Empirical Asset Pricing via Machine Learning: The Global Edition

Nusret Cakici[†], *Adam Zaremba*[‡⁑]

Abstract

We examine the cross-section of international equity risk premia with machine learning methods. We identify, classify, and calculate 88 market characteristics and use them to forecast country returns with various machine learning techniques. While all algorithms produce substantial economic gains, a two-layer neural network proves particularly effective. The associated long-short portfolio generates 1.69% per month at a Sharpe ratio of 1.57. Most models select a consistent group of leading predictors: long-run reversal, earnings yield, size, market breadth, and momentum. The return predictability is driven by mispricing rather than risk. In consequence, it is boosted by high limits to arbitrage but gradually diminishes over time as global markets mature.

Keywords: machine learning, factor investing, the cross-section of country stock returns, equity risk premia, international markets, return predictability, forecast combination

JEL classification: C52, C55, C58, G11, G12, G14, G17

This version: March 25, 2021

* We thank Azizjon Alimov, Turan G. Bali, Guillaume Coqueret, Jean Dessain, Hayette Gatfaoui, Campbell R. Harvey, Oskar Kowalewski, Joelle Miffre, Alexandre Rubesam, Clemens Struck, Ondrey Tobek, Pim van Vliet, and Goufu Zhou for helpful comments and suggestions, as well as seminar participants at IÉSEG School of Management. We are solely responsible for any remaining errors. Adam Zaremba acknowledges the support of the National Science Center of Poland [grant no. 2019/33/B/HS4/01021].

⁑ Corresponding author.

† Nusret Cakici, Gabelli School of Business, Fordham University, 45 Columbus Avenue, Room 510, New York, NY 10023, USA, tel. +1-212-636-6776, email: cakici@fordham.edu.

‡ Adam Zaremba, 1) Montpellier Business School, 2300 Avenue des Moulins, 34185 Montpellier, France, email: a.zaremba@montpellier-bs.com, tel. +33 (0)4 67 10 27 67; 2) Department of Investment and Financial Markets, Institute of Finance, Poznan University of Economics and Business, al. Niepodległości 10, 61-875 Poznań, Poland, email: adam.zaremba@ue.poznan.pl.

1. Introduction

Asset pricing literature documents a growing list of predictors of the cross-section of country equity risk premia. It contains not only counterparts of traditional stock level anomalies—such as value, momentum, reversal, or beta—but also market-specific features: political risk, sovereign risk, or the interest rates.[1] Though this abundance promises improved return predictability, it also gives rise to whole new questions. Which of these variables really matter? Do they interact with each other? How can they be integrated? While handling these problems with a traditional econometric toolset may be challenging, recent developments in machine learning offer promising solutions. Their capacity for feature selection and capturing interactions and nonlinearities appears well-suited to deal with the proliferation of country-level return predictors.

In this paper, we examine the cross-section of country equity returns with machine learning methods. Using nearly four decades of data from 71 markets, we identify, classify, and calculate a comprehensive set of 88 country-level return predictors. Next—building on Gu, Kelly, and Xiu's (2020) framework—we apply a repertoire of various machine learning algorithms. The aim of our study is twofold. First, we seek to scrutinize the performance of machine learning techniques in predicting the cross-section of country-equity returns. Second—taking advantage of their unique properties—we want to gain novel insights into the dynamics of country risk premia around the world.

Our explorations contribute in five major ways. First, we demonstrate substantial economic gains from using machine learning methods for forecasting country risk premia. In line with earlier stock-level evidence, accounting for interactions and nonlinearities brings significant benefits. The top-performing methods for data include neural networks and support vector machines. On the other hand, dimension reduction techniques visibly lag behind. Finally, the champion of this model horserace is the forecast combination. While individual algorithms fare either better or worse, combining them effectively reduces model variance and produces superior results.

When evaluating the relative performance of different prediction techniques, we do not limit our tests to the classical out-of-sample R^2 coefficient. While this measure is prevalent, its indications may be drowned in the noise of both return and forecast variance (Kelly et al., 2021; Coqueret, 2022). Hence, we supplement our results with a novel alternative measure: the cross-sectional R^2. This measure focuses on *cross-sectional*

[1] See, e.g., for value: Asness et al. (2013), Baltussen et al. (2021); for momentum: Chan et al. (2000), Bhojraj and Swaminathan (2006), Asness et al. (2013); for beta: Frazzini and Pedersen (2014); for reversal: Balvers et al. (2000); for political, sovereign, and economic risks: Erb et al. (1995, 1996), Diamonte et al. (1996), Bekaert et al. (1997), Avramov et al. (2012); for the influence of interest rates and bond markets: Hjalmarsson (2010), Pitkäjärvi et al. (2020).

fit rather than on *global fit*. Therefore, it guides how effectively a given method may sort assets into portfolios. Interestingly, we find that even the simplest methods (which produce seemingly low out-of-sample R^2) can still exhibit a close cross-sectional fit. In consequence, even if the standard R^2 is negative, the predictions could still be helpful in practice.

Second, machine learning models allow the extraction of the crucial country return predictors from the existing "anomaly zoo." The variable importance analysis reveals that a relatively sparse selection of covariates dominates the cross-section of country returns. Most models agree on several market characteristics that matter; these include long-term reversal, earnings yield, market value, market breadth, and momentum. These few variables capture most of the global variation in the cross-section of country equity risk premia. Notably, more sophisticated measures of momentum or value effects—as well as plenty of other political, credit, liquidity, or economic risks—are of secondary importance.

