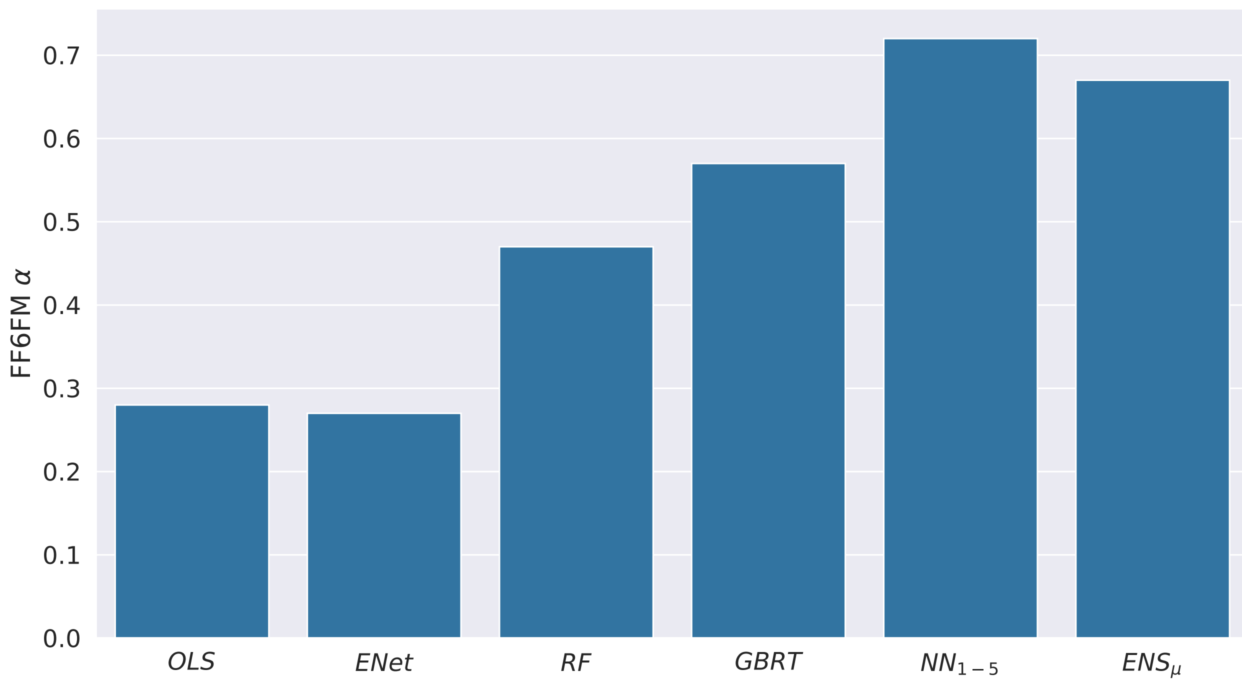
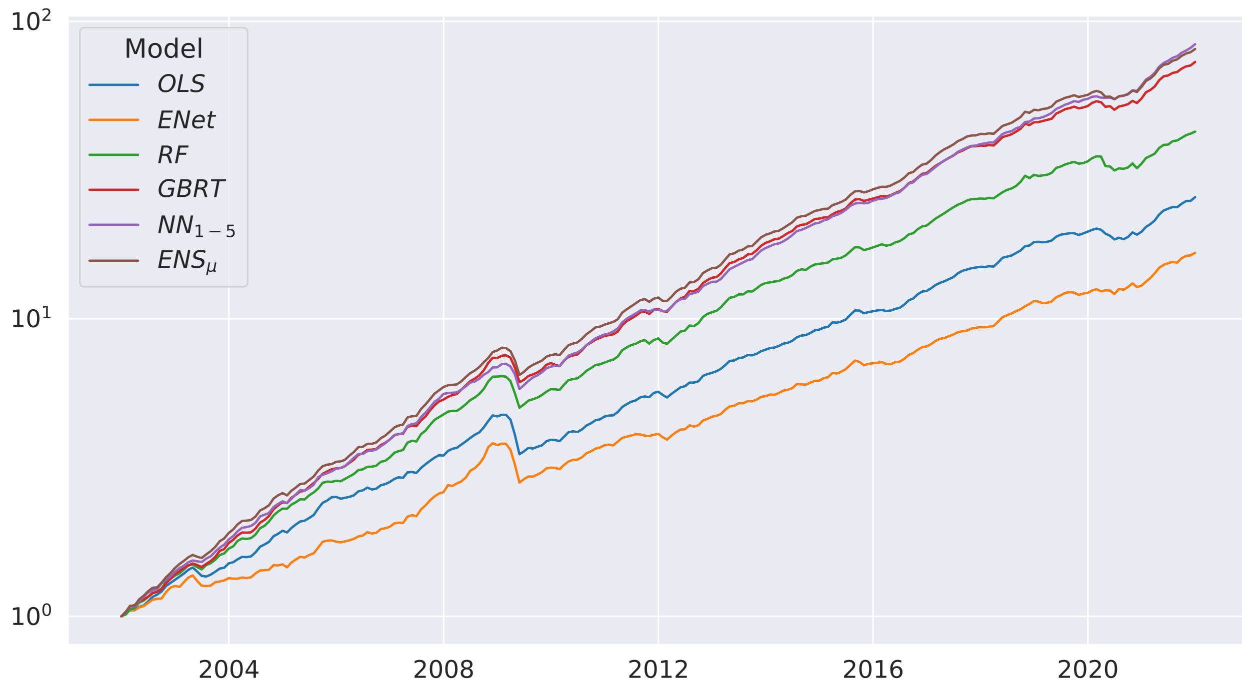


**Figure 5:** Cumulative return of machine learning portfolios

The figure shows the cumulative log returns of long-short quintile portfolios sorted on the out-of-sample machine learning return forecasts. Panel A shows equal-weighted returns, while Panel B shows value-weighted returns. The sample period is from January 2002 to December 2021.

**Panel A: Equal-Weighted**




**Panel B: Value-Weighted**

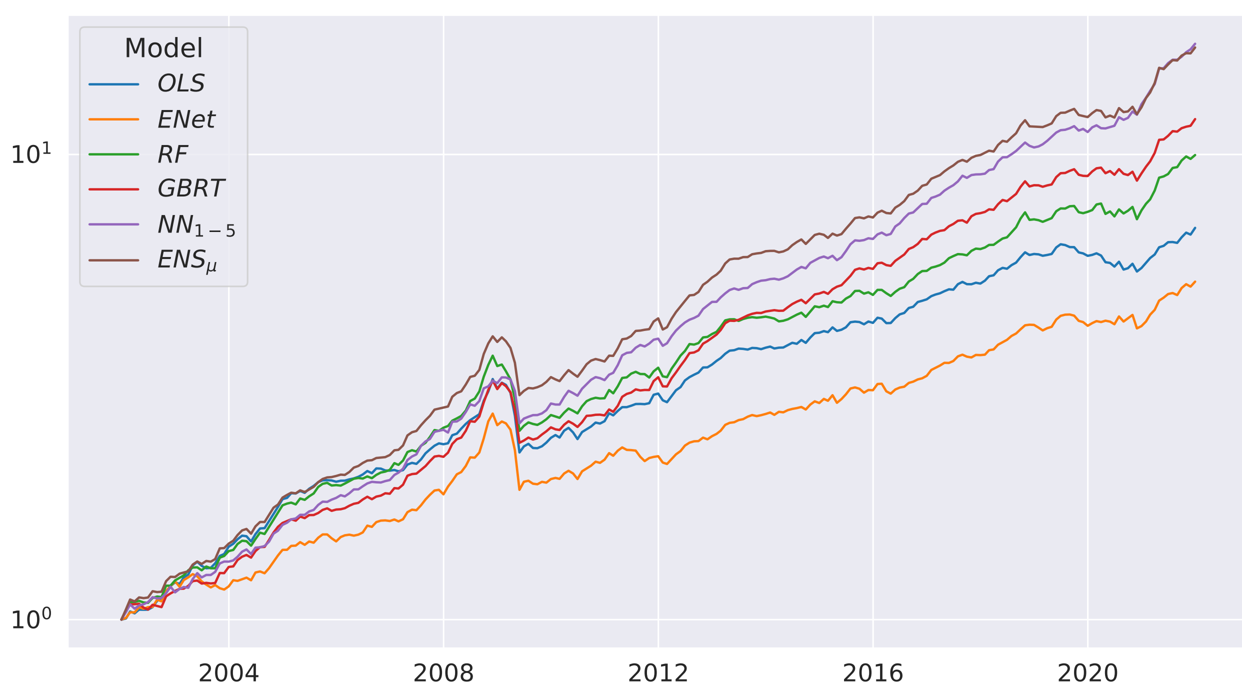


**Figure 6:** Long-horizon performance of machine learning forecasts

This figure shows the average cumulative risk-adjusted return of the different machine learning long-short portfolios. First, we obtain the return of the portfolio formed at the end of month t for month t+k, where k $\in \{1, ..., 36\}$. Second, we run a time-series regression with the Fama and French (2018) six-factor model for the corresponding months. The regression intercept is defined as the average risk-adjusted portfolio return for the long–short portfolio at month t + k. In the final step, we compute the average holding period (cumulative) risk-adjusted return for the next k months since formation as the sum over the previous k months.

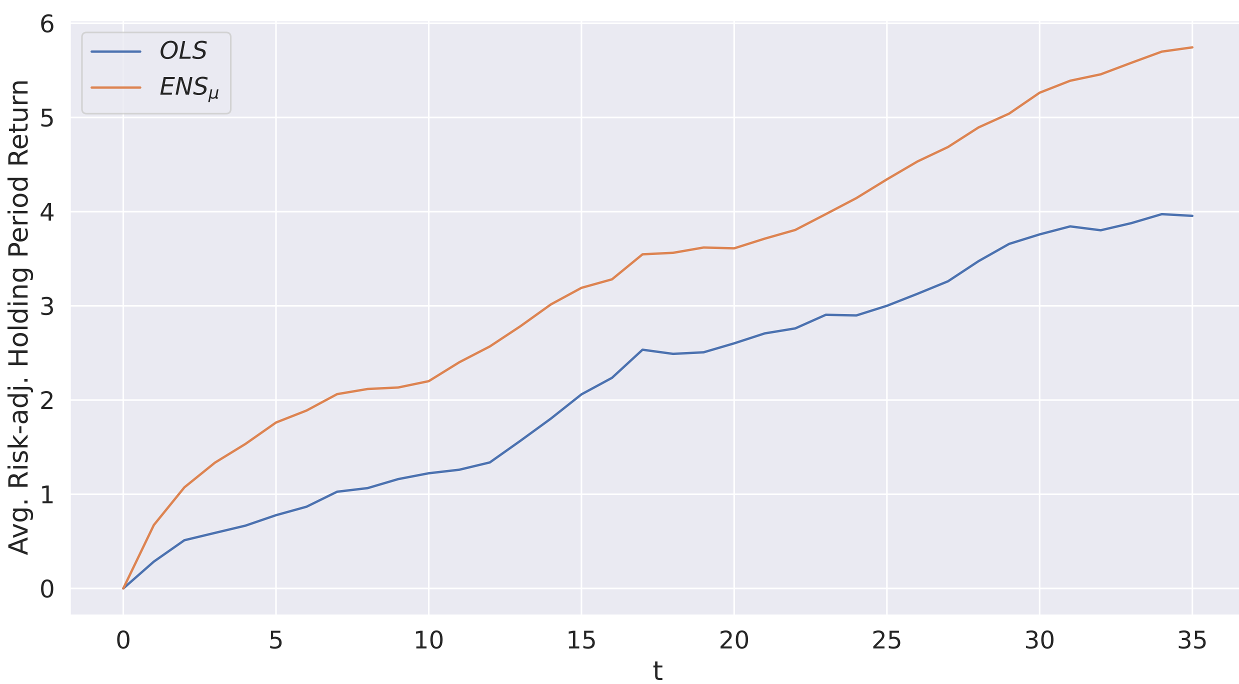


**Table 1:** Summary statistics by country

The table presents summary statistics for the 32 countries of our sample. Column 1 reports the country names, and Columns 2, 3, 4, and 5 report the total, minimum, mean, and maximum number of firms per country. Columns 6 and 7 state the average mean and median size per country-month. Column 8 shows the average total size per country-month and column 9 reports these values in percentage of the respective total across countries. Size is measured as market capitalization in million USD. The last two columns report the actual beginning and ending dates during which each country is included in our sample.

| Country | Total no. firms | Min no. firms | Mean no. firms | Max no. firms | Mean size | Median size | Total size | % of total size | Start date | End date |
|---|---|---|---|---|---|---|---|---|---|---|
| Argentina | 96 | 11 | 30 | 45 | 832 | 376 | 25467 | 1.12 | 91-05 | 21-12 |
| Brazil | 289 | 17 | 65 | 154 | 3121 | 1500 | 247409 | 4.08 | 94-09 | 21-12 |
| Chile | 201 | 53 | 74 | 102 | 1740 | 826 | 125716 | 4.26 | 90-07 | 21-12 |
| China | 28 | 10 | 16 | 24 | 2772 | 1292 | 45063 | 0.10 | 16-05 | 21-12 |
| Colombia | 50 | 14 | 19 | 25 | 2791 | 2178 | 53710 | 0.95 | 94-07 | 21-12 |
| Czechia | 89 | 10 | 36 | 77 | 662 | 234 | 13541 | 0.24 | 97-07 | 05-08 |
| Egypt | 199 | 51 | 81 | 123 | 577 | 213 | 46194 | 0.59 | 01-07 | 21-12 |
| Greece | 334 | 37 | 90 | 224 | 471 | 178 | 39442 | 1.56 | 90-07 | 21-12 |
| Hungary | 41 | 10 | 12 | 22 | 1921 | 486 | 22053 | 0.44 | 97-07 | 21-12 |
| India | 2238 | 356 | 593 | 893 | 1242 | 334 | 788478 | 13.09 | 94-07 | 21-12 |
| Indonesia | 649 | 35 | 150 | 296 | 1019 | 324 | 181003 | 4.50 | 90-07 | 21-12 |
| Israel | 634 | 173 | 245 | 331 | 270 | 60 | 62659 | 1.31 | 95-07 | 10-06 |
| Jordan | 161 | 10 | 98 | 119 | 328 | 70 | 32467 | 0.08 | 06-04 | 09-06 |
| Korea | 2972 | 394 | 803 | 1343 | 622 | 134 | 572232 | 12.87 | 92-07 | 21-12 |
| Kuwait | 81 | 73 | 75 | 78 | 1504 | 396 | 113350 | 0.02 | 21-07 | 21-12 |
| Malaysia | 1173 | 177 | 389 | 534 | 641 | 154 | 242228 | 9.97 | 90-07 | 21-12 |
| Mexico | 181 | 25 | 54 | 70 | 2692 | 1316 | 156068 | 4.92 | 90-07 | 21-12 |
| Morocco | 57 | 24 | 31 | 37 | 1323 | 658 | 42758 | 0.41 | 01-07 | 14-06 |
| Pakistan | 362 | 73 | 139 | 205 | 190 | 68 | 28089 | 0.50 | 94-07 | 21-12 |
| Peru | 103 | 17 | 28 | 40 | 1100 | 643 | 31521 | 0.61 | 94-07 | 21-12 |
| Philippines | 270 | 23 | 80 | 113 | 1087 | 433 | 101378 | 2.70 | 90-07 | 21-12 |
| Poland | 591 | 26 | 135 | 232 | 653 | 143 | 100667 | 1.74 | 95-07 | 21-12 |
| Portugal | 99 | 35 | 51 | 60 | 394 | 155 | 18182 | 0.84 | 90-07 | 98-06 |
| Qatar | 32 | 25 | 27 | 29 | 4969 | 2880 | 134437 | 0.42 | 14-07 | 21-12 |
| Russia | 232 | 10 | 54 | 102 | 4711 | 1958 | 285849 | 3.77 | 98-07 | 21-12 |
| Saudi Arabia | 89 | 34 | 52 | 84 | 8189 | 4936 | 398893 | 0.36 | 19-07 | 21-12 |
| South Africa | 517 | 80 | 131 | 240 | 2445 | 1101 | 274534 | 6.31 | 95-07 | 21-12 |
| Sri Lanka | 150 | 88 | 98 | 105 | 15 | 7 | 1530 | 0.03 | 94-07 | 01-06 |
| Taiwan | 1912 | 339 | 743 | 978 | 772 | 223 | 596406 | 11.36 | 97-07 | 21-12 |
| Thailand | 823 | 133 | 237 | 387 | 728 | 188 | 194393 | 6.48 | 90-07 | 21-12 |
| Turkey | 430 | 50 | 134 | 237 | 862 | 243 | 125222 | 3.77 | 90-07 | 21-12 |
| UAE | 69 | 33 | 43 | 49 | 4703 | 1653 | 201327 | 0.61 | 14-07 | 21-12 |
| Global | 15152 | 594 | 3763 | 5690 | 901 | 209 | 3909693 | 100.00 | 90-07 | 21-12 |