Third, further analyses uncover the practical implications of our findings. The predictions taken from the machine learning models can be effectively coined into successful investment strategies. Remarkably, all the forecasting methods translate into evident patterns in the cross-section of stock returns. Consequently, univariate portfolio sorts that are based on machine learning predictions produce substantial profits. Contrary to the standard narrative, even the mere ordinary least squares method exhibits sizeable alphas. Nonetheless, the center stage belongs to neural networks. The two-layer feed-forward network capitalizes on interactions and nonlinearities and, therefore, delivers the best results. An equal-weighted quintile of markets with the top forecasts outperforms the low-rated countries by 1.69% per month. The Sharpe ratio that is associated with such a long-short strategy equals 1.57.

Importantly, the impressive profits on machine learning portfolios do not come from their exposure to common factors. The abnormal returns survive even after controlling for stock-level and country-level value, momentum, size, profitability, and investment effects. The alphas remain both sizeable and robust.

However, from a practical perspective, the machine learning strategies come with a caveat. Because some of the market characteristics are short-term in nature, the portfolios exhibit substantial turnover. Similarly, as seen in the seminal stock-level study of Gu et al. (2020), the long-short strategies require replacing about half the portfolio each month. Although this may generate substantial trading costs, the portfolio rotation could be reduced in at least two ways. First, unlike typical stock-level anomalies (Stambaugh et al., 2012), most of the alphas on the long-short machine learning strategies come from the long side. Hence, the strategies can be effectively implemented via a long-only approach with a limited decline in risk-adjusted performance. Second, the machine learning signals prove relatively persistent through time. Consequently, even if the portfolios are reformed only once in 12 months, they continue to produce significant profits—albeit of limited magnitude.

Fourth, we shed light on the sources of the cross-sectional predictability of country equity returns. The popular narrative on stock return predictability is linked with two competing explanations: risk vs. mispricing. Having tested both, we find no convincing evidence in support of the risk story. Bivariate sorts on country risk changes and machine learning forecasts reveal no link between variation in sovereign, financial, or political risk and return predictability. On the other hand, the predictability of market returns visibly interacts with mispricing. The abnormal returns on machine learning strategies are higher in both overpriced and underpriced markets, and visibly weaker in the countries with neutral pricing. To sum up, our findings favor mispricing as the critical driver of the predictability of country equity returns.

Fifth, our final tests provide insights into time-series and cross-sectional variation in the predictability of country equity risk premia. The mispricing roots of the predictability have potential implications for its magnitude through time and across markets. To begin with the time-series dynamics, voluminous evidence from the security level points out to a gradual decay in return predictability; this is typically linked with investor learning, falling limits to arbitrage, or an improvement in market efficiency—which drive down asset mispricing overtime (e.g., Schwert, 2003; Chordia et al., 2014; McLean & Pontiff, 2016; Calluzzo et al., 2019). We find that the cross-section of country equity returns is, apparently, plagued by a similar problem. While the information content of country characteristics was clear in both the 1990s and 2000s, their relevance has declined over the past decade. In consequence, the weaker return predictability leads to lower—though still observable—profits on machine learning portfolios. Although the abnormal returns survive through the entire study period, in its second half, they are roughly 50% lower than in its first half.

The mispricing story also bears implications for international heterogeneity. If the machine learning alphas are driven by mispricing, they should be boosted by high limits to arbitrage. The empirical evidence supports this view. Although the return predictability is not limited to a particular market segment; it is measurably stronger in places where capital moves slower: across smaller and emerging markets with lower liquidity and higher idiosyncratic risk. The return predictability improves in these segments for most of the machine learning models we test.

Our findings contribute to two major strains of asset pricing literature. First, we extend the research on machine learning applications in the cross-section of returns. Specifically, we are the first to explore the international country equity risk premia. While earlier studies gained insights from including multiple other asset classes, such as U.S. stocks, (e.g., Freyberger et al., 2020; Gu et al., 2020, 2021; Avramov et al., 2021; Han et al., 2021), international equities (Leippold et al., 2021; Drobetz & Otto, 2021; Jiang et al., 2018; Tobek & Hronec, 2021; Choi et al., 2021), corporate bonds (Bali et al., 2021), U.S. Treasury bonds (Bianchi et al., 2021), commodities (Struck & Cheng, 2020; Rad et al., 2021), industries (Rapach et al. 2019), and currencies (Filippou et al., 2020), the cross-section of international risk premia remained unexplored.

Second, we add to the research on the predictability of cross-section of country equity returns. Earlier papers mainly focused on aggregate-level counterparts of individual stock-level anomalies, such as value, size, momentum, reversal, idiosyncratic and systematic risk, or seasonality.[2] Moreover, many articles considered the role of various country-specific political and economic risks in asset pricing (e.g., Erb et al., 1995, 1996; Diamonte et al., 1996; Bekaert et al., 1997; Avramov et al., 2012). Few studies of multiple predictors examined them mainly in the context of their replicability and reliability (e.g., Zaremba et al., 2020; Baltussen et al., 2021). On the contrary, we integrate numerous variables using machine learning models in order to better understand the dynamics of international country risk premia.