**Table 2:** Monthly out-of-sample stock-level prediction performance

This table summarizes the monthly out-of-sample stock-level prediction performance using OLS ($OLS$), elastic net ($ENet$), random forest ($RF$), gradient boosted regression trees ($GBRT$), neural networks with 1 to 5 layers ($NN_1$–$NN_5$), an ensemble of the different neural networks ($NN_{1-5}$), and an ensemble of the different non-linear machine learning algorithms ($ENS_\mu$). Panel A reports the monthly $R^2_{OOS}$ statistics for the full sample and within subsamples that include only large stocks or small stocks. Panel B reports pairwise Newey and West (1987) adjusted Diebold-Mariano test statistics comparing the out-of-sample stock-level prediction performance among each machine learning model. Positive numbers indicate the column model outperforms the row model. Bold font indicates the difference is significant at 1% level or better for individual tests, and an asterisk indicates significance at the 1% level for 10-way comparisons via our conservative Bonferroni adjustment. The out-of-sample period is from January 2002 to December 2021.

| | $OLS$ | $ENet$ | $RF$ | $GBRT$ | $NN_1$ | $NN_2$ | $NN_3$ | $NN_4$ | $NN_5$ | $NN_{1-5}$ | $ENS_\mu$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Panel A: Percentage $R^2_{OOS}$** | | | | | | | | | | | |
| Full Sample | 0.29 | 0.18 | 0.40 | 0.52 | 0.49 | 0.53 | 0.53 | 0.55 | 0.54 | 0.60 | 0.60 |
| Large firms | 0.12 | -0.01 | 0.25 | 0.30 | 0.19 | 0.24 | 0.27 | 0.31 | 0.31 | 0.34 | 0.38 |
| Small firms | 0.40 | 0.29 | 0.49 | 0.66 | 0.67 | 0.70 | 0.68 | 0.70 | 0.68 | 0.75 | 0.73 |
| **Panel B: Between-model comparison of predictive performance** | | | | | | | | | | | |
| $OLS$ | | -2.13 | **3.45*** | **6.35*** | **5.57*** | **5.42*** | **5.64*** | **6.05*** | **6.65*** | **7.34*** | **8.25*** |
| $ENet$ | | | **4.53*** | **6.50*** | **5.91*** | **6.08*** | **6.29*** | **7.21*** | **7.01*** | **7.68*** | **8.19*** |
| $RF$ | | | | **6.80*** | **2.96** | **3.96*** | **4.38*** | **5.27*** | **4.57*** | **6.59*** | **12.65*** |
| $GBRT$ | | | | | -0.72 | 0.71 | 0.68 | 1.74 | 1.00 | **3.49*** | **7.59*** |
| $NN_1$ | | | | | | 2.51 | 2.12 | **3.24** | **2.85** | **9.19*** | **4.38*** |
| $NN_2$ | | | | | | | -0.23 | 1.69 | 0.25 | **6.01*** | 2.37 |
| $NN_3$ | | | | | | | | 1.82 | 0.39 | **5.86*** | **2.82** |
| $NN_4$ | | | | | | | | | -1.42 | **3.12** | 1.45 |
| $NN_5$ | | | | | | | | | | **5.30*** | **2.60** |
| $NN_{1-5}$ | | | | | | | | | | | -0.45 |

**Table 3:** Drawdowns, turnover, and risk-adjusted performance of machine learning portfolios

This table reports the out-of-sample performance of the different machine learning long-short portfolios. Stocks are sorted into country-neutral quintiles based on their predicted returns for the next month. The sorting breakpoints are based on big stocks only, which are in the top 90% of a country's aggregated market capitalization. Panel A (Panel D) summarizes the quintile sort results from equal-weighting (value-weighting) and provides the predicted monthly returns for the long-short quintile (Pred), the average monthly returns of the long-short quintile (Avg), Newey and West (1987) adjusted $t$-statistics with 4 lags ($t$-stat), their standard deviations (SD), and annualized Sharpe ratios (SR), respectively. Panel B (Panel E) reports the average Fama and French (2018) six-factor model alpha ($\alpha_{FF6}$), corresponding Newey and West (1987) adjusted t-statistics with 4 lags ($t$-stat$_\alpha$), and corresponding $R^2$ using equal-weighting (value-weighting). Panel C (Panel F) describes the maximum drawdowns (Max DD), the most negative monthly return (Max 1M Loss), and the average monthly turnover in % of the equal-weighted (value-weighted) long-short portfolio. The sample period is from January 2002 to December 2021.

| | $OLS$ | $ENet$ | $RF$ | $GBRT$ | $NN_1$ | $NN_2$ | $NN_3$ | $NN_4$ | $NN_5$ | $NN_{1-5}$ | $ENS_\mu$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Panel A: Quntile sorts performance - Equal-weighted** | | | | | | | | | | | |
| Pred | 1.93 | 1.97 | 1.50 | 1.80 | 2.61 | 2.60 | 2.41 | 2.29 | 2.25 | 2.30 | 1.71 |
| Avg | 1.38 | 1.20 | 1.60 | 1.82 | 1.89 | 1.91 | 1.84 | 1.86 | 1.85 | 1.88 | 1.86 |
| $t$-stat | 7.82 | 6.83 | 9.33 | 11.57 | 14.01 | 15.75 | 14.81 | 13.82 | 13.50 | 13.50 | 11.79 |
| SD | 2.04 | 2.12 | 2.04 | 1.88 | 1.68 | 1.58 | 1.60 | 1.66 | 1.69 | 1.72 | 1.87 |
| SR | 2.34 | 1.96 | 2.71 | 3.35 | 3.91 | 4.21 | 4.00 | 3.89 | 3.78 | 3.79 | 3.44 |
| **Panel B: Risk-adjusted performance - Equal-weighted** | | | | | | | | | | | |
| $\alpha_{FF6}$ | 0.97 | 0.83 | 1.19 | 1.40 | 1.47 | 1.55 | 1.49 | 1.48 | 1.44 | 1.46 | 1.43 |
| $t$-stat$_\alpha$ | 8.02 | 6.94 | 14.10 | 15.65 | 15.67 | 19.02 | 16.93 | 16.72 | 15.79 | 15.81 | 15.66 |
| $R^2$ | 62.42 | 55.79 | 59.81 | 60.89 | 53.21 | 48.65 | 52.70 | 54.66 | 56.15 | 55.67 | 58.17 |
| **Panel C: Drawdowns and turnover - Equal-weighted** | | | | | | | | | | | |
| Max DD (%) | 26.35 | 26.23 | 21.69 | 18.84 | 16.70 | 13.45 | 16.00 | 16.07 | 17.61 | 17.82 | 19.04 |
| Max 1M loss (%) | 13.97 | 12.96 | 10.53 | 10.70 | 9.37 | 7.65 | 10.20 | 10.17 | 10.25 | 10.75 | 10.68 |
| Turnover (%) | 89.27 | 96.38 | 89.61 | 97.39 | 101.87 | 102.02 | 100.80 | 99.21 | 99.50 | 99.72 | 95.77 |
| **Panel D: Quntile sorts performance - Value-weighted** | | | | | | | | | | | |
| Pred | 1.85 | 1.89 | 1.39 | 1.61 | 2.30 | 2.21 | 2.04 | 1.94 | 1.93 | 1.97 | 1.52 |
| Avg | 0.84 | 0.73 | 0.99 | 1.06 | 1.04 | 1.12 | 1.12 | 1.20 | 1.17 | 1.21 | 1.21 |
| $t$-stat | 4.64 | 4.01 | 5.28 | 6.14 | 7.00 | 9.47 | 7.91 | 8.35 | 8.17 | 8.55 | 7.04 |
| SD | 2.22 | 2.36 | 2.32 | 2.17 | 1.95 | 1.75 | 2.01 | 1.97 | 1.87 | 1.98 | 2.20 |
| SR | 1.31 | 1.07 | 1.48 | 1.69 | 1.85 | 2.23 | 1.93 | 2.11 | 2.17 | 2.12 | 1.91 |
| **Panel E: Risk-adjusted performance - Value-weighted** | | | | | | | | | | | |
| $\alpha_{FF6}$ | 0.28 | 0.27 | 0.47 | 0.57 | 0.57 | 0.71 | 0.66 | 0.73 | 0.71 | 0.72 | 0.67 |
| $t$-stat$_\alpha$ | 2.72 | 2.28 | 5.24 | 6.73 | 4.83 | 9.16 | 6.55 | 7.76 | 8.21 | 8.26 | 8.29 |
| $R^2$ | 68.25 | 56.39 | 67.61 | 68.50 | 52.97 | 48.50 | 51.79 | 58.50 | 59.39 | 56.86 | 67.17 |
| **Panel F: Drawdowns and turnover - Value-weighted** | | | | | | | | | | | |
| Max DD (%) | 30.49 | 31.45 | 31.07 | 26.54 | 23.63 | 15.44 | 23.19 | 21.81 | 20.81 | 20.36 | 25.28 |
| Max 1M loss (%) | 16.60 | 17.82 | 14.46 | 14.73 | 16.45 | 9.58 | 17.36 | 15.90 | 12.83 | 14.81 | 14.81 |
| Turnover (%) | 91.28 | 97.18 | 90.46 | 101.10 | 103.81 | 106.35 | 106.02 | 104.35 | 104.43 | 101.46 | 96.85 |

## Table 4: Robustness

This table reports robustness tests for the out-of-sample performance of equal- and value-weighted long-short portfolios. All stocks are sorted into country-neutral portfolios based on their predicted returns for the next month. We investigate predictions from a linear OLS model and an ensemble ($ENS$) of non-linear machine learning models ($RF$, $GBRT$, and $NN_{1-5}$). The sorting breakpoints are based on big stocks only, which are in the top 90% of a country's aggregated market capitalization. Panel A summarizes the baseline results as presented in Table 3. Panel B reports results on using decile sorts. Panel C uses an extended feature set following Hanauer and Lauterbach (2019). Panel D applies a feature selection before training the machine learning algorithms. Panel E uses predictions stemming from machine learning algorithms only trained on developed market data. Panel F excludes the high-turnover characteristics `Idiovol`, `LTurnover`, $\mathtt{r}_{2-1}$, `SUV`, `Illiqu` from the feature set. Pangel G shows the results for models trained on emerging market subregions. The first two rows of each panel provide the average monthly return of the long-short quintile (Avg), corresponding $t$-statistics ($t$), the average Fama and French (2018) six-factor alpha ($\alpha$), corresponding $t$-statistics ($t_\alpha$), and $R^2$. The next two rows show spanning alpha ($\alpha$), corresponding $t$-statistic ($t_\alpha$), and $R^2$ when regressing the long-short $ENS$ returns on $OLS$ returns and vice versa. All $t$-statistics are calculated using Newey and West (1987) adjusted standard errors with 4 lags. The sample period is from January 2002 to December 2021.

| | Equal-weighted | | | | | Value-weighted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Avg | $t$ | $\alpha$ | $t_\alpha$ | $R^2$ | Avg | $t$ | $\alpha$ | $t_\alpha$ | $R^2$ |
| **Panel A: Baseline** | | | | | | | | | | |
| $OLS$ | 1.38 | 7.82 | 0.97 | 8.02 | 62.42 | 0.84 | 4.64 | 0.28 | 2.72 | 68.25 |
| $ENS_\mu$ | 1.86 | 11.79 | 1.43 | 15.66 | 58.17 | 1.21 | 7.04 | 0.67 | 8.29 | 67.17 |
| $ENS_\mu \sim OLS$ | | | 0.73 | 9.01 | 80.28 | | | 0.49 | 7.83 | 74.66 |
| $OLS \sim ENS_\mu$ | | | -0.44 | -2.10 | 80.28 | | | -0.22 | -1.50 | 74.66 |
| **Panel B: Decile sorts** | | | | | | | | | | |
| $OLS$ | 1.84 | 10.09 | 1.41 | 11.30 | 52.39 | 1.18 | 5.91 | 0.55 | 4.66 | 62.37 |
| $ENS_\mu$ | 2.50 | 13.93 | 2.02 | 18.31 | 54.16 | 1.66 | 8.12 | 1.10 | 10.30 | 57.37 |
| $ENS_\mu \sim OLS$ | | | 1.01 | 7.10 | 71.88 | | | 0.78 | 6.00 | 56.06 |
| $OLS \sim ENS_\mu$ | | | -0.38 | -2.76 | 71.88 | | | -0.07 | -0.36 | 56.06 |
| **Panel C: Extended feature set** | | | | | | | | | | |
| $OLS$ | 1.51 | 9.51 | 1.14 | 11.05 | 52.93 | 0.87 | 5.22 | 0.36 | 3.17 | 55.86 |
| $ENS_\mu$ | 1.96 | 13.14 | 1.57 | 19.57 | 56.22 | 1.22 | 6.80 | 0.71 | 7.55 | 63.06 |
| $ENS_\mu \sim OLS$ | | | 0.69 | 9.72 | 80.98 | | | 0.45 | 5.37 | 73.16 |
| $OLS \sim ENS_\mu$ | | | -0.38 | -2.43 | 80.98 | | | -0.13 | -1.26 | 73.16 |
| **Panel D: Feature selection** | | | | | | | | | | |
| $OLS$ | 1.36 | 8.03 | 0.97 | 8.94 | 61.23 | 0.82 | 4.51 | 0.28 | 2.61 | 65.34 |
| $ENS_\mu$ | 1.83 | 12.31 | 1.43 | 17.52 | 58.83 | 1.23 | 7.36 | 0.73 | 8.71 | 63.59 |
| $ENS_\mu \sim OLS$ | | | 0.75 | 8.96 | 80.53 | | | 0.60 | 9.09 | 69.83 |
| $OLS \sim ENS_\mu$ | | | -0.50 | -2.34 | 80.53 | | | -0.29 | -1.60 | 69.83 |
| **Panel E: Trained on developed markets** | | | | | | | | | | |
| $OLS$ | 1.29 | 6.71 | 0.93 | 6.76 | 62.40 | 0.89 | 4.67 | 0.38 | 3.37 | 68.17 |
| $ENS_\mu$ | 1.67 | 10.55 | 1.23 | 10.12 | 59.97 | 1.20 | 6.64 | 0.62 | 5.31 | 61.15 |
| $ENS_\mu \sim OLS$ | | | 0.72 | 6.99 | 79.14 | | | 0.43 | 4.75 | 74.18 |
| $OLS \sim ENS_\mu$ | | | -0.50 | -3.84 | 79.14 | | | -0.15 | -1.20 | 74.18 |
| **Panel F: Excluding short-term feature set** | | | | | | | | | | |
| $OLS$ | 1.36 | 7.12 | 0.89 | 8.85 | 63.66 | 0.77 | 4.25 | 0.24 | 3.37 | 74.39 |
| $ENS_\mu$ | 1.59 | 9.13 | 1.17 | 11.62 | 59.27 | 1.00 | 5.56 | 0.46 | 5.91 | 70.92 |
| $ENS_\mu \sim OLS$ | | | 0.45 | 7.31 | 88.62 | | | 0.32 | 3.93 | 82.31 |
| $OLS \sim ENS_\mu$ | | | -0.32 | -4.99 | 88.62 | | | -0.16 | -2.18 | 82.31 |
| **Panel G: Subregional training** | | | | | | | | | | |
| $OLS$ | 1.09 | 6.29 | 0.77 | 6.23 | 53.92 | 0.78 | 4.42 | 0.29 | 2.57 | 56.92 |
| $ENS_\mu$ | 1.35 | 8.88 | 0.97 | 9.75 | 57.22 | 0.97 | 5.95 | 0.44 | 4.59 | 58.10 |
| $ENS_\mu \sim OLS$ | | | 0.49 | 7.78 | 77.56 | | | 0.28 | 4.46 | 75.77 |
| $OLS \sim ENS_\mu$ | | | -0.23 | -1.76 | 77.56 | | | -0.06 | -0.51 | 75.77 |

**Table 5:** Regional performance

This table reports the out-of-sample performance of equal- and value-weighted long-short portfolios for emerging market subregions. All stocks are sorted into country-neutral portfolios based on their predicted returns for the next month. We investigate predictions from a linear OLS model and an ensemble ($ENS$) of non-linear machine learning models ($RF$, $GBRT$, and $NN_{1-5}$). The sorting breakpoints are based on big stocks only, which are in the top 90% of the country's aggregated market capitalization. Panel A summarizes the baseline results as presented in Table 3, and Panel B shows the result for all countries being part of emerging Americas, Panel C combines all emerging Asian countries, and Panel D reports results for emerging countries from Europe, the Middle East, and Africa. The first two rows of each panel provide the average monthly return of the long-short quintile (Avg), corresponding $t$-statistics ($t$), the average Fama and French (2018) six-factor alpha ($\alpha$), corresponding $t$-statistics ($t_\alpha$), and $R^2$. The next two rows show spanning alpha ($\alpha$), corresponding $t$-statistic ($t_\alpha$), and $R^2$ when regressing the long-short $ENS$ returns on $OLS$ returns and vice versa. All $t$-statistics are calculated using Newey and West (1987) adjusted standard errors with 4 lags. The sample period is from January 2002 to December 2021.

| | Equal-weighted | | | | | Value-weighted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Avg | $t$ | $\alpha$ | $t_\alpha$ | $R^2$ | Avg | $t$ | $\alpha$ | $t_\alpha$ | $R^2$ |
| **Panel A: Emerging Markets** | | | | | | | | | | |
| $OLS$ | 1.38 | 7.82 | 0.97 | 8.02 | 62.42 | 0.84 | 4.64 | 0.28 | 2.72 | 68.25 |
| $ENS_\mu$ | 1.86 | 11.79 | 1.43 | 15.66 | 58.17 | 1.21 | 7.04 | 0.67 | 8.29 | 67.17 |
| $ENS_\mu \sim OLS$ | | | 0.73 | 9.01 | 80.28 | | | 0.49 | 7.83 | 74.66 |
| $OLS \sim ENS_\mu$ | | | -0.44 | -2.10 | 80.28 | | | -0.22 | -1.50 | 74.66 |
| **Panel B: Americas** | | | | | | | | | | |
| $OLS$ | 0.70 | 2.73 | 0.51 | 2.56 | 39.51 | 0.75 | 2.83 | 0.37 | 1.70 | 39.83 |
| $ENS_\mu$ | 0.88 | 4.06 | 0.69 | 3.90 | 25.45 | 0.85 | 3.20 | 0.57 | 2.73 | 33.47 |
| $ENS_\mu \sim OLS$ | | | 0.45 | 3.45 | 46.98 | | | 0.33 | 1.95 | 48.58 |
| $OLS \sim ENS_\mu$ | | | 0.03 | 0.15 | 46.98 | | | 0.16 | 0.82 | 48.58 |
| **Panel C: Asia** | | | | | | | | | | |
| $OLS$ | 1.46 | 7.59 | 1.13 | 9.34 | 61.82 | 0.84 | 3.95 | 0.37 | 2.93 | 66.71 |
| $ENS_\mu$ | 1.98 | 11.18 | 1.63 | 17.32 | 60.28 | 1.34 | 7.02 | 0.87 | 8.81 | 67.14 |
| $ENS_\mu \sim OLS$ | | | 0.74 | 7.97 | 79.70 | | | 0.62 | 6.85 | 75.00 |
| $OLS \sim ENS_\mu$ | | | -0.40 | -1.83 | 79.70 | | | -0.33 | -1.64 | 75.00 |
| **Panel D: Europe, the Middle East and Africa** | | | | | | | | | | |
| $OLS$ | 1.12 | 6.46 | 0.96 | 6.23 | 18.96 | 0.82 | 4.00 | 0.37 | 2.00 | 26.18 |
| $ENS_\mu$ | 1.57 | 10.27 | 1.32 | 9.33 | 16.23 | 1.13 | 5.54 | 0.59 | 3.38 | 29.44 |
| $ENS_\mu \sim OLS$ | | | 0.83 | 7.42 | 51.63 | | | 0.57 | 4.21 | 46.30 |
| $OLS \sim ENS_\mu$ | | | -0.11 | -0.58 | 51.63 | | | 0.05 | 0.38 | 46.30 |

**Table 6:** Higher-risk versus lower-risk periods

This table reports the equal-weighted and value-weighted performance of the long-short prediction-sorted portfolios over the 20-year out-of-sample testing period in higher- versus lower-risk months. All stocks are sorted into country-neutral portfolios based on their predicted returns for the next month. The sorting breakpoints are based on big stocks only, which are in the top 90% of the country's aggregated market capitalization. Risk proxies are whether emerging markets as a whole go up or down ($Mkt$), the rate on long-term U.S. government bonds is going up or down ($\Delta Yield$), whether the TED spread is below or above its median value ($TED$), and whether the time-varying risk aversion index proposed by Bekaert et al. (2022) (RAbex) is below or above its median value ($RAbex$). Panel A (Panel B) summarizes the quintile sort results from equal-weighting (value-weighting). The first two rows of each panel provide the average monthly long-short returns and corresponding $t$-statistics. The next two rows show spanning alpha and corresponding $t$-statistic when regressing the long-short $ENS$ returns on $OLS$ returns and vice versa. All $t$-statistics are calculated using Newey and West (1987) adjusted standard errors with 4 lags. The sample period is from January 2002 to December 2021.

| Model | $Mkt_{up}$ | $Mkt_{down}$ | $\Delta Yield_{up}$ | $\Delta Yield_{down}$ | $TED_{high}$ | $TED_{low}$ | $RAbex_{high}$ | $RAbex_{low}$ |
|---|---|---|---|---|---|---|---|---|
| **Panel A: Equal-Weighted** | | | | | | | | |
| $OLS$ | 1.07 | 1.82 | 1.60 | 1.17 | 1.19 | 1.55 | 1.13 | 1.63 |
| | (3.82) | (11.81) | (7.25) | (6.20) | (3.60) | (9.93) | (3.55) | (12.48) |
| $ENS_\mu$ | 1.73 | 2.06 | 2.00 | 1.74 | 1.68 | 2.02 | 1.70 | 2.02 |
| | (7.01) | (14.26) | (9.52) | (9.85) | (6.07) | (12.72) | (6.15) | (14.41) |
| $ENS_\mu \sim OLS$ | 0.84 | 0.47 | 0.63 | 0.81 | 0.73 | 0.68 | 0.80 | 0.56 |
| | (7.60) | (5.43) | (6.31) | (6.60) | (7.08) | (4.95) | (8.07) | (4.25) |
| $OLS \sim ENS_\mu$ | -0.64 | 0.03 | -0.34 | -0.52 | -0.58 | -0.13 | -0.66 | 0.05 |
| | (-2.56) | (0.23) | (-1.46) | (-2.13) | (-2.38) | (-0.81) | (-2.93) | (0.53) |
| **Panel B: Value-Weighted** | | | | | | | | |
| $OLS$ | 0.64 | 1.12 | 1.06 | 0.63 | 0.69 | 0.96 | 0.74 | 0.93 |
| | (2.13) | (5.51) | (4.69) | (3.31) | (2.03) | (6.49) | (2.29) | (6.30) |
| $ENS_\mu$ | 1.12 | 1.34 | 1.40 | 1.03 | 1.12 | 1.29 | 1.07 | 1.35 |
| | (4.12) | (5.84) | (5.53) | (6.40) | (3.55) | (7.95) | (3.65) | (8.38) |
| $ENS_\mu \sim OLS$ | 0.57 | 0.41 | 0.47 | 0.52 | 0.53 | 0.45 | 0.45 | 0.48 |
| | (6.47) | (3.88) | (4.36) | (6.37) | (6.84) | (5.03) | (4.94) | (5.27) |
| $OLS \sim ENS_\mu$ | -0.37 | 0.08 | -0.13 | -0.30 | -0.44 | 0.10 | -0.27 | -0.00 |
| | (-2.01) | (0.54) | (-0.71) | (-1.78) | (-3.35) | (0.75) | (-1.55) | (-0.00) |

## **Table 7:** Market integration

This table reports summary statistics for regressions of emerging market regions' long-short returns on developed market's long-short returns. The long-short returns are based on ensemble ($ENS$) return forecasts of non-linear machine learning models ($RF$, $GBRT$, and $NN_{1-5}$) separately estimated for emerging and developed markets. All stocks are sorted into country-neutral quintile portfolios based on their predicted returns for the next month. The sorting breakpoints are based on big stocks only, which are in the top 90% of a country's aggregated market capitalization. Panel A (Panel B) summarizes the results of equal-weighting (value-weighting) of the prediction-sorted portfolios based on the different regional subsets. Each Panel provides the average monthly return of the long-short quintile (Avg), the alpha ($\alpha$), beta ($\beta$), their corresponding $t$-statistics, and $R^2$ with respect to the developed market ensemble machine-learning factor. All $t$-statistics are calculated using Newey and West (1987) adjusted standard errors with 4 lags. The sample period is from January 2002 to December 2021.

| | Avg | $t$ | $\alpha$ | $t_\alpha$ | $\beta$ | $t_\beta$ | $R^2$ |
|---|---|---|---|---|---|---|---|
| **Panel A: Equal-weighted** | | | | | | | |
| $Global_{EM}$ | 1.86 | 11.79 | 1.39 | 9.24 | 0.50 | 7.01 | 32.34 |
| $AME_{EM}$ | 0.88 | 4.06 | 0.35 | 1.73 | 0.56 | 5.57 | 19.44 |
| $ASIA_{EM}$ | 1.98 | 11.18 | 1.47 | 8.66 | 0.55 | 6.71 | 28.42 |
| $EMEA_{EM}$ | 1.57 | 10.27 | 1.24 | 7.39 | 0.36 | 4.64 | 12.78 |
| **Panel B: Value-weighted** | | | | | | | |
| $Global_{EM}$ | 1.21 | 7.04 | 0.89 | 5.73 | 0.49 | 4.95 | 28.17 |
| $AME_{EM}$ | 0.85 | 3.20 | 0.53 | 2.30 | 0.49 | 3.68 | 14.37 |
| $ASIA_{EM}$ | 1.34 | 7.02 | 0.98 | 4.92 | 0.54 | 3.95 | 22.47 |
| $EMEA_{EM}$ | 1.13 | 5.54 | 0.87 | 4.36 | 0.40 | 4.55 | 10.73 |

**Table 8:** Limits to abitrage

This table reports the results of a Fama and MacBeth (1973) regression of future returns on machine learning return forecasts (ML), proxies for limits to arbitrage, and their interaction. Each month, we run a cross-sectional regression of excess stock returns in month t + 1 on a firm's ML value and on interaction terms between ML and proxies for limits to arbitrage constructed at the end of the previous month t. The proxies for limits to arbitrage are: -1 × market capitalization ($SIZE$), idiosyncratic volatility ($IVOL$), Amihud illiquidity ($ILLIQ$), and a combination of the different proxies. All proxies for limits to arbitrage are ranked into the [-1,1] interval for each month and country. The $t$-statistics in parentheses are the corresponding Newey and West (1987) adjusted $t$-statistics with 4 lags. The sample period is from January 2002 to December 2021.

| | $SIZE$ | | $IVOL$ | | $ILLIQ$ | | $COMBO$ | |
|---|---|---|---|---|---|---|---|---|
| | ( 1 ) | ( 2 ) | ( 3 ) | ( 4 ) | ( 5 ) | ( 6 ) | ( 7 ) | ( 8 ) |
| $OLS$ | 0.72 | | 0.72 | | 0.74 | | 0.74 | |
| | (9.43) | | (9.16) | | (9.55) | | (9.27) | |
| $ENS_{\mu}$ | | 1.11 | | 1.13 | | 1.15 | | 1.13 |
| | | (13.81) | | (14.31) | | (14.58) | | (12.88) |
| $LTA \times ML$ | 0.27 | 0.19 | 0.48 | 0.21 | 0.01 | -0.06 | 0.52 | 0.27 |
| | (5.36) | (3.27) | (9.65) | (3.85) | (0.23) | (-0.90) | (7.36) | (2.80) |
| $LTA$ | 0.15 | 0.12 | 0.06 | 0.12 | 0.21 | 0.19 | 0.27 | 0.26 |
| | (2.42) | (1.93) | (0.68) | (1.45) | (2.78) | (2.48) | (3.06) | (2.94) |
| Country | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adj. $R^2$ (%) | 15.00 | 15.22 | 15.10 | 15.30 | 15.10 | 15.34 | 15.01 | 15.23 |
| Avg. Obs | 4419 | 4419 | 4419 | 4419 | 4419 | 4419 | 4419 | 4419 |

## **Table 9:** Further investment frictions

This table reports results for returns on different buy/hold long-only strategies before and after transaction-cost. We report the strategies' gross returns in excess of the market, average turnover, transaction costs, net returns in excess of the market, and net Fama and French (2018) six-factor models alphas. We estimate one-way transaction costs as half of a stock's bid-ask spread, estimated as in Ardia et al. (2022). All $t$-statistics are Newey and West (1987) adjusted with 4 lags. Panel A summarizes results from equal-weighting, while Panel B shows results from value-weighting. The sample period is from January 2002 to December 2021.

| | $OLS$ | | $ENS_{\mu}$ | |
|---|---|---|---|---|
| | 20%/20% | 10%/30% | 20%/20% | 10%/30% |
| **Panel A: Equal-weighted** | | | | |
| $r^e_{gross} - Mkt$ | 0.49 | 0.46 | 0.78 | 0.79 |
| | (5.46) | (5.33) | (8.07) | (7.42) |
| TO (in %) | 44.29 | 24.86 | 45.20 | 27.53 |
| T-cost (in %) | 0.31 | 0.18 | 0.32 | 0.20 |
| $r^e_{net} - Mkt$ | 0.19 | 0.28 | 0.46 | 0.59 |
| | (2.11) | (3.30) | (4.88) | (5.63) |
| $\alpha^{FF6}_{net}$ | 0.29 | 0.39 | 0.55 | 0.67 |
| | (4.91) | (6.19) | (7.66) | (8.28) |
| **Panel B: Value-weighted** | | | | |
| $r^e_{gross} - Mkt$ | 0.32 | 0.32 | 0.47 | 0.48 |
| | (3.39) | (3.47) | (4.88) | (4.66) |
| TO (in %) | 44.48 | 22.16 | 45.51 | 23.40 |
| T-cost (in %) | 0.25 | 0.13 | 0.26 | 0.14 |
| $r^e_{net} - Mkt$ | 0.07 | 0.19 | 0.21 | 0.34 |
| | (0.75) | (2.08) | (2.20) | (3.31) |
| $\alpha^{FF6}_{net}$ | 0.06 | 0.17 | 0.22 | 0.34 |
| | (1.26) | (3.15) | (3.91) | (5.18) |

# Appendix A - Filter Datastream

## Constituent lists

Datastream comprises three types of constituent lists: (1) research lists, (2) Worldscope lists, and (3) dead lists. By using dead lists, we ensure that any survivorship bias is obviated. For each country, we use the union of all available lists and eliminate any duplicates. As a result, one list remains for each country to be used in the subsequent static filter process. Table A.1 provides an overview of the constituent lists for emerging markets that are used in our study.

[Table A.1 about here.]

## Static screens

We restrict our sample to common equity stocks by applying several static screens, as shown in Table A.2. Screens (1) to (7) are straightforward to apply and common in the literature.

[Table A.2 about here.]

Screen (8) relates to, among others, to work by the following: Ince and Porter (2006), Campbell et al. (2010), Griffin et al. (2010), Karolyi et al. (2012). The authors provide generic filter rules to exclude non-common equity securities from Refinitiv Datastream. We apply the identified keywords and match them with the security names provided by Datastream. A security is excluded from the sample in the event that a keyword coincides with part of the security name. The following three Datastream items store security names and are applied to the keyword filters: 'NAME', 'ENAME', and 'ECNAME'. Table A.3 gives an overview of the keywords used.

[Table A.3 about here.]

In addition, Griffin et al. (2010) introduce specific keywords for individual countries. The keywords are thus applied to the security names of single countries only. For example, German security names are parsed to contain the word 'GENUSSSCHEINE', which declares the security to be a non-common equity. In Table A.4, we give an overview of country-specific keyword deletions conducted in our study.