The remainder of the paper proceeds as follows. Section 2 presents the data and methods. Section 3 reports the major empirical findings. Section 4 discusses the portfolio implementation. Section 5 focuses on the sources of the return predictability, and Section 6 explores its variation across space and time. Finally, Section 7 concludes.

2. Research Design

We start by outlining the data and variables utilized in this study. Subsequently, we discuss the machine learning methods that are employed.

2.1. Data Sources and Sample Preparation

Our sample encompasses a total of 71 country stock markets; the detailed composition is stipulated in Table A1 in the Online Appendix. The study period runs from January 1985 to April 2021; however, older data is also used to calculate various variables when necessary. In general, we aim to build a possibly comprehensive representation of global stock markets, and its timeline and composition are dictated by data availability.

[2] See, e.g., for value: Asness, Moskowitz, and Pedersen (2013), Baltussen, Swinkels, and Van Vliet (2021); for size: Asness, Liew, and Stevens (1997), Fisher, Shah, and Titman (2017); for momentum: Chan, Hameed, and Tong (2000), Bhojraj and Swaminathan (2006), Asness, Moskowitz, and Pedersen (2013), Geczy and Samonov (2017), Pitkäjärvi et al. (2020); for idiosyncratic risk: Bali and Cakici (2010), Umutlu (2015); for systematic risk: Frazzini and Pedersen (2014); for reversal: Balvers et al. (2000); and for seasonality: Keloharju, Linnainmaa, and Nyberg (2016, 2021).

As in Baltussen et al. (2021) and Zhang and Jacobsen (2021), we enhance market coverage by combining data from different sources. We calculate stock market returns using Datastream Global Equity Indices, representing value-weighted portfolios covering most of the investable equity universe in their respective countries (Thomson Reuters, 2008). Thanks to their comparability across countries, the Datastream indices are a common choice in the studies of global equity risk premia (e.g., Chan et al., 2000; Ferreira & Gama, 2007; Bali & Cakici, 2010; Brusa et al., 2020; Zhang & Jacobsen, 2021). In the case of data unavailability, we extend the Datastream time-series (typically backfill) with Global Financial Data (GFD) Equity Indices. By assuring an extensive long-run historical coverage, the GFD indices have recently gained popularity in the examinations of global stock returns (e.g., Hjalmarsson, 2010; Zhang & Jacobsen, 2013; Albuquerque et al., 2015; Muir, 2017; Danielsson et al., 2018; Bekaert & Mehl, 2019; Miranda-Agrippino & Rey, 2020; Cortes et al., 2021).

To assure the data quality, we replicate the screens from Baltussen et al. (2021). Specifically, we ascertain that there are no zero, missing, or stale returns, nor any return interpolation. Furthermore, we eliminate the hyperinflation episodes. Building on the definition of Cagan (1956), if the ex-ante level of monthly inflation rate exceeds 50%, we discard all the observations within the subsequent 12 months.

As in Fama and French (2012, 2017), we express all the stock market data (including the returns) in U.S. dollars. This approach allows us to cope with all the issues associated with foreign-exchange conversions and currency risk, as well as align our paper with a practical perspective of a U.S. investor. Consistent with this framework, we represent the risk-free return with the one-month U.S. Treasury bill rate from French (2022).

The number of countries in the sample increases gradually along with the evolution of global stock markets, from 31 in 1985 to reach 71 in 2009, with the time-series average of 59. The total number of return-month observations is 25,789; however, the data available for specific variables may be lower. Figure 1 displays the size of our sample over time. In addition, Table A1 from the Online Appendix details the statistical properties.

[Insert Figure 1 about here]

2.2. Stock Market Characteristics

With the country sample at hand, we form a collection of return predictors for the aggregate stock market returns. To this end, we identify, classify, and reproduce 88 country characteristics from the asset pricing literature. These variables could be broadly categorized into two major groups: 1) replications of firm-level anomalies at the aggregate stock market level; and 2) country-specific macroeconomic or political features.

Within the first category, we consider the anomalies and risk factors that are documented in major finance journals. To keep our examination meaningful, manageable, and of practical relevance, we impose several conditions to include an anomaly in the sample. First, the stock market anomaly needs to have been demonstrated to hold—in a direct or a closely-related form—also at the level of country equity indices. Second, the return predicting signal can be derived from market or accounting data using standard databases, such as Datastream, GFD, or Bloomberg. Third, the anomaly pertains to the cross-section of returns—rather than time-series or seasonal patterns—and can be implemented via traditional quantile portfolios. Fourth, it can be captured at a monthly frequency. Our final selection return predictors can be classified into several groups that share a similar economic intuition: a) value vs. growth; b) size and liquidity; c) price risk; d) momentum; e) seasonality; f) profitability; g) indebtedness; h) skewness; i) long-term reversal; j) technical analysis; and k) investment and issuance.

The second major category contains variables that exist only at the level of countries and does not have their explicit counterparts for individual stocks. Again, we solely focus on country characteristics that have been explored within finance literature for their predictive powers over the country equity returns. This class of features encompasses principally macroeconomic conditions—variables derived from government bond and bill markets—as well as financial, economic, and political risks.