[Table A.4 about here.]

### Dynamic screens

For the securities remaining from the static screens above, we obtained return and market capitalization data from Datastream and accounting data from Worldscope. Several dynamic screens that are common in the literature were installed in order to account for data errors, mainly within return characteristics. The dynamic screens are shown in Table A.5.

[Table A.5 about here.]

**Table A.1:** Constituent lists: Emerging markets

The table contains the research lists, Worldscope lists and dead lists of emerging markets countries in our sample.

| Country | List | Country | List | Country | List |
|---|---|---|---|---|---|
| **Argentina** | DEADAR | **Israel** | DEADIL | **Portugal** | WSCOPEPT |
| | FARALL | | WSCOPEIS | | FPTALL |
| | WSCOPEAR | | FILALL | | DEADPT |
| **Brazil** | DEADBR | **Jordan** | DEADJO | **Qatar** | DEADQA |
| | FBRALL | | FJOALL | | FQAALL |
| | WSCOPEBR | | WSCOPEJO | | WSCOPEQA |
| **Chile** | DEADCL | **Korea** | DEADKR | **Russia** | DEADRU |
| | FCLALL | | FKRALL | | FRUSXALL |
| | WSCOPECL | | WSCOPEKO | | WSCOPERS |
| **China** | DEADCN | **Kuwait** | DEADKW | **Saudi Arabia** | DEADSA |
| | FCNALL | | FKWALL | | FSAALL |
| | WSCOPECH | | WSCOPEKW | | WSCOPESI |
| **Colombia** | DEADCO | **Malaysia** | DEADMY | **South Africa** | DEADZA |
| | FCOALL | | FACE | | FZAALL |
| | WSCOPECB | | FMYALL | | WSCOPESA |
| **Czechia** | DEADCZ | | WSCOPEMY | **Sri Lanka** | DEADLK |
| | FCZALL | **Mexico** | DEADMX | | FLKALL |
| | WSCOPECZ | | FMXALL | | WSCOPECY |
| **Egypt** | DEADEG | | WSCOPEMX | **Taiwan** | DEADTW |
| | FEGALL | **Morocco** | DEADMA | | FROCOALL |
| | WSCOPEEY | | FMAALL | | FTWALL |
| **Greece** | DEADGR | | WSCOPEMC | | WSCOPETA |
| | FGRALL | **Pakistan** | DEADPK | **Thailand** | DEADTH |
| | WSCOPEGR | | FPKALL | | FTHALL |
| **Hungary** | DEADHU | | WSCOPEPK | | WSCOPETH |
| | FHUALL | **Peru** | DEADPE | **Turkey** | DEADTR |
| | WSCOPEHN | | FPEALL | | FTRALL |
| **India** | DEADIN | | WSCOPEPE | | WSCOPETK |
| | FINALL | **Philippines** | DEADPH | **UAE** | DEADAE |
| | FINCONS | | FPHALL | | FAEALL |
| | FXBOMALL | | WSCOPEPH | | FXADSALL |
| | FXNSEALL | **Poland** | DEADPL | | FXDFMALL |
| | WSCOPEIN | | FPLALL | | WSCOPEAE |
| **Indonesia** | DEADID | | FPOLCM | | |
| | FIDALL | | WSCOPEPO | | |
| | WSCOPEID | | | | |

**Table A.2: Static screens**

The table displays the static screens applied in our study, mainly following Ince and Porter (2006), Schmidt et al. (2017) and Griffin et al. (2010). Column 3 lists the Datastream items involved (on the left of the equals sign) and the values which we set them to in the filter process (to the right of the equals sign). Column 4 indicates the source of the screens.

| Nr. | Description | Datastream item(s) involved | Source |
|---|---|---|---|
| (1) | For firms with more than one security, only the one with the biggest market capitalization and liquidity is used. | MAJOR = Y | Schmidt et al. (2017) |
| (2) | The type of security must be equity. | TYPE = EQ | Ince and Porter (2006) |
| (3) | Only the primary quotations of a security are analyzed. | ISINID = P | Fong et al. (2017) |
| (4) | Firms are located in the respective domestic country. | GEOGN = country shortcut | Ince and Porter (2006) |
| (5) | Securities are listed in the respective domestic country. | GEOLN = country shortcut | Griffin et al. (2010) |
| (6) | Securities whose quoted currency is different to the one of the associated country are disregarded.[a] | PCUR = currency shortcut of the country | Griffin et al. (2010) |
| (7) | Securities whose ISIN country code is different to the one of the associated country are disregarded.[b] | GGISN = country shortcut | Annaert et al. (2013) |
| (8) | Securities whose name fields indicate non-common stock affiliation are disregarded. | NAME, ENAME, ECNAME | Ince and Porter (2006), Campbell et al. (2010), Griffin et al. (2010) and Karolyi et al. (2012) |

[a] In this filter rule, the respective pre-euro currencies are also accepted for countries within the euro-zone. Moreover, in Russia 'USD' is accepted as currency, in addition to 'RUB'.
[b] In Hong Kong, ISIN country codes equal to 'BM' or 'KY' and in the Czech Republic ISIN country codes equal to 'CS' are also accepted.

**Table A.3: Generic keyword deletions**
The table reports generic keywords searched for in the names of all stocks of all countries. If a harmful keyword is detected as part of the name of a stock, the respective stock is removed from the sample.

| Non-common equity | Keywords |
|---|---|
| **Duplicates** | 1000DUPL, DULP, DUP, DUPE, DUPL, DUPLI, DUPLICATE, XSQ, XETa |
| **Depository receipts** | ADR, GDR |
| **Preferred stock** | PF, ’PF’, PFD, PREF, PREFERRED, PRF |
| **Warrants** | WARR, WARRANT, WARRANTS, WARRT, WTS, WTS2 |
| **Debt** | %, DB, DCB, DEB, DEBENTURE, DEBENTURES, DEBT |
| **Unit trusts** | .IT, .ITb, TST, INVESTMENT TRUST, RLST IT, TRUST, TRUST UNIT, TRUST UNITS, TST, TST UNIT, TST UNITS, UNIT, UNIT TRUST, UNITS, UNT, UNT TST, UT |
| **ETFs** | AMUNDI, ETF, INAV, ISHARES, JUNGE, LYXOR, X-TR |
| **Expired securities** | EXPD, EXPIRED, EXPIRY, EXPY |
| **Miscellaneous (mainly taken from Ince and Porter (2006))** | ADS, BOND, CAP.SHS, CONV, DEFER, DEP, DEPY, ELKS, FD, FUND, GW.FD, HI.YIELD, HIGH INCOME, IDX, INC.&GROWTH, INC.&GW, INDEX, LP, MIPS, MITS, MITT, MPS, NIKKEI, NOTE, OPCVM, ORTF, PARTNER, PERQS, PFC, PFCL, PINES, PRTF, PTNS, PTSHP, QUIBS, QUIDS, RATE, RCPTS, REAL EST, RECEIPTS, REIT, RESPT, RETUR, RIGHTS, RST, RTN.INC, RTS, SBVTG, SCORE, SPDR, STRYPES, TOPRS, UTS, VCT, VTG.SAS, XXXXX, YIELD, YLD |

**Table A.4: Country-specific keyword deletions**
The table reports country-specific keywords searched for in the names of all stocks of the respective countries. If a harmful keyword is detected as part of the name of a stock, the respective stock is removed from the sample.

| Country | Keywords |
|---|---|
| **Brazil** | PN, PNA, PNB, PNC, PND, PNE, PNF, PNG, RCSA, RCTB |
| **Greece** | PR |
| **Indonesia** | FB DEAD, FOREIGN BOARD |
| **Israel** | P1, 1, 5 |
| **Korea** | 1P |
| **Mexico** | ’L’, ’C’ |
| **Peru** | INVERSION, INVN, INV |
| **Philippines** | PDR |
| **South Africa** | N’, OPTS\\., CPF\\., CUMULATIVE PREFERENCE |

**Table A.5: Dynamic screens**

The table displays the dynamic screens applied to the data in our study, following Ince and Porter (2006), Griffin et al. (2010), Jacobs (2016) and Schmidt et al. (2017). Column 3 lists the respective Datastream items. Column 4 refers to the source of the screens.

| Nr. | Description | Datastream item(s) involved | Source |
|---|---|---|---|
| (1) | We delete the zero returns at the end of the return time-series that exist because in the case of a delisting, Datastream displays stale prices from the date of delisting until the end of the respective time-series. We also delete the associated market capitalizations. | RI, MV | Ince and Porter (2006) |
| (2) | We delete the associated returns and market capitalizations in case of abnormal prices (unadjusted prices $>$ 1000000). | RI, MV, UP | The screen originally stems from Schmidt et al. (2017), however we employ it on unadjusted price. |
| (3) | We delete monthly (daily) returns and the associated market capitalizations if returns exceed 990% (200%). | RI, MV | Griffin et al. (2010); Schmidt et al. (2017) |
| (4) | We delete monthly returns and the associated market capitalizations in the case of strong return reversals, defined as $(1+r_{t-1})(1+r_t)-1 < 0.5$ given that either $r_{t-1}$ or $r_t \geq 3.0$. | RI, MV | Ince and Porter (2006) |
| (5) | We delete daily returns and the associated market capitalizations in the case of strong return reversals, defined as $(1+r_{t-1})(1+r_t)-1 < 0.2$ with $r_{t-1}$ or $r_t \geq 1.0$. | RI, MV | Griffin et al. (2010); Jacobs (2016) |
| (6) | We delete observations of stocks that show non-zero price changes in less than 50% of the traded months in the previous 12 months. | RI, MV | Griffin et al. (2011) |
| (7) | We delete observations of stocks in the lowest 3% of a country's aggregated market capitalization. | MV | Hanauer and Lauterbach (2019) |

# Appendix B - Characteristics definition

This section outlines the construction of characteristic variables that we use in the paper. For each characteristic, we give the respective Datastream and Worldscope items in parentheses, the category (past returns, investment, profitability, intangibles, value, or trading frictions) and frequency (monthly vs. yearly), plus the relevant reference. As described in Section 2, we use balance-sheet data from December in year t-1 for the stock returns from July of year t to June of year t + 1 as in Fama and French (1993).

**A2ME (assets-to-market), Value, Yearly** Assets-to-market cap is the ratio of total assets (WC02999) to market capitalization as at December t-1, as in Bhandari (1988).

**AT (total assets), Trading Frictions, Yearly** Total assets measured in USD (WC02999) as in Gandhi and Lustig (2015).

**ATO (sales-to-assets), Profitability, Yearly** As in Soliman (2008), we calculate net sales (WC01001) over lagged net operating assets. Net operating assets are defined following Hirshleifer et al. (2004) and are explained in the construction of NOA.

**BEME (book-to-market), Value, Yearly** Book-to-market is the ratio of book value of equity to market value of equity. We define the book value of equity as common equity (WC03501) plus deferred taxes (WC03263). If no deferred taxes are given, the book value of equity equals common equity (WC03501). The market value of equity is as of December t-1. See Rosenberg et al. (1985) and Davis et al. (2000).

**BEME$_\mathbf{m}$ (monthly updated book-to-market), Value, Monthly** Monthly updated book-to-market is the ratio of book value of equity to the most recent market value of equity. Book value of equity is defined as for $BEME$. The most recent market value of equity is of the end of month t to predict returns of month t+1 as in Asness (2011).

**Beta (market beta), Trading Frictions, Monthly** Following Lewellen and Nagel (2006), we calculate beta daily as the sum of the regression coefficients of daily excess returns on the local market excess return and one lag of the local market excess return for the past 12 months. We require at least 126 observations for valid beta estimates, as in Welch (2020).

**C (cash-and-short-term-investment-to-assets), Value, Yearly** The ratio of cash and short-term investments (WC02001) to total assets (WC02999), as in Palazzo (2012).

**CbOPtA (cash-based operating profits-to-asset), Profitability, Yearly** As in Ball et al. (2016), cash-based operating profits-to-asset is operating profits converted to a cash basis divided by total assets (WC02999). Following Ball et al. (2015), operating profits is net sales or revenues (WC01001) minus cost of goods sold (WC01501) minus selling, general, and administrative expenses (WC01101), excluding research and development expense (WC01201). The cash-based adjustment is the year-on-year change in deferred income (WC03262), plus change in accounts payable (WC03040), plus change in accrued expenses (WC03054 + WC03069), minus change in accounts receivable (WC02051), minus change in inventory (WC02101), minus prepaid expenses (WC02140), all divided by total assets. All changes are set to zero if missing.

**CEI (composite equity issuance), Intangibles, Monthly** Similar to Daniel and Titman (2006), we define composite equity issuance as the growth rate in the market capitalization not attributable to the total stock return $R$: $log(MC_{t-1}/MC_{t-13}) - R_{(t-13,t-1)}$. To predict the returns of month $t$, $R_{(t-13,t-1)}$ is the cumulative log return (calculated via the total return index, Datastream item RI) from month $t-13$ to month $t-1$ and $MC_{t-1}$ is the market capitalization (Datastream item MV) from the end of month $t-1$.

**CF2P (cash flow-to-price), Value, Yearly** Cash flow to price is the ratio of net cash flow from operating activities (WC04860) to the market capitalization as at December t-1, as in Lakonishok et al. (1994).

**CTO (capital turnover), Profitability, Yearly** We define capital turnover as the ratio of net sales (WC01001) to lagged total assets (WC02999), as in Haugen and Baker (1996).

**D2A (capital intensity), Intangibles, Yearly** Capital intensity is the ratio of depreciation and amortization (WC01151) over total assets (WC02999), as in Gorodnichenko and Weber (2016).

**Debt2P (leverage), Value, Yearly** Following Litzenberger and Ramaswamy (1979), debt to price is the ratio of total assets (WC02999) minus common equity (WC03501) to the market capitalization as of December t-1.

**DPI2A (ratio of change in property, plants & equipment to total assets), Investment, Yearly** Following Lyandres et al. (2007), we define the changes in PP&E and inventory as the annual change in gross property, plant, and equipment (WC02301) plus the annual change in inventory (WC02101) over lagged total assets (WC02999).

**E2P (earnings-to-price), Value, Yearly** Earnings to price is the ratio of income before extraordinary items (WC01551) to the market capitalization as at December t-1, as in Basu (1983).

**FC2Y (fixed costs-to-sales), Profitability, Yearly** As in Gorodnichenko and Weber (2016), fixed costs to sales is the sum of selling, general and administrative expenditures (WC01101) and research and development expenses (WC01201) over net sales (WC01001).

**FreeCF (cash flow-to-book), Value, Yearly** Following Hou et al. (2011), we define cash flow to book as free cash flow to book value of equity. Free cash flow is calculated as net income (WC01551) plus depreciation and amortization (WC01151) minus changes in working capital minus capital expenditure (WC04601). The book value of equity is defined in the construction of BEME.

**GP2A (gross profits-to-assets), Profitability, Yearly** Gross profits-to-assets is net sales (WC01001) minus costs of goods sold (WC01051) divided by total assets (WC02999), as in Novy-Marx (2013).

**Idiovol (idiosyncratic volatility with respect to the Fama and French (1993) three-factor model), Trading Frictions, Monthly** As in Ang et al. (2006), we define idiosyncratic volatility as the standard deviation of the residuals from a regression of excess returns on a local Fama and French (1993) three-factor model. We use one month of daily data and require at least fifteen non-missing observations.

**INV (investment), Investment, Yearly** Investment is the percentage year-to-year growth rate of total assets (WC02999) following Cooper et al. (2008).

**LME (market capitalization), Trading Frictions, Monthly** Size is a stock's market capitalization at the end of the previous month and measured in USD, as in Fama and French (1992).

**LTurnover (turnover), Trading Frictions, Monthly** Turnover is a stock's trading volume (VO) divided by its shares outstanding (NOSH) during the last month, as in Datar et al. (1998).

**NOA (net operating assets), Investment, Yearly** Following Hirshleifer et al. (2004), net operating assets are defined as the difference between operating assets and operating liabilities, scaled by lagged total assets. Operating assets are total assets (WC02999) minus cash and short-term investments (WC02001). Operating liabilities are total assets (WC02999), minus total debt (WC03255), minus minority interest (WC03426), minus preferred stock and common equity (WC03995).