Overall, our sample comprises 88 variables, forming—to the best of our knowledge—the most comprehensive sample of equity country predictors ever considered. Table 1 contains their brief summary; furthermore, Table A2 in the Online Appendix details the calculation procedures, along with the essential literature references and data sources. The variables are calculated using various data sources; besides Datastream and GFD, we also rely on Bloomberg, PRS Group, or Varieties of Democracy (V-Dem)—where needed.[3] All the accounting and macro are based on lagged data to avoid a look-ahead bias. As in Gu et al. (2020), any missing values are replaced by the cross-sectional median. Finally, for each month, we standardize all the variables cross-sectionally to have a zero mean and a standard deviation of 1.[4]

[Insert Table 1 about here]

[3] For PRS Group, see: https://www.prsgroup.com/; for V-Dem: https://www.v-dem.net/.

[4] Notably, our approach here departs from Kelly et al. (2019), Gu et al. (2020), and Leippold et al. (2021), who cross-sectionally rank all the characteristics month-by-month, subsequently mapping them into the [-1,1] interval. By using standardization, we seek to keep the information on the magnitude of different variables—which is otherwise lost in the ranking process. In an unreported analysis, we find and compare the two methods and find the results qualitatively similar; furthermore, the standardization leads to only marginally better performance.

2.3. Machine Learning Methods

Following Gu et al. (2020), we employ a general additive prediction model to describe the association between stock markets' excess return and its different characteristics:

$$r_{i,t+1} = E_t(r_{i,t+1}) + \varepsilon_{i,t+1}, \tag{1}$$

where $r_{i,t+1}$ denotes the excess return on index $i = 1, \dots, N_T$ in month $t = 1, \dots, T$. The expected excess returns are calculated as a constant function of predictor variables available at period t:

$$E_t(r_{i,t+1}) = g(z_{i,t}), \tag{2}$$

where $z_{i,t}$ indicates a P-dimensional vector of return predicting variables. Notably, as in Gu et al. (2020), $g(z_{i,t})$ estimates the expected returns independently of any information before t or from other markets than i. The vector N_{t+1} comprises the 88 market characteristics from Table 1.

The precise form of the model $g(z_{i,t})$ is left unspecified. Hence, the approximation functions are both flexible and family-specific and can be parametric and non-parametric, as well as linear or nonlinear. Despite these differences, all prediction models are constructed to approximate the true returns by minimizing the out-of-sample mean squared forecast error:

$$MSFE_{t+1} = \tfrac{1}{N_{t+1}} \textstyle\sum_{i=1}^{N_{t+1}} (\hat{\varepsilon}_{i,t+1})^2, \tag{3}$$

where $\hat{\varepsilon}_{i,t+1}$ represents the individual prediction error for the country stock market i coming from the forecast of a given model, and N_{t+1} is the number of markets at period t+1. Our overall aim is to search for the forecasting model from a pool of candidates that exhibits a superior prediction performance.

Our selection of machine learning models builds on the works of Gu et al. (2020), Bali et al. (2021), and Leippold et al. (2021). Specifically, we adopt 12 different methods: ordinary least squares (OLS) regression, partial least squares (PLS), principal component analysis (PCA), least absolute shrinkage and selection operator (LASSO), elastic net (ENET), support vector machines (SVM), gradient boosted regression trees (GBRT), random forest (RF), and feed-forward neural networks with one to three layers (FFN1, FFN2, FFN3). Moreover—following the arguments seen in Rapach et al. (2010) and Chen et al. (2020)—we also calculate a combination forecast (COMB) that averages individual return predictions from the 11 machine learning models stated above. A detailed description of the models that are employed is provided in Section B of the Online Appendix.

We estimate the models, select the hyperparameters, and assess their performance following the typical methods found in the literature. We pursue an increasing window approach and split our study period into three separate subsamples while holding the temporal ordering: the training sample (1985 to 1991), the validation sample (1992 to 1994), and the testing sample (1995 to 2021). In the first step, the training sample is used to estimate the model parameters that are subject to some pre-specified model family-specific hyperparameters. Subsequently, the validation sample is utilized to tune the model's hyperparameters subject to the objective loss function (Section B of the Online Appendix contains further details on models' hyperparameters).[5] Last, we test the model using the single month right after the validation sample; this testing month never enters the training and validation samples.

Notably, Gu et al. (2020) only refit the prediction models annually (rather than monthly) due to the substantial computational intensity of their machine learning models. Since our sample of country indices is cross-sectionally smaller, we re-estimate the models each month. In line with the increasing window approach, whenever we refit the model, we increase the training period by one month while holding the length of the validation sample constant (three years).

3. Baseline Empirical Findings

We begin by exploring the forecasting abilities of different factor models; next, we explore the major drivers of predictability of the cross-section of country equity returns.

3.1. Predictive Performance of the Machine Learning Models

Table 2 presents the overall assessment and comparison of the machine learning models' predictive performance. We run four different tests. First—as in Gue et al. (2020)—we compute out-of-sample predictive R^2 metrics. Second, building on Lewellen (2015) and Drobetz et al. (2019), we estimate out-of-sample predictive slopes. Third, we introduce a new rank-based R^2 evaluation metric. Finally, to evaluate the relative forecasting effectiveness of different models, we conduct pairwise comparisons using a modified Diebold and Mariano (1995) test.