**OA (operating accruals), Intangibles, Yearly** Following Sloan (1996), operating accruals are calculated as changes in working capital minus depreciation (WC01151) scaled by lagged total assets (WC02999). Changes in operating working capital are changes in current assets (WC02201) minus changes in cash and short-term investments (WC02001) minus changes in current liabilities (WC03101), plus changes in debt in current liabilities (WC03051) plus changes in income taxes payable (WC03063).

**OL (operating leverage), Intangibles, Yearly** We define operating leverage as the sum of costs of goods sold (WC01051) and selling, general, and administrative expenses (WC01101) over total assets (WC02999), as in Novy-Marx (2010).

**P2P52WH (price relative to its 52-week high), Trading Frictions, Monthly** Rel to high price is the ratio of the unadjusted stock price (UP) at the end of the previous calendar month to the past 52-weeks high, as in George and Hwang (2004).

**PCM (price-to-cost margin), Profitability, Yearly** As in Gorodnichenko and Weber (2016) and D'Acunto et al. (2018), the price-to-cost margin is net sales (WC01001) minus costs of goods sold (WC01051), divided by net sales (WC01001).

**PM (profit margin), Profitability, Yearly** As in Soliman (2008), we calculate the profit margin as operating income after depreciation or EBIT (WC18191) over sales (WC01001).

**Prof (gross profitability), Profitability, Yearly** Profitability is net sales (WC01001) minus costs of goods sold (WC01051) divided by the book value of equity, following Ball et al. (2015). The book value of equity is defined in the construction of BEME.

**Q (Tobin's Q), Value, Yearly** As in Freyberger et al. (2020), we define Tobin's Q as total assets (WC02999) plus the market capitalization as of December t-1 minus cash and short-term investments (WC02001) and minus deferred taxes (WC03263), scaled by total assets (WC02999).

**$r_{12\text{-}2}$ (momentum), Past Returns, Monthly** Momentum is the cumulative return from month t-12 to t-2 as in Fama and French (1996).

**$r_{12\text{-}7}$ (intermediate momentum), Past Returns, Monthly** Intermediate momentum is the cumulative return from t-12 to t-7 as in Novy-Marx (2012).

**$r_{2\text{-}1}$ (short-term reversal), Past Returns, Monthly** Short-term reversal is the lagged one-month return as in Jegadeesh (1990).

**$r_{36\text{-}13}$ (long-term reversal), Past Returns, Monthly** Long-term reversal is the cumulative return from t-36 to t-13 as in De Bondt and Thaler (1985).

**RNA (return on net operating assets), Profitability, Yearly** As in Soliman (2008), we calculate the return on net operating assets as the ratio of operating income after depreciation or EBIT (WC18191) to lagged net operating assets. Net operating assets are defined following Hirshleifer et al. (2004) and explained in the construction of NOA.

**ROA (return on assets), Profitability, Yearly** Following Balakrishnan et al. (2010), return-on-assets is the ratio of earnings before extraordinary items (WC01551) to lagged total assets (WC02999).

**ROE (return on equity), Profitability, Yearly** Following Haugen and Baker (1996), return-on-equity are earnings before extraordinary items (WC01551) to lagged book equity. The book value of equity is defined in the construction of BEME.

**S2P (sales-to-price), Value, Yearly** Following Lewellen (2015), sales-to-price is the ratio of net sales (WC01001) to the market capitalization as of December t-1.

**SGA2S (sales and general administrative costs to sales), Intangibles, Yearly** As in Freyberger et al. (2020), we define SG&A to sales as the ratio of selling, general and administrative expenses (WC01101) to net sales (WC01001).

**Illiqu (Amihud (2002) illiquidity), Trading Frictions, Monthly** We calculate illiquidity according to Amihud (2002) as the arithmetic mean of the following ratio for the past month: the daily absolute return divided by the product of the end-of-day stock price (UP) and the daily trading volume (VO).

**SUV (unexplained volume), Trading Frictions, Monthly** Following Garfinkel (2009), standard unexplained volume is the difference between actual volume and predicted volume in the previous month. Predicted volume comes from a regression of daily volume on a constant and the absolute values of positive and negative returns. We use two months of data to estimate the model parameters (data from t-2 and t-1) and estimate the predicted volume using data from the previous month (t-1). I require at least fifteen daily observations in the previous month. Unexplained volume is standardized by the standard deviation of the residuals from the regression.

# Appendix C - Methodology

## C.1 Simple linear regression

The least complex method in our analysis and most widely used in the context of empirical asset pricing is the simple linear regression model estimated via the ordinary least squares (OLS) method. We will use it as a benchmark to compare the more complex machine learning models to it. In the case of the simple linear regression, the conditional expectations $f^*(x)$ can be modeled using the following linear model:

$$f(x_{i,t,c}, \theta) = \theta^T x_{i,t,c}, \tag{17}$$

where $\theta$, $\theta^T = (\theta_1, \theta_2, ..., \theta_p) \in \mathbb{R}^p$, is the column vector of coefficients that can be estimated with OLS by minimizing the loss function:

$$L_{MSE}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (r_{i,t+1,c}^{abn} - f(x_{i,t,c}, \theta)^2, \tag{18}$$

which is also known as the Mean Squared Error (MSE). The OLS has the big advantage that it does not require any hyperparameter input from the user. Further, by minimizing the loss function $L_{MSE}$ a unique analytical solution can be extracted, which is easy to interpret as the coefficients, $\theta$ directly describe how a change in the stock characteristics affects the expected return. Additionally, if the number of observations in the underlying dataset is larger than the number of coefficients that need to be estimated, the OLS yields an efficient and unbiased estimator according to Wooldridge (2001). But if the number of characteristics approaches the number of observations in the dataset, the OLS has issues distinguishing between signal and noise. While the signal is the portion we can understand, model, and predict, noise consists of the unpredictable component of price movements. In the case of a small sample or a large number of characteristics, the OLS starts with over-fitting the coefficients to noise rather than extracting the signal. This is of particular importance in the field of asset pricing, which empirically relies on a low signal-to-noise ratio. This overfitting yields a higher in-sample performance but a poor out-of-sample performance. Further, multicollinearity between the different characteristics can lead to a fallacious interpretation of test statistics as well as misleading coefficients. Lastly, the OLS does not model or evaluate any non-linearities of the characteristics nor any potential interactions between them. Any non-linearity would have to be imputed by the user.

## C.2 Regularized regression

To avoid overfitting in the case of empirical asset pricing, the user could increase the training sample, reduce the number of characteristics used to predict future returns, or utilize regularized regression techniques that identify which characteristics are informative and omits those that are not. Classical regularized regression techniques are ridge regression, lasso regression, or elastic net. To limit the number of machine learning methods, we concentrate on the elastic net, which is a combination of ridge and lasso regression. While the different regularized regression models have the same linear functional form as the simple linear regression, they differ with respect to the loss function by adding a penalty term ($\phi_{ENet}(\theta, \lambda, \alpha)$) to it:

$$L_{ENet}(\theta, \lambda, \alpha) = L_{MSE}(\theta) + \phi_{ENet}(\theta, \lambda, \alpha). \tag{19}$$

This penalty term reduces the model's in-sample performance and increases its out-of-sample stability by shrinking the coefficients of noisy characteristics, improving the signal-to-noise ratio. The penalty function of the elastic net is defined as:

$$\phi_{ENet}(\theta, \lambda, \alpha) = (1 - \alpha)\lambda \sum_{j=1}^{P} \mid \theta_j \mid + \frac{1}{2}\alpha\lambda \sum_{j=1}^{P} \theta_j^2, \tag{20}$$

where $\lambda$, $\lambda \in \mathbb{R}^+$ defines the magnitude of shrinkage and $\alpha$, $\alpha \in \{0, ..., 1\}$ which determines the relative weight between the two penalty components of the ridge and lasso regression. In the case of $\lambda = 0$ the regularized regression models yield a simple linear regression model. The coefficients are shrunk towards zero by setting $\lambda > 0$. As these two hyperparameters have to be set by the user, we utilize our validation sample to find the optimal in-sample $\lambda$ and $\alpha$ in the first run. We determine the optimal $\theta$ in the second run using the full training and validation sample.

## C.3 Tree-based regression

Tree-based models represent the first non-parametric regression model as their structure is decided by the training data. For our return prediction, we will utilize two tree-based methods: the random forest, as well as the gradient boosted regression tree. Compared to the linear methods, one advantage of these tree methods is that the user does not have to manually add any potential non-linearities or interactions to the data as the tree methods build these by construction.

Regression trees follow the idea of sequentially partitioning the underlying data into groups that behave similarly to each other based on a selected characteristic with regard to the future return. By sequentially separating the data, the tree "grows" and new "branches" are created each time the data is split into new groups. The tree can grow to a depth of $D$ based on the user input. At each new branch, the characteristic is picked that causes the biggest separation in the data based on an optimized cut-off value.[15] As soon as the data can not be split into subgroups or the depth $D$ is reached, a "leave" is created. In asset pricing, the tree yields a return that is clustered by the underlying characteristics.

The following equation describes a tree with a depth of $D$ and $K$ leaves:

$$\begin{aligned} f(x_{i,t,c}, \theta, D, K) &= \sum_{k=1}^{K} \theta_k 1_{\{x_{i,t,c} \in C_k(D)\}} \\ \theta_k &= \frac{1}{N_k} \sum_{x_{i,t,c} \in C_k(D)} r^{abn}_{i,t+1,c}, \end{aligned} \tag{21}$$

where $D$ is the depth of the tree measured as the maximum number of separations following the longest branch, $C_k(D)$ indicates the $k$-th separation of the characteristics, $\theta_k$ is average abnormal return within the partition, and $1_{\{x_{i,t,c} \in C_k(D)\}}$ indicates if $x_{i,t,c}$ is part of $C_k(D)$. Following this methodology, a tree of depth $D$ can capture up to $D-1$ interactions. To avoid overfitting, the tree must be regularized. We follow two different approaches in our analysis.

The first regularization approach uses bootstrap aggregation, or "bagging," developed by Breiman (2001). In this approach, each of the $T$ trees starts with a share of $B$ bootstrap samples from the data and fits an individual regression tree to the bootstrapped data. Afterward, the forecasts from

[15] In our case, for each separation, the characteristic is selected that minimizes the MSE.

the individual trees are averaged. This reduces the variation in the prediction and stabilizes the prediction performance. In the case of the random forest, the trees additionally use random subsets $R$ of characteristics to grow the branches. This reduces the impact of certain dominant return characteristics and creates de-correlated trees.

The second regularization approach is "boosting." It starts by training a weak and shallow regression tree on the full training data. In the next step, a second regression tree with the same depth $D$ is trained on the residuals of the first tree. The prediction of these two trees is then averaged while the contribution of the second tree is shrunken by a factor $LR$ (learning rate), $LR \in (0, 1)$ to avoid the model overfitting the residuals. At each new step $b$, till the model reaches a total of $B$ trees, a new shallow tree is fitted to the residual, which is based on the $b-1$-th model and added to it with a shrinkage weight of $LR$.

Both regression trees share the two main hyperparameters: the number of trees in the forest $T$, $T \in \mathbb{Z}^+$ and the maximum depth $D$, $D \in \mathbb{Z}^+$. While the random forest additionally requires the share of the bootstrapped samples $B$, $0 > B \leq 1$, the gradient boosted regression tree requires a certain learning rate $LR$, $0 > B \leq 1$. These hyperparameters are optimized through the validation step. Additionally, we can provide the share $R$, $0 > R \leq 1$, of randomly selected characteristics that are used in each tree of the random forest.

## C.4 Neural networks

Neural networks are another highly flexible but opposed to the regression trees, a parametric model. While these models can have various forms, we focus on the standard structure of a feed-forward neural network. A feed-forward neural network consists of an "input" layer of input characteristics and the intercept, at least one "hidden" layer compromising activation functions, and an "output" layer that aggregates the outcome of the last hidden layer into a return prediction.

A feedforward neural network consists of several subsequent layers $l$, $l = 0, 1, ..., L$, one input layer ($l = 0$), $L - 1$ hidden layers ($l = 1, 2, ..., L - 1$) and one output layer $l = L$. Each layer $l$ contains $n^l$ nodes. In the case of the input layer, the number of nodes is equal to the number of characteristics, including an intercept, while the output layer contains due to the regression setting one node. In the case of the hidden layer, we consider an architecture of up to five hidden layers while the first hidden layer contains 32 nodes and each additional hidden layer divides the number of nodes by two compared to the previous layer following the geometric pyramid rule according to Masters (1993). This results in the following number of nodes per layer:

$$\begin{aligned} n^0 &= p + 1, \\ n^1 &= 32, \\ n^l &= \frac{n^{l-1}}{2} \forall l \in \{2, ..., L-1\}, \\ n^L &= 1. \end{aligned} \tag{22}$$

Each of the nodes in the hidden layer contains an activation function. In our case we follow Gu et al. (2020) and Leippold et al. (2022) and choose the rectified linear unit defined as:

$$\mathrm{ReLU}(x) = \max(0, x), \tag{23}$$

As in De Nard et al. (2022), we adopt the Adam optimization algorithm (Kingma and Ba, 2014), early stopping, batch normalization (Ioffe and Szegedy, 2015), ten ensembles with individual seeds (Hansen and Salamon, 1990; Dietterich, 2000) and dropout (Srivastava et al., 2014) when training our models.

## C.5 Hyperparameters

We will use the following hyperparameters based on the hyperparameter range in Gu et al. (2020), Tobek and Hronec (2020), Drobetz and Otto (2021), and Leippold et al. (2022):

- Elastic net
  - $\lambda$: $[1x10^{-5}, 2x10^{-5}, ..., 1x10^{-2}]$
  - $\alpha$: $[0, 0.01, ..., 1]$
- Random forest
  - $R$: $[0.01, 0.02, ..., 1]$
  - $B$: 1

  - $T$: $[100, 102, ..., 600]$
  - $D$: $[1, 2, ..., 8]$

- Gradient boosted regression tree
  - $LR$: $[0.01, 0.02, ..., 0.1]$
  - $T$: $[50, 52, ..., 500]$
  - $D$: $[1, 2, ..., 8]$

- Neural networks
  - $l_1$: $[0.00001, ..., 0.001]$
  - $LR$: $[0.001, 0.1]$
  - Batch Size: 10000
  - Epochs: 100

# Appendix D - Factor construction

We calculate the market factor as the value-weighted returns of all available stocks in excess of the risk-free rate. For the factors value, profitability, investment, and momentum, we estimate the portfolio breakpoints using the country-specific 30% and 70% percentile of the underlying characteristic using only the big-stock sample. In the case of the value stocks, we use the book-to-market ratio to categorize the stocks as Growth ($G$), Neutral ($N$), and Value ($V$). For profitability, we use the cash-based profitability as an underlying characteristic which enables us to sort the stocks into the extreme portfolios Weak ($W$) and Robust ($R$). In the case of the investment factor, we base the sorting on the stock's asset growth, which yields a Conservative ($C$) and Aggressive ($A$) portfolio. The last factor is based on the stock's momentum and sorts the stocks into the Winner ($W$) and Loser ($L$) portfolios. Finally, in the case of the size factor, we classify stocks into big ($B$) and small ($S$) as described in Section 2. The final factor calculation is based on the intersection of the different portfolios while the portfolio returns are value-weighted,

$$
\begin{aligned}
SMB &= (SV + SN + SG)/3 - (BV + BN + BG)/3, \\
HML &= (BV + SV)/2 - (BG + SG)/2, \\
RMW &= (BR + SR)/2 - (BW + SW)/2, \\
CMA &= (BC + SC)/2 - (BA + SA)/2, \\
MOM &= (BW + SW)/2 - (BL + SL)/2.
\end{aligned} \tag{24}
$$

# Appendix E - Figures

**Figure E.1:** Variable importance by model

Individual characteristics importance for the characteristics in each model. Characteristics importance is an average over all training samples. Variable importance within each model is normalized to sum to one.

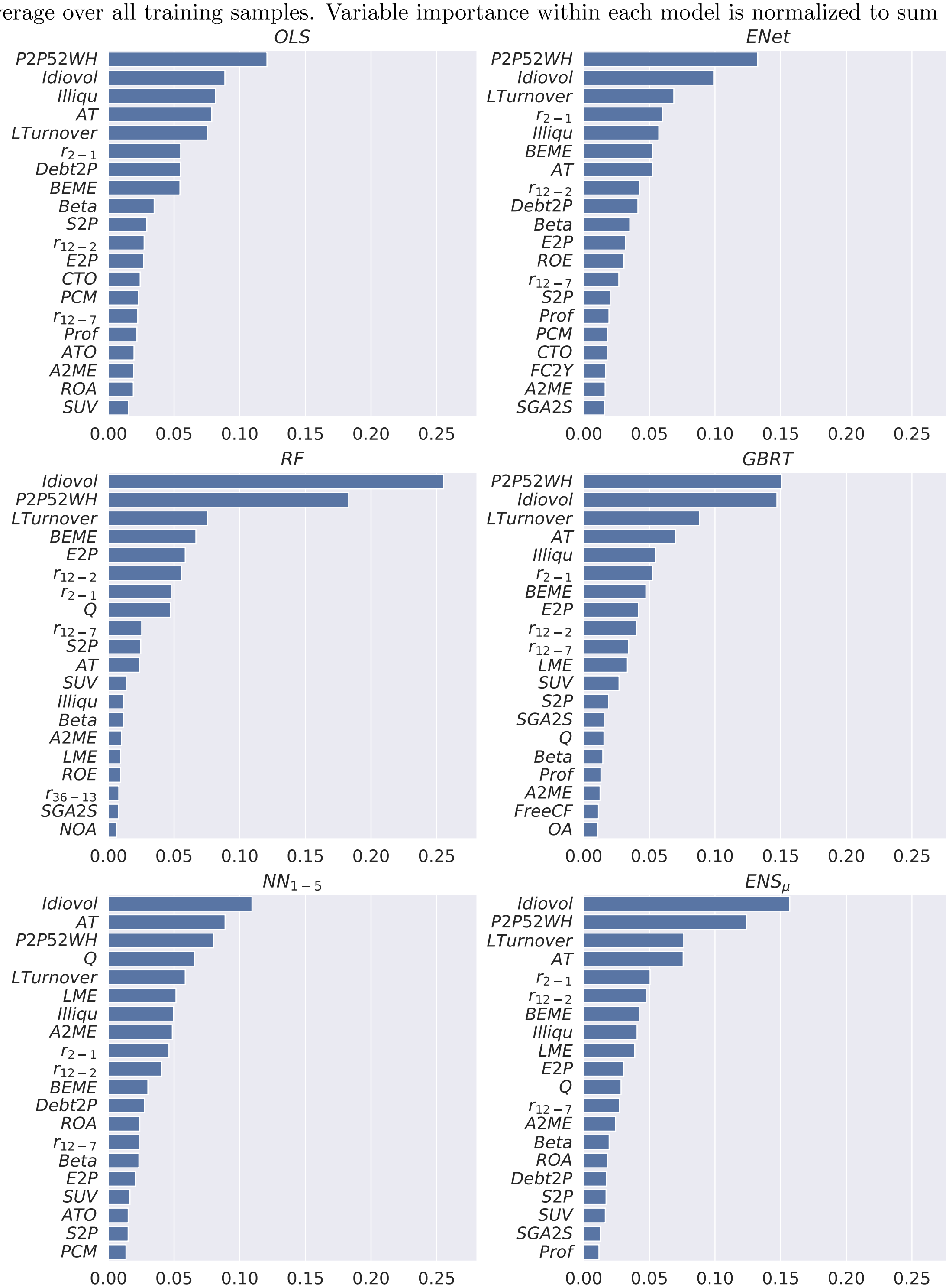
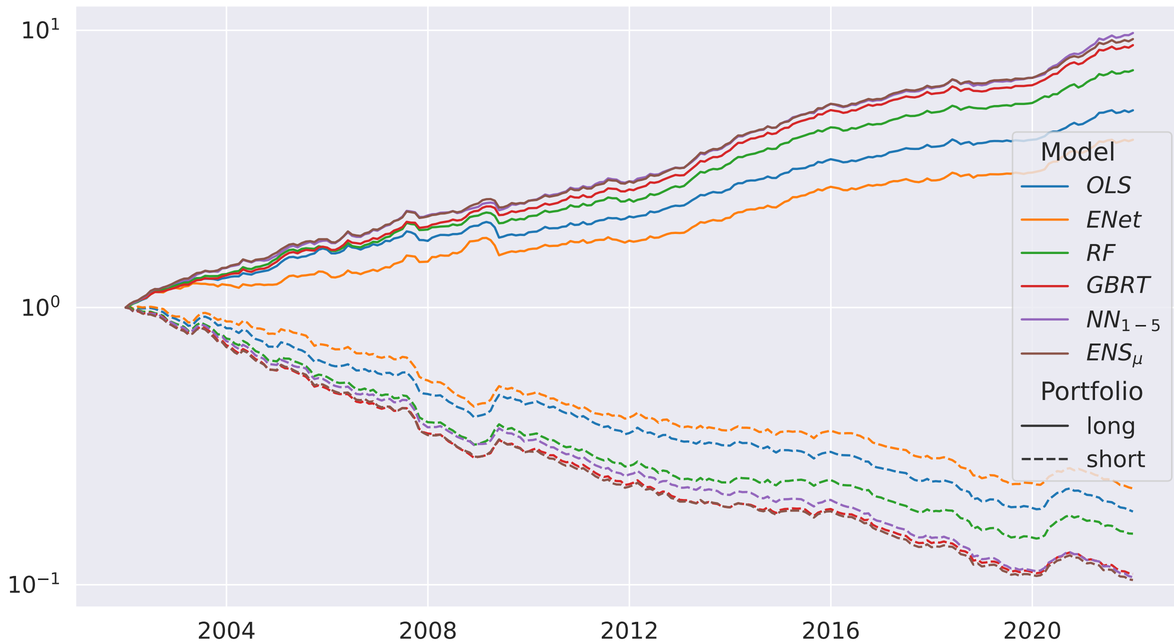


**Figure E.2:** Cumulative return of machine learning portfolios

The figure shows the cumulative log returns in excess of the market of portfolios sorted on out-of-sample machine learning return forecasts. The solid and dashed lines represent long (top quintile) and short (bottom quintile) positions, respectively. In Panel A equal-weighted cumulative log returns are shown while in Panel B the long and short positions are value-weighted. The sample period is from January 2002 to December 2021.

**Panel A: Equal-weighted**




**Panel B: Value-weighted**

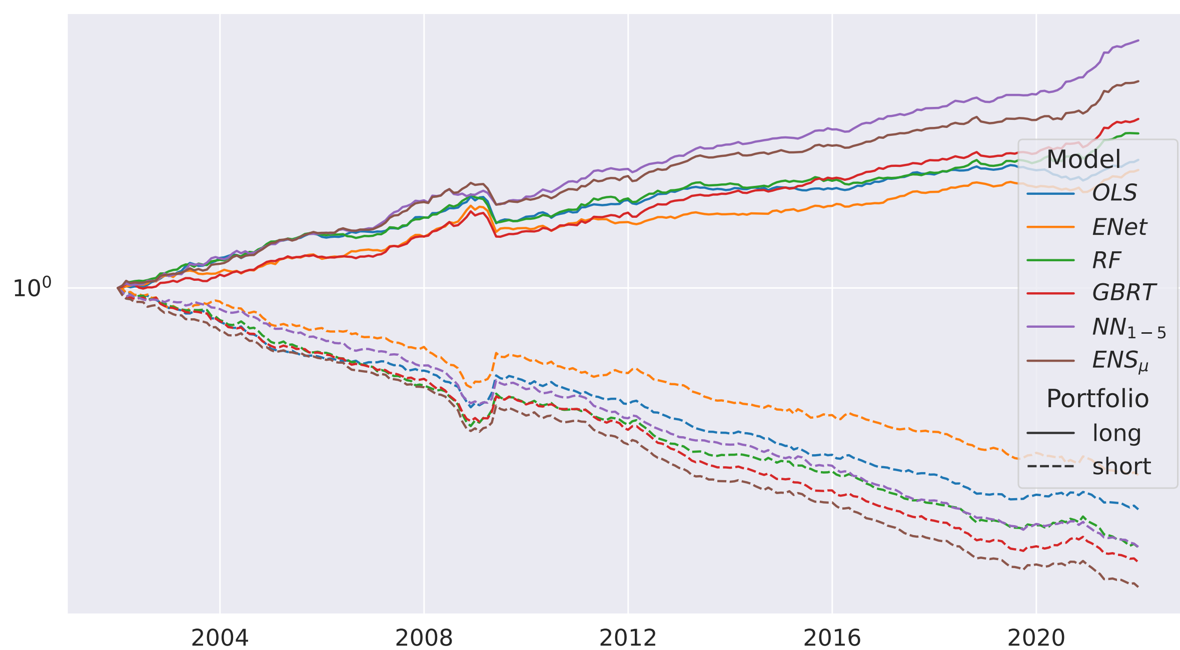
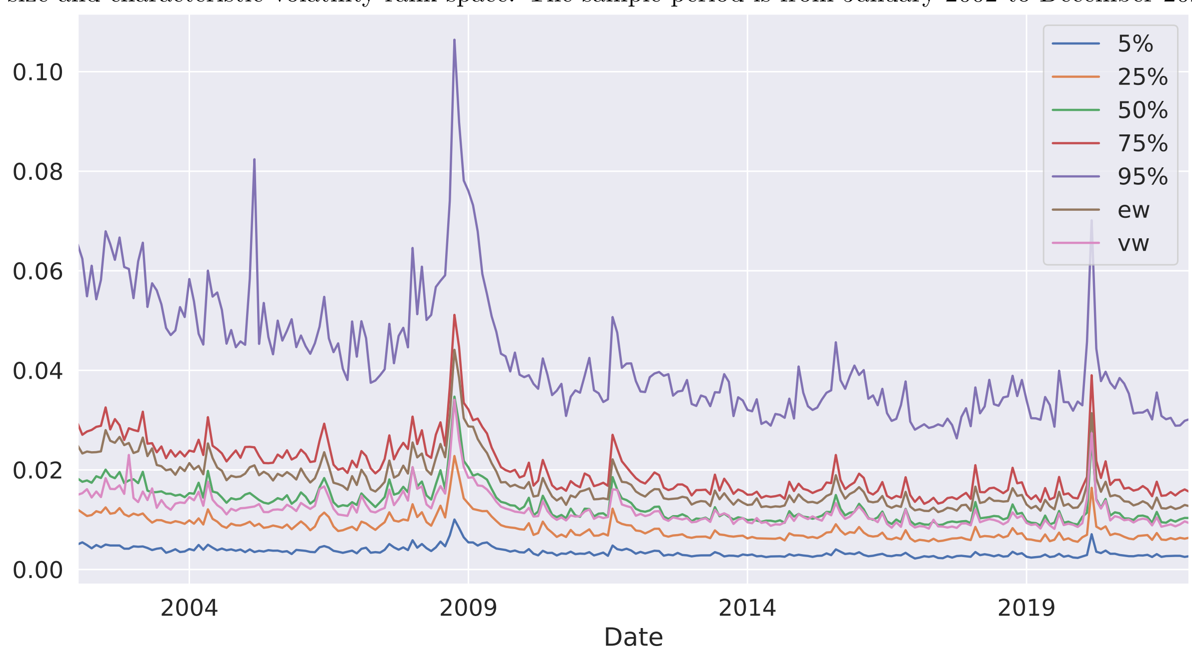


**Figure E.3:** Estimated bid-ask spreads based on the EDGE estimator

This figure shows the cross-sectional distribution of estimated bid-ask spreads for big stocks in emerging markets. Thereby, big stocks are defined as the biggest stocks, which together account for 90% of a country's aggregated market capitalization. For each stock and month, we compute the efficient discrete generalized estimator (EDGE) of the bid-ask spread, proposed in Ardia et al. (2022). The estimators are based on daily prices using a monthly estimation window. Following Novy-Marx and Velikov (2016), we replace zero estimates with the non-zero estimate of the stock of the same country with the shortest Euclidean distance in size and characteristic volatility rank space. The sample period is from January 2002 to December 2021.



# Appendix F - Table

**Table F.1:** Detail performance of the machine learning portfolios

This table reports the out-of-sample performance of the different machine learning quintile portfolios. Stocks are sorted into country-neutral quintiles based on their predicted returns for the next month. The sorting breakpoints are based on big stocks only, which are in the top 90% of a country's aggregated market capitalization. Each Panel provides the predicted monthly returns (Pred), the average monthly excess returns (Avg), corresponding $t$-statistics ($t$), the Fama and French (2018) six-factor model alpha ($\alpha$), and corresponding $t$-statistics. All $t$-statistics are calculated using Newey and West (1987) adjusted standard errors with 4 lags. The sample period is from January 2002 to December 2021.

| | Equal-weighted | | | | | Value-weighted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pred | Avg | $t$ | $\alpha$ | $t_\alpha$ | Pred | Avg | $t$ | $\alpha$ | $t_\alpha$ |
| **Panel A:** *OLS* | | | | | | | | | | |
| Low (L) | -1.07 | 0.26 | 0.50 | -0.39 | -3.43 | -0.99 | 0.43 | 0.83 | -0.23 | -3.40 |
| 2 | -0.38 | 0.90 | 1.85 | 0.11 | 1.16 | -0.37 | 0.82 | 1.70 | -0.00 | -0.05 |
| 3 | 0.00 | 1.15 | 2.51 | 0.26 | 3.08 | 0.01 | 0.95 | 2.09 | 0.09 | 2.55 |
| 4 | 0.36 | 1.34 | 3.03 | 0.41 | 4.91 | 0.37 | 1.13 | 2.55 | 0.17 | 3.45 |
| High (H) | 0.86 | 1.64 | 3.73 | 0.58 | 6.39 | 0.87 | 1.25 | 2.82 | 0.06 | 1.05 |
| H-L | 1.93 | 1.38 | 7.76 | 0.97 | 8.02 | 1.85 | 0.83 | 4.57 | 0.28 | 2.72 |
| **Panel B:** *ENet* | | | | | | | | | | |
| Low (L) | -1.07 | 0.34 | 0.65 | -0.33 | -3.05 | -0.99 | 0.51 | 0.98 | -0.18 | -2.33 |
| 2 | -0.37 | 0.92 | 1.86 | 0.12 | 1.32 | -0.36 | 0.81 | 1.70 | 0.02 | 0.32 |
| 3 | 0.02 | 1.16 | 2.48 | 0.27 | 3.19 | 0.03 | 0.96 | 2.06 | 0.09 | 1.95 |
| 4 | 0.39 | 1.32 | 2.97 | 0.38 | 4.52 | 0.40 | 1.09 | 2.43 | 0.11 | 2.21 |
| High (H) | 0.90 | 1.53 | 3.61 | 0.50 | 4.99 | 0.91 | 1.23 | 2.86 | 0.09 | 1.56 |
| H-L | 1.97 | 1.20 | 6.79 | 0.83 | 6.94 | 1.90 | 0.72 | 3.95 | 0.27 | 2.28 |
| **Panel C:** *RF* | | | | | | | | | | |
| Low (L) | -0.79 | 0.18 | 0.35 | -0.47 | -4.05 | -0.69 | 0.34 | 0.66 | -0.30 | -4.37 |
| 2 | -0.21 | 0.82 | 1.74 | 0.07 | 0.81 | -0.21 | 0.86 | 1.82 | 0.09 | 1.77 |
| 3 | 0.09 | 1.12 | 2.41 | 0.23 | 2.62 | 0.09 | 0.94 | 2.06 | 0.03 | 0.61 |
| 4 | 0.37 | 1.35 | 2.92 | 0.40 | 5.16 | 0.37 | 1.15 | 2.51 | 0.09 | 1.45 |
| High (H) | 0.71 | 1.78 | 3.96 | 0.72 | 8.34 | 0.70 | 1.32 | 2.99 | 0.17 | 3.38 |
| H-L | 1.50 | 1.60 | 9.29 | 1.19 | 14.10 | 1.39 | 0.99 | 5.24 | 0.47 | 5.24 |
| **Panel D:** *GBRT* | | | | | | | | | | |
| Low (L) | -0.87 | 0.04 | 0.08 | -0.59 | -5.24 | -0.74 | 0.30 | 0.58 | -0.38 | -5.57 |
| 2 | -0.18 | 0.87 | 1.82 | 0.11 | 1.21 | -0.17 | 0.81 | 1.71 | 0.09 | 1.68 |
| 3 | 0.13 | 1.13 | 2.42 | 0.25 | 3.25 | 0.13 | 0.94 | 2.04 | 0.03 | 0.73 |
| 4 | 0.43 | 1.36 | 2.98 | 0.39 | 4.76 | 0.42 | 1.12 | 2.47 | 0.11 | 1.62 |
| High (H) | 0.93 | 1.86 | 4.12 | 0.81 | 8.77 | 0.86 | 1.35 | 3.09 | 0.19 | 4.06 |
| H-L | 1.80 | 1.82 | 11.49 | 1.40 | 15.65 | 1.61 | 1.05 | 6.06 | 0.57 | 6.73 |