The first row of Table 2, Panel A reports the out-of-sample predictive R^2 measures (R^2_{OOS}). We closely follow Gu et al. (2020) and estimate the R^2 based on our test sample and re-estimation dates. Overall, our results resemble earlier applications of the machine

[5] If a model does not involve a validation sample, as in the case of OLS, then the training sample is extended to include the original validation period. For example, the first training sample is 1985-1992.

learning methods to the cross-section returns (Gu et al., 2020; Drobetz & Otto, 2021; Leippold et al., 2021), and the R^2_{OOS} exhibit a similar order of magnitude.

[Insert Table 2 about here]

The simple OLS method employing all 88 market characteristics yields the R^2_{OOS} of -0.14%. The poor performance, matching the earlier findings of Gu et al. (2020), signifies that the OLS is beaten by a mere naïve forecast that assumes zero returns on all stocks. The OLS lacks any form of regularization, so the reliance on numerous potential predictors makes it prone to overfitting. This weakness, resulting in low R^2_{OOS} readings, may be overcome via dimension reduction techniques or enforcing a sparser model by penalizing excessive covariates.

The dimension reduction methods do a mixed job in improving the OLS performance. While PLS fails to exhibit a substantial improvement, PCA shows a positive R^2_{OOS} of 0.28%. The penalized regressions seem to be a more effective method regularization technique, effectively boosting the predictive abilities further. The R^2_{OOS} measures for LASSO and ENET amount to 0.91% and 0.90%, respectively. Both algorithms display very similar performance, suggesting that the precise form of the penalty term in these functions is of little importance. Finally, the SVM method leads to even further improvement—raising R^2_{OOS} to 1.47%.

Noteworthily, the overall predictive performance of the regularized techniques dominates the seminal findings from the U.S. market by Gu et al. (2020). For example, their baseline R^2_{OOS} for the elastic net reaches the level of 0.11%; i.e., more than 80% lower than in our case. This may be unsurprising as the cross-section of country equity returns is much narrower, encompassing considerably fewer assets relative to the number of available market characteristics. Furthermore, aggregation of individual stocks into country portfolios diminishes the impact of extreme observations.

The regression tree methods, RF and GBRT, fail to demonstrate their competitiveness when compared to simple regularized regressions. The R^2 does not reveal a substantial improvement that is relative to penalized regressions, suggesting that the two techniques may be prone to overfitting despite the boosting (Friedman et al., 2000; Fiedman, 2001) and bagging (Breiman, 2001) regularizers that are embedded in these methods. Our country-level findings, in this regard, are visibly weaker than in earlier stock-level research (Gu et al., 2020).

On the other hand, the neural networks exhibit sizeable R^2_{OOS}—especially in the case of multiple hidden layers. The FFN1, comprising one hidden layer, has the R^2_{OOS} value of 1.52%. For the FFN2, this metric equals 1.29%. Finally, for FFN3—which contains three hidden layers—R^2_{OOS} reaches 1.89%. Consequently, according to this metric, FFN3 exhibits the best predictive performance among all the *individual* models. Unlike simple regularized (or unregularized) regressions, neural networks effectively capture both nonlinear relationships and complex feature interactions. These benefits increase along with the depth of the neural network. The superiority of the neural networks is in line with both Gu et al. (2020) and Leippold et al. (2021), who also count it among the most effective forecasting techniques.

Last, the top-performing prediction method is COMB. The combination of every forecasting technique produces R^2_{OOS} of 2.21%, noticeably dominating the individual methods. The greatest benefit of forecast combination is that it effectively reduces forecast variance that is associated with particular models. In consequence, reducing the impact of uncorrelated prediction errors generates more accurate forecasts (Rapach et al., 2010; Chen et al., 2020). Our findings match the observations of Bali et al. (2021), who also document substantial gains from averaging forecasts from different models and superior performance of the ensemble methods. In a nutshell, no model is the best; however, when combined, their performance thrives.

The out-of-sample R^2 coefficient is the most popular evaluation metric in machine learning literature. However, it is not free of flaws. While the pure R^2_{OO} measure may be disappointing for specific models, it can also be irrelevant. A large portion of the global fit is driven by the variance of the forecasts and realized returns; hence, the picture of the actual correlation between both the predicted and realized returns may be blurred (Coqueret, 2022). In consequence, investors may realize large economic gains—even if R^2_{OOS} is large and negative (Kelly et al., 2021). To cope with these issues, we supplement a predictive power assessment with two further measures: predictive slopes and rank-based correlations.

The second row of Table 2, Panel A, uncovers the predictive slopes (PS_{OOS}) originating from Drobetsz and Otto (2021). These measures are calculated based on pooled regressions of the monthly realized excess returns on the corresponding predictions from the machine learning models. The slopes close to one indicate that the forecast dispersion essentially mirrors the cross-sectional variation in country risk premia. On the other hand, the predictive slopes are larger (smaller) than one—implying overly narrow (wide) predictions.

A quick overview of the predictive slopes broadly confirms the conclusions from the R^2_{OOS} coefficients. OLS, PLS, and PCA display low PS_{OOS} levels of approximately 0.5. This suggests a substantially lower realized return dispersion than what is seen in the models' forecasts. On the contrary, however, the LASSO and ENET predictions typically undershoot the actual returns. Their PS_{OOS} equal 1.24 and 1.25, respectively. These elevated values contain a clue that the traditional predictive R^2 may undervalue the actual economic gains from utilizing the forecasts; this is because the global fit may be drowned in the noise of the variance of realized returns.