Continued on next page

Table F.1 continued

| | Equal-weighted | | | | | Value-weighted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pred | Avg | $t$ | $\alpha$ | $t_\alpha$ | Pred | Avg | $t$ | $\alpha$ | $t_\alpha$ |
| **Panel E:** $NN_1$ | | | | | | | | | | |
| Low (L) | -1.27 | 0.03 | 0.06 | -0.62 | -5.23 | -1.05 | 0.41 | 0.81 | -0.26 | -2.74 |
| 2 | -0.28 | 0.80 | 1.71 | 0.04 | 0.54 | -0.26 | 0.76 | 1.68 | -0.01 | -0.28 |
| 3 | 0.16 | 1.07 | 2.31 | 0.21 | 2.60 | 0.16 | 0.94 | 2.08 | 0.05 | 1.15 |
| 4 | 0.58 | 1.34 | 2.93 | 0.40 | 4.92 | 0.57 | 1.07 | 2.35 | 0.05 | 0.78 |
| High (H) | 1.34 | 1.91 | 4.09 | 0.86 | 8.93 | 1.25 | 1.45 | 3.09 | 0.30 | 5.86 |
| H-L | 2.61 | 1.88 | 13.92 | 1.47 | 15.67 | 2.30 | 1.04 | 6.94 | 0.57 | 4.83 |
| **Panel F:** $NN_2$ | | | | | | | | | | |
| Low (L) | -1.27 | -0.01 | -0.03 | -0.68 | -6.33 | -1.01 | 0.37 | 0.74 | -0.35 | -5.42 |
| 2 | -0.23 | 0.82 | 1.74 | 0.06 | 0.70 | -0.21 | 0.80 | 1.72 | 0.03 | 0.52 |
| 3 | 0.17 | 1.01 | 2.18 | 0.16 | 2.10 | 0.18 | 0.93 | 2.08 | 0.03 | 0.65 |
| 4 | 0.56 | 1.35 | 3.06 | 0.42 | 4.76 | 0.55 | 1.08 | 2.42 | 0.06 | 0.87 |
| High (H) | 1.32 | 1.90 | 3.99 | 0.86 | 8.87 | 1.21 | 1.49 | 3.08 | 0.36 | 6.84 |
| H-L | 2.60 | 1.91 | 15.67 | 1.55 | 19.02 | 2.21 | 1.11 | 9.40 | 0.71 | 9.16 |
| **Panel G:** $NN_3$ | | | | | | | | | | |
| Low (L) | -1.20 | 0.02 | 0.05 | -0.66 | -5.87 | -0.94 | 0.38 | 0.74 | -0.35 | -4.59 |
| 2 | -0.18 | 0.82 | 1.75 | 0.06 | 0.74 | -0.16 | 0.73 | 1.60 | -0.01 | -0.34 |
| 3 | 0.17 | 1.08 | 2.32 | 0.24 | 3.47 | 0.17 | 0.97 | 2.17 | 0.09 | 2.29 |
| 4 | 0.51 | 1.31 | 2.92 | 0.39 | 4.42 | 0.50 | 1.08 | 2.36 | 0.06 | 1.13 |
| High (H) | 1.22 | 1.86 | 3.99 | 0.83 | 8.34 | 1.11 | 1.49 | 3.17 | 0.32 | 5.34 |
| H-L | 2.41 | 1.84 | 14.72 | 1.49 | 16.93 | 2.04 | 1.12 | 7.85 | 0.66 | 6.55 |
| **Panel H:** $NN_4$ | | | | | | | | | | |
| Low (L) | -1.14 | 0.04 | 0.07 | -0.63 | -5.57 | -0.90 | 0.32 | 0.62 | -0.38 | -5.18 |
| 2 | -0.17 | 0.79 | 1.67 | 0.03 | 0.34 | -0.14 | 0.73 | 1.59 | -0.02 | -0.50 |
| 3 | 0.16 | 1.09 | 2.37 | 0.26 | 3.34 | 0.17 | 0.95 | 2.14 | 0.08 | 1.87 |
| 4 | 0.48 | 1.31 | 2.91 | 0.35 | 4.39 | 0.47 | 1.11 | 2.43 | 0.05 | 0.81 |
| High (H) | 1.15 | 1.89 | 4.07 | 0.86 | 8.26 | 1.04 | 1.52 | 3.25 | 0.35 | 6.35 |
| H-L | 2.29 | 1.86 | 13.75 | 1.48 | 16.72 | 1.94 | 1.20 | 8.30 | 0.73 | 7.76 |
| **Panel I:** $NN_5$ | | | | | | | | | | |
| Low (L) | -1.12 | 0.04 | 0.08 | -0.62 | -5.43 | -0.90 | 0.36 | 0.68 | -0.37 | -4.83 |
| 2 | -0.17 | 0.82 | 1.74 | 0.04 | 0.46 | -0.15 | 0.71 | 1.56 | 0.01 | 0.29 |
| 3 | 0.18 | 1.10 | 2.37 | 0.27 | 3.68 | 0.18 | 0.91 | 1.97 | 0.02 | 0.50 |
| 4 | 0.50 | 1.32 | 2.96 | 0.38 | 4.59 | 0.49 | 1.13 | 2.54 | 0.08 | 1.47 |
| High (H) | 1.14 | 1.88 | 4.04 | 0.82 | 8.38 | 1.04 | 1.52 | 3.26 | 0.34 | 7.61 |
| H-L | 2.26 | 1.84 | 13.42 | 1.44 | 15.79 | 1.93 | 1.17 | 8.11 | 0.71 | 8.21 |

Continued on next page

Table F.1 continued

| | Equal-weighted | | | | | Value-weighted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pred | Avg | $t$ | $\alpha$ | $t_\alpha$ | Pred | Avg | $t$ | $\alpha$ | $t_\alpha$ |
| **Panel J:** $NN_{1-5}$ | | | | | | | | | | |
| Low (L) | -1.13 | 0.03 | 0.06 | -0.62 | -5.41 | -0.91 | 0.34 | 0.65 | -0.36 | -4.82 |
| 2 | -0.19 | 0.77 | 1.65 | -0.00 | -0.00 | -0.16 | 0.73 | 1.62 | 0.02 | 0.42 |
| 3 | 0.17 | 1.10 | 2.38 | 0.27 | 3.57 | 0.17 | 0.95 | 2.10 | 0.06 | 1.35 |
| 4 | 0.51 | 1.31 | 2.92 | 0.39 | 4.50 | 0.50 | 1.07 | 2.39 | 0.03 | 0.54 |
| High (H) | 1.17 | 1.90 | 4.05 | 0.84 | 8.70 | 1.07 | 1.54 | 3.22 | 0.36 | 7.02 |
| H-L | 2.30 | 1.87 | 13.42 | 1.46 | 15.81 | 1.97 | 1.21 | 8.48 | 0.72 | 8.26 |
| **Panel K:** $ENS_\mu$ | | | | | | | | | | |
| Low (L) | -0.85 | 0.02 | 0.04 | -0.61 | -5.32 | -0.71 | 0.24 | 0.47 | -0.42 | -5.89 |
| 2 | -0.17 | 0.85 | 1.78 | 0.09 | 0.99 | -0.16 | 0.80 | 1.71 | 0.09 | 1.85 |
| 3 | 0.13 | 1.13 | 2.46 | 0.29 | 3.44 | 0.13 | 0.90 | 1.98 | 0.01 | 0.29 |
| 4 | 0.41 | 1.38 | 3.02 | 0.40 | 5.17 | 0.41 | 1.15 | 2.59 | 0.12 | 1.95 |
| High (H) | 0.87 | 1.88 | 4.12 | 0.82 | 9.02 | 0.81 | 1.45 | 3.17 | 0.25 | 5.82 |
| H-L | 1.71 | 1.86 | 11.73 | 1.43 | 15.66 | 1.52 | 1.20 | 6.97 | 0.67 | 8.29 |
| **Panel L:** $\mu_{sign(c)}$ | | | | | | | | | | |
| Low (L) | -0.20 | 0.22 | 0.42 | -0.41 | -3.16 | -0.19 | 0.48 | 0.93 | -0.22 | -3.42 |
| 2 | -0.08 | 0.79 | 1.57 | 0.00 | 0.02 | -0.08 | 0.90 | 1.84 | 0.10 | 2.10 |
| 3 | -0.01 | 1.05 | 2.20 | 0.18 | 2.19 | -0.01 | 0.96 | 2.10 | 0.08 | 1.75 |
| 4 | 0.06 | 1.25 | 2.77 | 0.32 | 3.72 | 0.06 | 1.08 | 2.45 | 0.05 | 1.06 |
| High (H) | 0.15 | 1.56 | 3.67 | 0.57 | 7.27 | 0.15 | 1.24 | 2.91 | 0.11 | 2.18 |
| H-L | 0.35 | 1.34 | 6.48 | 0.98 | 7.26 | 0.34 | 0.77 | 4.16 | 0.32 | 3.32 |

**Table F.2:** Regional performance - Subregion

This table reports the individual regional equal-weighted and value-weighted performance of the long-short prediction-sorted portfolios over the 20-year out-of-sample testing period. All stocks are sorted into country-neutral portfolios based on their predicted returns for the next month. The sorting breakpoints are based on big stocks only, which are in the top 90% of the country's aggregated market capitalization. Panel A summarizes the baseline results, and Panel B shows the result for all countries being part of emerging Americas, Panel C combines all emerging Asian countries, and Panel D reports results for emerging countries from Europe, the Middle East, and Africa. The first two rows of each panel provide the average monthly return of the long-short quintile (Avg), corresponding $t$-statistics ($t$), the average Fama and French (2018) six-factor alpha ($\alpha$), corresponding $t$-statistics ($t_\alpha$), and $R^2$. The next two rows show spanning alpha ($\alpha$), corresponding $t$-statistic ($t_\alpha$), and $R^2$ when regressing the long-short $ENS$ returns on $OLS$ returns and vice versa. All $t$-statistics are calculated using Newey and West (1987) adjusted standard errors with 4 lags. The sample period is from January 2002 to December 2021.

| | Equal-weighted | | | | | Value-weighted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Avg | $t$ | $\alpha$ | $t_\alpha$ | $R^2$ | Avg | $t$ | $\alpha$ | $t_\alpha$ | $R^2$ |
| **Panel A: Emerging Markets** | | | | | | | | | | |
| $OLS$ | 1.09 | 6.29 | 0.77 | 6.23 | 53.92 | 0.78 | 4.42 | 0.29 | 2.57 | 56.92 |
| $ENS_\mu$ | 1.35 | 8.88 | 0.97 | 9.75 | 57.22 | 0.97 | 5.95 | 0.44 | 4.59 | 58.10 |
| $ENS_\mu \sim OLS$ | | | 0.49 | 7.78 | 77.56 | | | 0.28 | 4.46 | 75.77 |
| $OLS \sim ENS_\mu$ | | | -0.23 | -1.76 | 77.56 | | | -0.06 | -0.51 | 75.77 |
| **Panel B: Americas** | | | | | | | | | | |
| $OLS$ | 0.76 | 3.23 | 0.36 | 1.82 | 46.02 | 0.60 | 2.67 | 0.04 | 0.22 | 49.64 |
| $ENS_\mu$ | 0.70 | 3.37 | 0.41 | 2.52 | 36.93 | 0.64 | 3.12 | 0.20 | 1.14 | 35.55 |
| $ENS_\mu \sim OLS$ | | | 0.19 | 1.56 | 56.53 | | | 0.22 | 1.66 | 49.56 |
| $OLS \sim ENS_\mu$ | | | 0.17 | 1.02 | 56.53 | | | 0.14 | 0.83 | 49.56 |
| **Panel C: Asia** | | | | | | | | | | |
| $OLS$ | 1.43 | 7.73 | 1.12 | 9.62 | 60.08 | 0.81 | 4.03 | 0.41 | 3.22 | 63.14 |
| $ENS_\mu$ | 1.95 | 11.16 | 1.61 | 17.97 | 60.73 | 1.27 | 6.62 | 0.83 | 8.24 | 67.10 |
| $ENS_\mu \sim OLS$ | | | 0.65 | 8.64 | 83.54 | | | 0.52 | 5.07 | 71.55 |
| $OLS \sim ENS_\mu$ | | | -0.36 | -2.10 | 83.54 | | | -0.17 | -1.11 | 71.55 |
| **Panel D: Europe, the Middle East and Africa** | | | | | | | | | | |
| $OLS$ | 1.06 | 5.97 | 0.91 | 5.67 | 19.35 | 0.92 | 4.43 | 0.47 | 2.39 | 26.25 |
| $ENS_\mu$ | 1.39 | 8.11 | 1.06 | 6.59 | 20.58 | 0.99 | 4.75 | 0.34 | 1.97 | 37.32 |
| $ENS_\mu \sim OLS$ | | | 0.61 | 5.47 | 56.62 | | | 0.25 | 2.01 | 59.45 |
| $OLS \sim ENS_\mu$ | | | -0.01 | -0.06 | 56.62 | | | 0.19 | 1.45 | 59.45 |

**Table F.3:** Further investment frictions

This table reports the out-of-sample performance for returns on different buy/hold long-only strategies before when accounting for transaction costs and limiting our investment universe to big stocks only. We investigate predictions from a linear OLS model and an ensemble ($ENS$) of non-linear machine learning models ($RF$, $GBRT$, and $NN_{1-5}$). Every month the portfolio consists of the stocks that currently belong to the top X% plus the stocks selected in previous months that have not deteriorated beyond the top (bottom) Y%. We report the strategies' gross returns in excess of the market, average two-way turnover, transaction costs, net returns in excess of the market, and net Fama and French (2018) six-factor models alphas. We assume one-way transaction costs of 100 basis points. All $t$-statistics are Newey and West (1987) adjusted with 4 lags. Panel A summarizes results from equal-weighting while Panel B shows results from value-weighting. The sample period is from January 2002 to December 2021.

| | $OLS$ | | $ENS_{\mu}$ | |
|---|---|---|---|---|
| | 20%/20% | 10%/30% | 20%/20% | 10%/30% |
| **Panel A: Equal-weighted** | | | | |
| $r^e_{gross} - Mkt$ | 0.47 | 0.44 | 0.78 | 0.78 |
| | (5.31) | (5.07) | (7.88) | (7.37) |
| TO (in %) | 44.14 | 24.68 | 45.13 | 27.38 |
| T-cost (in %) | 0.44 | 0.25 | 0.45 | 0.27 |
| $r^e_{net} - Mkt$ | 0.03 | 0.19 | 0.33 | 0.51 |
| | (0.37) | (2.26) | (3.35) | (4.82) |
| $\alpha^{FF6}_{net}$ | 0.17 | 0.33 | 0.43 | 0.62 |
| | (2.77) | (5.17) | (5.84) | (7.35) |
| **Panel B: Value-weighted** | | | | |
| $r^e_{gross} - Mkt$ | 0.27 | 0.29 | 0.45 | 0.46 |
| | (2.99) | (3.13) | (4.67) | (4.53) |
| TO (in %) | 44.09 | 21.86 | 45.46 | 23.28 |
| T-cost (in %) | 0.44 | 0.22 | 0.45 | 0.23 |
| $r^e_{net} - Mkt$ | -0.17 | 0.07 | -0.01 | 0.23 |
| | (-1.88) | (0.74) | (-0.08) | (2.26) |
| $\alpha^{FF6}_{net}$ | -0.16 | 0.06 | 0.02 | 0.25 |
| | (-3.27) | (1.14) | (0.36) | (3.72) |

## References


Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Anand, V., Brunner, R., Ikegwu, K., Sougiannis, T., 2019. Predicting profitability using machine learning. SSRN Working Paper no. 3466478 .

Ang, A., Hodrick, R. J., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. The Journal of Finance 61, 259–299.

Annaert, J., Ceuster, M. D., Verstegen, K., 2013. Are extreme returns priced in the stock market? European evidence. Journal of Banking & Finance 37, 3401–3411.

Ardia, D., Guidotti, E., Kroencke, T. A., 2022. Efficient estimation of bid-ask spreads from open, high, low, and close prices. SSRN Working Paper no. 3892335 .

Asness, C. S., 2011. Momentum in Japan: The exception that proves the rule. Journal of Portfolio Management 37, 67–75.

Asness, C. S., Frazzini, A., 2013. The devil in HML's details. Journal of Portfolio Management 39, 49–68.

Asness, C. S., Frazzini, A., Pedersen, L. H., 2019. Quality minus junk. Review of Accounting Studies 24, 34–112.

Avramov, D., Cheng, S., Metzker, L., 2022. Machine learning vs. economic restrictions: Evidence from stock return predictability. Management Science , forthcoming.

Azevedo, V., Kaiser, G. S., Müller, S., 2022. Stock market anomalies and machine learning across the globe. SSRN Working Paper no. 4071852 .

Balakrishnan, K., Bartov, E., Faurel, L., 2010. Post loss/profit announcement drift. Journal of Accounting and Economics 50, 20–41.

Bali, T. G., Beckmeyer, H., Moerke, M., Weigert, F., 2022a. Option return predictability with machine learning and big data. SSRN Working Paper no. 3895984 .

Bali, T. G., Goyal, A., Huang, D., Jiang, F., Wen, Q., 2022b. Predicting corporate bond returns: Merton meets machine learning. SSRN Working Paper no. 3686164 .

Ball, R., Gerakos, J., Linnainmaa, J. T., Nikolaev, V. V., 2015. Deflating profitability. Journal of Financial Economics 117, 225–248.

Ball, R., Gerakos, J., Linnainmaa, J. T., Nikolaev, V. V., 2016. Accruals, cash flows, and operating profitability in the cross section of stock returns. Journal of Financial Economics 121, 28–45.

Basu, S., 1983. The relationship between earnings yield, market value and return for nyse common stocks : Further evidence. Journal of Financial Economics 12, 129–156.

Bekaert, G., Engstrom, E. C., Xu, N. R., 2022. The time variation in risk appetite and uncertainty. Management Science 68, 3975–4004.

Bekaert, G., Harvey, C. R., 1995. Time-varying world market integration. The Journal of Finance 50, 403–444.

Bhandari, L. C., 1988. Debt/equity ratio and expected common stock returns: Empirical evidence. The Journal of Finance 43, 507–528.

Bianchi, D., Büchner, M., Hoogteijling, T., Tamoni, A., 2021a. Corrigendum: Bond risk premiums with machine learning. The Review of Financial Studies 34, 1090–1103.

Bianchi, D., Büchner, M., Tamoni, A., 2021b. Bond risk premiums with machine learning. The Review of Financial Studies 34, 1046–1089.

Blitz, D., Hanauer, M. X., Honarvar, I., Huisman, R., van Vliet, P., 2022. Beyond Fama-French factors: Alpha from short-term signals. SSRN Working Paper no. 4115411 .

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Cakici, N., Fabozzi, F. J., Tan, S., 2013. Size, value, and momentum in emerging market stock returns. Emerging Markets Review 16, 46–65.

Cakici, N., Fieberg, C., Metko, D., Zaremba, A., 2022a. Machine learning goes global: Cross-sectional return predictability in international stock markets. SSRN Working Paper no. 4141663 .

Cakici, N., Shahzad, S. J. H., Bedowska-Sojka, B., Zaremba, A., 2022b. Machine learning and the cross-section of cryptocurrency returns. SSRN Working Paper no. 4295427 .

Cakici, N., Zaremba, A., 2022. Empirical asset pricing via machine learning: The global edition. SSRN Working Paper no. 4028525 .

Campbell, C. J., Cowan, A. R., Salotti, V., 2010. Multi-country event-study methods. Journal of Banking & Finance 34, 3078–3090.

Carhart, M. M., 1997. On persistence in mutual fund performance. The Journal of Finance 52, 57–82.

Chan, L. K. C., Karceski, J., Lakonishok, J., 1998. The risk and return from factors. Journal of Financial and Quantitative Analysis 33, 159–188.

Chen, L., Pelger, M., Zhu, J., 2021. Deep learning in asset pricing. SSRN Working Paper no. 3350138 .

Chen, X., Cho, Y. H. T., Dou, Y., Lev, B., 2022. Predicting future earnings changes using machine learning and detailed financial data. Journal of Accounting Research 60, 467–515.

Cooper, M. J., Gulen, H., Schill, M. J., 2008. Asset growth and the cross-section of stock returns. The Journal of Finance 63, 1609–1651.

Daniel, K., Moskowitz, T. J., 2016. Momentum crashes. Journal of Financial Economics 122, 221–247.

Daniel, K., Titman, S., 2006. Market reactions to tangible and intangible information. The Journal of Finance 61, 1605–1643.

Datar, V. T., Y. Naik, N., Radcliffe, R., 1998. Liquidity and stock returns: An alternative test. Journal of Financial Markets 1, 203–219.

Davis, J. L., Fama, E. F., French, K. R., 2000. Characteristics, covariances, and average returns: 1929 to 1997. The Journal of Finance 55, 389–406.

De Bondt, W. F. M., Thaler, R., 1985. Does the stock market overreact? The Journal of Finance 40, 793–805.

De Nard, G., Hediger, S., Leippold, M., 2022. Subsampled factor models for asset pricing: The rise of vasa. Journal of Forecasting 41, 1217–1247.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13, 134–144.

Dietterich, T. G., 2000. Ensemble methods in machine learning. In: *Multiple Classifier Systems*, Springer Berlin Heidelberg, Berlin, Heidelberg, p. 1.

Drobetz, W., Hollstein, F., Otto, T., Prokopczuk, M., 2021. Estimating security betas via machine learning. SSRN Working Paper no. 3933048 .

Drobetz, W., Otto, T., 2021. Empirical asset pricing via machine learning: evidence from the European stock market. Journal of Asset Management 22, 507–538.

D'Acunto, F., Liu, R., Pflueger, C., Weber, M., 2018. Flexible prices and leverage. Journal of Financial Economics 129, 46–68.

Erel, I., Stern, L. H., Tan, C., Weisbach, M. S., 2021. Selecting directors using machine learning. The Review of Financial Studies 34, 3226–3264.

Estrada, G. B., Park, D., Ramayandi, A., 2016. Taper tantrum and emerging equity market slumps. Emerging Markets Finance and Trade 52, 1060–1071.

Fama, E. F., French, K. R., 1992. The cross-section of expected stock returns. The Journal of Finance 47, 427–465.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E. F., French, K. R., 1996. Multifactor explanations of asset pricing anomalies. The Journal of Finance 51, 55–84.

Fama, E. F., French, K. R., 2008. Dissecting anomalies. The Journal of Finance 63, 1653–1678.

Fama, E. F., French, K. R., 2012. Size, value, and momentum in international stock returns. Journal of Financial Economics 105, 457–472.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Fama, E. F., French, K. R., 2017. International tests of a five-factor asset pricing model. Journal of Financial Economics 123, 441–463.

Fama, E. F., French, K. R., 2018. Choosing factors. Journal of Financial Economics 128, 234–252.

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

Fong, K. Y. L., Holden, C. W., Trzcinka, C. A., 2017. What are the best liquidity proxies for global research? Review of Finance 21, 1355–1401.

Frazzini, A., Pedersen, L. H., 2014. Betting against beta. Journal of Financial Economics 111, 1 – 25.

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. The Review of Financial Studies 33, 2326–2377.

Gandhi, P., Lustig, H., 2015. Size anomalies in U.S. bank stock returns. The Journal of Finance 70, 733–768.

Garfinkel, J. A., 2009. Measuring investors' opinion divergence. Journal of Accounting Research 47, 1317–1348.

George, T. J., Hwang, C.-Y., 2004. The 52-week high and momentum investing. The Journal of Finance 59, 2145–2176.

Gorodnichenko, Y., Weber, M., 2016. Are sticky prices costly? Evidence from the stock market. American Economic Review 106, 165–99.

Griffin, J. M., Hirschey, N. H., Kelly, P. J., 2011. How important is the financial media in global markets? The Review of Financial Studies 24, 3941–3992.

Griffin, J. M., Kelly, P. J., Nardari, F., 2010. Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. The Review of Financial Studies 23, 3225–3277.

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

Hanauer, M., 2020. A comparison of global factor models. Working Paper.

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

Hanauer, M. X., Lauterbach, J. G., 2019. The cross-section of emerging market stock returns. Emerging Markets Review 38, 265–286.

Hanauer, M. X., Linhart, M., 2015. Size, value, and momentum in emerging market stock returns: Integrated or segmented pricing? Asia-Pacific Journal of Financial Studies 44, 175–214.

Hansen, L., Salamon, P., 1990. Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence 12, 993–1001.

Harvey, C. R., 1995. Predictable risk and returns in emerging markets. The Review of Financial Studies 8, 773–816.

Haugen, R. A., Baker, N. L., 1996. Commonality in the determinants of expected stock returns. Journal of Financial Economics 41, 401–439.

Hirshleifer, D., Kewei Hou, Teoh, S. H., Yinglei Zhang, 2004. Do investors overvalue firms with bloated balance sheets? Journal of Accounting and Economics 38, 297–331.

Hou, K., Karolyi, G. A., Kho, B.-C., 2011. What factors drive global stock returns? The Review of Financial Studies 24, 2527–2574.

Hou, K., Xue, C., Zhang, L., 2020. Replicating anomalies. The Review of Financial Studies 33, 2019–2133.

Ince, O. S., Porter, R. B., 2006. Individual equity return data from Thomson Datastream: Handle with care! Journal of Financial Research 29, 463–479.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift.

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

Jansen, M., Swinkels, L., Zhou, W., 2021. Anomalies in the china a-share market. Pacific-Basin Finance Journal 68, 101607.

Jaquart, P., Dann, D., Weinhardt, C., 2021. Short-term bitcoin market prediction via machine learning. The Journal of Finance and Data Science 7, 45–66.

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. The Journal of Finance 45, 881–898.

Kaniel, R., Lin, Z., Pelger, M., Van Nieuwerburgh, S., 2022. Machine-learning the skill of mutual fund managers. NBER Working Paper No. 29723 .

Karolyi, G. A., Lee, K.-H., van Dijk, M. A., 2012. Understanding commonality in liquidity around the world. Journal of Financial Economics 105, 82–112.

Kaufmann, H., Messow, P., Vogt, J., 2021. Boosting the equity momentum factor in credit. Financial Analysts Journal 77, 83–103.

Kelly, B. T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134, 501–524.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization.

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

Lakonishok, J., Shleifer, A., Vishny, R. W., 1994. Contrarian investment, extrapolation, and risk. The Journal of Finance 49, 1541–1578.

Leippold, M., Wang, Q., Zhou, W., 2022. Machine learning in the chinese stock market. Journal of Financial Economics 145, 64–82.

Leung, E., Lohre, H., Mischlich, D., Shea, Y., Stroh, M., 2021. The promises and pitfalls of machine learning for predicting stock returns. The Journal of Financial Data Science 3, 21–50.

Lewellen, J., 2015. The cross-section of expected stock returns. Critical Finance Review 4, 1–44.

Lewellen, J., Nagel, S., 2006. The conditional capm does not explain asset-pricing anomalies. Journal of Financial Economics 82, 289 – 314.

Litzenberger, R. H., Ramaswamy, K., 1979. The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. Journal of Financial Economics 7, 163–195.

Lyandres, E., Sun, L., Zhang, L., 2007. The New Issues Puzzle: Testing the Investment-Based Explanation. The Review of Financial Studies 21, 2825–2855.

Masters, T., 1993. Practical Neural Network Recipes in C++. Academic Press Professional, Inc., USA.

Medhat, M., Schmeling, M., 2022. Short-term momentum. The Review of Financial Studies 35, 1480–1526.

Moritz, B., Zimmermann, T., 2016. Tree-based conditional portfolio sorts: The relation between past and future stock returns. SSRN Working Paper no. 2740751 .

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Novy-Marx, R., 2010. Operating leverage. Review of Finance 15, 103–134.

Novy-Marx, R., 2012. Is momentum really momentum? Journal of Financial Economics 103, 429–453.

Novy-Marx, R., 2013. The other side of value: The gross profitability premium. Journal of Financial Economics 108, 1–28.

Novy-Marx, R., Velikov, M., 2016. A taxonomy of anomalies and their trading costs. The Review of Financial Studies 29, 104–147.

Palazzo, B., 2012. Cash holdings, risk, and expected returns. Journal of Financial Economics 104, 162–185.

Pontiff, J., 2006. Costly arbitrage and the myth of idiosyncratic risk. Journal of Accounting and Economics 42, 35–52, conference Issue on Implications of Changing Financial Reporting Standards.

Rapach, D. E., Strauss, J. K., Tu, J., Zhou, G., 2019. Industry return predictability: A machine learning approach. The Journal of Financial Data Science 1, 9–28.

Rasekhschaffe, K. C., Jones, R. C., 2019. Machine learning for stock selection. Financial Analysts Journal 75, 70–88.

Roon, F. A. D., Nijman, T. E., Werker, B. J. M., 2001. Testing for mean-variance spanning with short sales constraints and transaction costs: The case of emerging markets. The Journal of Finance 56, 721–742.

Rosenberg, B., Reid, K., Lanstein, R., 1985. Persuasive evidence of market inefficiency. Journal of Portfolio Management 11, 9–16.

Rossi, A. G., 2018. Predicting stock market returns with machine learning. Working paper .

Rouwenhorst, K. G., 1999. Local return factors and turnover in emerging stock markets. The Journal of Finance 54, 1439–1464.

Rubesam, A., 2022. Machine learning portfolios with equal risk contributions: Evidence from the Brazilian market. Emerging Markets Review 51, 100891.

Sadhwani, A., Giesecke, K., Sirignano, J., 2021. Deep learning for mortgage risk. Journal of Financial Econometrics 19, 313–368.

Schmidt, P. S., Von Arx, U., Schrimpf, A., Wagner, A. F., Ziegler, A., 2017. On the construction of common size, value and momentum factors in international stock markets: A guide with applications. Swiss Finance Institute Research Paper 10.

Schmidt, P. S., von Arx, U., Schrimpf, A., Wagner, A. F., Ziegler, A., 2019. Common risk factors in international stock markets. Financial Markets and Portfolio Management 33, 213–241.

Sloan, R. G., 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? The Accounting Review 71, 289–315.

Smajlbegovic, E., 2019. Regional economic activity and stock returns. Journal of Financial and Quantitative Analysis 54, 1051—1082.

Soliman, M. T., 2008. The use of DuPont analysis by market participants. The Accounting Review 83, 823–853.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15, 1929–1958.

Stambaugh, R. F., Yu, J., Yuan, Y., 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. The Journal of Finance 70, 1903–1948.

Struck, C., Cheng, E., 2020. The cross section of commodity returns: A nonparametric approach. The Journal of Financial Data Science 2, 86–103.

Tobek, O., Hronec, M., 2020. Does it pay to follow anomalies research? Machine learning approach with international evidence. Journal of Financial Markets p. 100588.

Van Binsbergen, J. H., Han, X., Lopez-Lira, A., 2020. Man vs. machine learning: The term structure of earnings expectations and conditional biases. NBER Working Paper No. 27843 .

van der Hart, J., de Zwart, G., van Dijk, D., 2005. The success of stock selection strategies in emerging markets: Is it risk or behavioral bias? Emerging Markets Review 6, 238–262.

van der Hart, J., Slagter, E., van Dijk, D., 2003. Stock selection strategies in emerging markets. Journal of Empirical Finance 10, 105–132.

Welch, I., 2020. Simpler better market betas. Working Paper.

Windmüller, S., 2022. Firm characteristics and global stock returns: A conditional asset pricing model. The Review of Asset Pricing Studies 12, 447–499.

Wooldridge, J. M., 2001. Applications of generalized method of moments estimation. Journal of Economic Perspectives 15, 87–100.

Wu, W., Chen, J., Yang, Z. B., Tindall, M. L., 2021. A cross-sectional machine learning approach for hedge fund return prediction and selection. Management Science 67, 4577–4601.

Zaremba, A., Czapkiewicz, A., 2017. Digesting anomalies in emerging European markets: A comparison of factor pricing models. Emerging Markets Review 31, 1–15.

Zhang, X., 2006. Specification tests of international asset pricing models. Journal of International Money and Finance 25, 275–307.