Next, SVM works relatively well, with an average slope of 0.85. On the other hand, both GBRT and RF's performances are particularly disappointing—which corroborates the observations from the out-of-sample R^2 coefficients. The respective PS_{OOS} do not exceed 0.48. Turning to the neural networks, their accuracy appears to be better than simple regularized regressions or tree methods; their slopes range from 0.6 to 0.7. Finally, the two top-performing methods—according to the predictive slopes metric—seem to be SVM and COMB. The combination forecast exhibits a slope of 0.83, highlighting the benefits of averaging the individual predictions again.

Quantitative portfolio managers typically form portfolios by sorting stocks on their expected returns. Hence, from a practical perspective, it is of interest—not only by how accurately a model predicts future returns—but whether it can currently rank the stock from the best to the worst. In other words, to what extent the model can effectively separate losers from winners. As noted by Coqueret (2022) and Kelly et al. (2021), this information—oftentimes—cannot be inferred from the traditional R^2_{OOS}.

To shed light on this issue, we propose a new metric that is based on rank correlation. We aim to capture to what extent a model ranks the assets consistently with their *ex-post* realized returns. For each month t, via the use of the test sample, we transform the predicted and realized returns into ranks i from 1 to N_t—where N_t denotes the number of available markets. Next, we map both the predictions and realizations into the interval [0,1]. Finally, we calculate the pseudo R^2 metric of Cox and Snell (1989) in order to gauge the link between the order of forecasted and realized payoffs.

The third row of Table 2, Panel A tabulates the outcomes of this exercise. The conclusions differ partly from the earlier analysis of the traditional R^2_{OOS}. First and foremost, all the techniques yield positive and sizeable R^2 values. In other words, all the methods do a decent job in ordering the markets. Notably, even if some models display negative classical R^2_{OOS}, they may still be quite effective in separating losers from winners. Looking further into the details, the relative efficiency of different methods resembles our previous observations. The worst performing algorithm is GBRT, suggesting that its predictions may not always translate into successful portfolios from one-way sorts. Conversely, the top performers among the individual techniques are LASSO, ENET, and SVM. Moreover, in line with our earlier findings, the combination method (COMB) also performs very well.[6]

[6] Importantly—despite its outstanding performance—the COMB model is dominated in this test by LASSO, ENET, and SVM. As noted by Bali et al. (2021), this is because its efficiency depends on the tradeoff between the reduction in model bias and variance (Rapach et al., 2010). The forecast combination

Last, Table 2, Panel B displays the pairwise comparisons of the predictions from different machine learning models using the modified test of Diebold and Mariano (2021), abbreviated DM. In essence, the DM test statistic compares the mean squared forecast errors to gauge which candidate produces more accurate forecasts. Our implementation closely follows Drobetz and Otto (2021); the DM statistic is computed as:

$$DM_{a,b} = \frac{\bar{d}_{a,b}}{\hat{\sigma}_{\bar{d}_{a,b}}}, \tag{4}$$

where $d_{a,b,t+1} = MSFE_{t+1}^{(a)} - MSFE_{t+1}^{(b)}$ denotes the differences in the monthly mean squared forecast errors of models a and b, $\bar{d}_{a,b} = \frac{d_{a,b,t+1}}{T}$ indicates the time-series average of these differences; furthermore, $\hat{\sigma}_{\bar{d}_{a,b}}$ is the Newey-West (1987) adjusted standard error. The DM test statistics follow the standard normal distribution. It is worth noting that we interpret them in two separate ways. First, to facilitate individual pairwise comparisons, we determine the standalone 5%-significance threshold corresponding with the |t-stat| of 1.96. Second, since we explore 12 models jointly, we address the multiple hypothesis problem by applying the Bonferroni correction (for discussion—see, e.g., Harvey et al., 2016). The adjusted hurdle for the *t*-statistics equals 2.87.

The conclusions from the DM tests are broadly in line with our earlier findings that pertain to predictive R^2 and slopes. Though not all differences are significant, we observe measurable gains from combining different forecasting methods together. The performance of the forecast combination method noticeably stands out. The COMB model reliably outperforms the individual algorithms in most cases. Again, this corroborates our earlier finding that whereas individual models have their ups and downs, the combination effectively extracts their strengths.

3.2. Which Market Characteristics Matter?

Having tested the overall predictive abilities of different machine learning models, we now explore the relative importance of individual country characteristics. We want to identify the crucial drivers of the cross-section of country returns while accounting for the impact of the entire "zoo" of predictors in the system. To ascertain the contribution of individual covariates, we follow the approach originating from Kelly et al. (2019). We compute the variable importance, denoted VI, of a given predictor as the reduction in the predictive out-of-sample R^2 from setting all its values to zero while holding the other model estimates as fixed.

is an effective tool in decreasing the prediction variance; however, it may simultaneously augment the model's bias. At the same time, some individual models may exhibit a superior ability in reducing biases—overcoming the costs associated with elevated variance.

We begin by presenting a simple ranking of variable importance for the 12 machine learning methods. Figure 2 depicts the model-specific hierarchy of characteristics by assigning the color gradient to covariates, where the darkest (lightest) hue stipulates the most (least) important predictors. The variables are sorted according to their average rank across the 12 methods.

Interestingly, the various machine learning techniques are in close agreement on the essential variables. The most influential predictors are the market size (*MV*) and long-run return reversal (*LtRev*). Furthermore, many key variables pertain to the short-term past performance and belong to the momentum or technical analysis categories. This comprises various variants of momentum (*LtMom*, *MtMom*). In addition, two popular technical indicators are also included: the first indicator, market breadth (*BRTH*), represents the differences in the numbers of rising and falling stocks; the second indicator, moving average difference (*MAD*), compares the levels of long- and short-term averages. Among the valuation ratios, the earnings yield (*EP*) plays the first fiddle. Several models also emphasize net share issuance (*NSI*), mirroring the analogous firm-level anomaly (Pontiff & Woodgate, 2008). The role of macroeconomic variables is of lesser importance; furthermore, the top positions are taken by the inflation rate (*Infl*) and the real effective exchange rate dynamics (*REERCh*). Interestingly, numerous popular risk factors—such as credit, liquidity, political risk, or overall idiosyncratic risk—reach lower grades in the importance ranking. Only bureaucracy quality (*BurQual*) and control of corruption (*Corr*) appear to play some role.

[Insert Figure 2 about here]

As previously noted, most of the models designate similar features like the essential drivers of stock returns; this places most of the weight on the combination of size, value, long-term reversal, and momentum variables. On the other hand, PCA and the tree methods—including RF in particular—are more democratic, spreading the importance weights across other covariates.

Figure 3 sheds further light on the issue of variable importance by depicting the specific and precise R^2 reduction for the top 10 variables of each of the models. Most commonly, the leading variable is *LtRev*; it is then closely followed by predictors such as *MV*, *EP*, *BRTH*, or *LtMom*. RF sorts the variables differently, placing *LtRev* at a lower position. Yet, still, the top ranks in this method include technical analysis signals.

[Insert Figure 3 about here]

Notably, most methods favor a sparse selection with just a few factors that explain most of the cross-section of returns. While the top predictors are associated with very high R^2 reductions, the importance of the remaining positions declines rapidly. The average aggregate importance of the 10 (five) top variables across all 12 models equals 57% (38%). This concentration is particularly pronounced for the regularized regressions and support vector machines. For example, in the case of ENET, the variable importance of the top five predictors (*LtRev*, *EP*, *BRTH*, *NSI*, and *LtMom*) adds up to 67%.

The observations above lead to a surprising conclusion concerning asset pricing in global markets. Although the finance literature has cataloged a plethora of predictors of the cross-section of country risk premia, it appears that only a handful of them really matter. This apparent multidimensionality can be potentially reduced to just a few fundamental phenomena (such as size, value, momentum, and reversal) that effectively capture the most cross-sectional variability in country equity returns.

Last, to supplement our analyses so far, we explore the importance of different groups of covariates. This additional test helps to uncover some variables that may be of minor importance on a standalone basis; however, as groups, they exert a measurable impact on asset pricing. To achieve this, we add the variable importance by category—as defined in Table 2. Figure 4 summarizes the results of this experiment.

[Insert Figure 4 about here]

In total, the most important groups of covariates pertain to long-term reversal and value versus growth phenomena. Not surprisingly, this is closely followed by both momentum and technical analysis variables. Interestingly, the regression trees models and neural networks also emphasize political risk and regimes. These two classes of machine learning techniques effectively integrate nonlinearities and interactions. Hence, they may capture—for example—the heightened importance of political risk in smaller markets, which evades the estimations in simple linear models. To conclude, once considered together as a group, political risks may contain incremental information pertaining to asset pricing in global equity markets.

4. Portfolio Analysis

Having established the basic properties of the machine learning predictions, we are now interested in whether they can be exploited in practice. Hence, we examine the profitability of machine learning strategies. Furthermore, we explore further practical aspects of portfolio implementation and their stability over time.

4.1. Machine Learning Portfolios

To capture the economic implications of the return predictability, we now continue with portfolio analysis. To keep our research both simple and intuitive, we form portfolios from one-way sorts on the predictions from the machine learning models. To this end, each month, we rank all the countries in the sample on their return forecasts for one month ahead. Subsequently, we sort the markets into quintiles and form both equal- and value-weighted portfolios.[7] Furthermore, we calculate a zero-investment hedge portfolio that assumes a long position in the quintile of markets with the highest returns predictions and, vice versa, a short position in the countries with the lowest forecasted payoffs.

Table 3 presents the performance of portfolios from univariate sorts on machine learning predictions. Specifically, we report the average realized and predicted returns per market quintile—as well as their annualized Sharpe ratios. We also calculate alphas from the global CAPM, where the market risk factor is proxied by the excess return on a value-weighted portfolio of global stocks.[8]

[Insert Table 3 about here]

A quick overview of the results indicates that all machine learning techniques can be coined into effective country allocation strategies. We can observe a monotonic (or nearly monotonic) pattern in the cross-section of realized returns in all cases. Moreover, in all circumstances, the long-short portfolios produce sizeable and significant abnormal returns—albeit their magnitude differs across the prediction techniques.

The average return on the equal-weighted spread portfolio across all 12 models amounts to 1.44% per month. Interestingly, even the modest OLS proves very efficient—producing both robust and profitable portfolios. The equal-weighted long-short portfolio yields a mean monthly return of 1.39% (t-stat = 6.31) and an associated alpha of 1.44 (t-stat = 6.44). The corresponding Sharpe ratio equals 1.21.

The simple dimension reduction techniques and regularized regressions do not significantly improve strategy performance. Both the return and alphas on PLS, PCA, LASSO, or ENET portfolios are qualitatively similar to OLS. Apparently, overfitting is not a major issue that is dampening the performance of international country allocation.

On the other hand, what does make a difference is effective accounting for nonlinearities and variable interactions. In consequence, neural network predictions prove highly effective in portfolio formation. The model with two hidden layers, FFN2, produces the best portfolios across all the considered techniques. The average monthly return on the equal-weighted long-short strategy is 1.69% (t-stat = 7.57) and the corresponding alpha equals 1.75% (t-stat = 7.95). Furthermore, FFN2 is also the winner in terms of the

[7] To assure that the biggest countries do not dominate the portfolios, we closely follow Jensen, Kelly, and Pedersen (2021) and winsorize the market equity of the largest markets at the 80th percentile. This operation seeks to form tradable, yet balanced, strategies.

[8] We represent the global portfolio with the Datastream World Market Index.

Sharpe ratio—which equals 1.57. The superior performance of neural networks matches the findings of Gu et al. (2020) and Leippold et al. (2021)—who also deem these methods highly successful. Nonetheless, contrary to their findings, we do not observe substantial benefits of including additional hidden layers beyond two. The performance of FFN3 does not beat FFN2. Apparently, the nonlinearities and interactions in the universe of country equity indices—which is considerably lower than the universe of individual stocks—can be effectively handled by just two hidden layers.

So far, our considerations have focused on individual machine learning techniques. Nonetheless, besides FFN2, another champion in the portfolio horserace also the *COMB* strategy. Blending individual predictions into a combination produces an impressive portfolio performance. The average monthly return equals 1.64% (t-stat = 6.75) and the associated alpha is 1.71% (t-stat = 7.07). Hence, the overall conclusion from this exercise is similar to the context of prediction accuracy: forecast combinations effectively eliminate the noise of individual models. In consequence, while different techniques have their pros and cons, the combination method clearly stands on the podium.

The discussion has, so far, concentrated on equal-weighted portfolios. Yet, all the strategies also work effectively in the value-weighted framework—even though the abnormal returns are somewhat lower. For example, the value-weighted spread portfolio based on the COMB model displays a mean return of 0.98% (t-stat = 3.26) and an alpha of 0.87% (t-stat = 3.04). Overall, across all the strategies, the equal-weighted hedge portfolios produce average returns approximately 67% higher than their value-weighted counterparts. Our findings in this regard are qualitatively similar to the stock-level evidence. For example, Drobetz and Otto (2021) also found that the equal-weighted strategies beat the capitalization-weighted ones by more than 75%. The difference is associated with stronger return predictability in smaller firms and markets.

Last, the final insight from Table 3 concerns the asymmetry in the cross-section of market returns. Across virtually all strategies, the abnormal returns on the spread portfolios principally come from the long side rather than short trades. The abnormal returns, in absolute terms, are typically higher for the top quintile than the bottom ones. On the one hand, this differs from the firm-level research—which typically attributes mispricing to the short legs (Stambaugh et al., 2012). On the other hand, this phenomenon has critical practical implications. Specifically, it allows investors to capture larger parts of the abnormal returns with the necessity of short-selling—which may be costly or even unavailable.

4.2. Practical Investor Perspective

To reflect deeper on the practical aspects of the international equity strategies building on machine learning, we run several additional calculations. First, following Gu et al.

(2020), we compute maximum monthly losses and drawdowns during the examination period. Second, we scrutinize the risk-adjusted performance in terms of multifactor models. Third, we check the portfolio turnover to understand the impact of trading costs. Finally, we examine the performance of strategies with extended holding periods.

Table 4 reveals the first set of results of these tests, with Panels A and B concerning equa-weighted and value-weighted strategies, respectively. Panels A.1 and B.1 repor the maximum monthly losses and total drawdowns during the test period from 1995 to 2021. The worst months for the equal-weighted portfolios (Table 4, Panel A) were associated with losses in the range of 9.82% to 13.01%, depending on the machine learning technique. The drawdowns, in turn, ranged from 24.37% to 27.93%. The similar numbers of the value weighed portfolios were, on average, slightly higher; for example, the maximum daily losses were between 10.47% and 18.91%. This riskier behavior is associated with lower diversification of these portfolios, as they tend to be more concentrated in a few large countries.

[Insert Table 4 about here]

Comparing the risk metrics that were mentioned above with the U.S. market evidence, our strategies appear substantially safer. Only the most sophisticated neural networks techniques in Gu et al. (2020) may compete with our portfolios in terms of drawdowns or maximum losses. The superior performance of our strategies stem, unsurprisingly, from their vast international diversification across multiple developed and emerging markets.

Next—as in Gu et al. (2020)—we are interested in whether the machine learning portfolios span popular factor strategies. Therefore, we test their performance with the Fama-French (2018) six-factor model; i.e., the five-factor model that is extended with momentum. We conduct this exercise in two ways. First, we utilize the standard stock-level international factors from French (2022). Second, we form analogous ad-hoc country-level factors. This alternative set builds on the same variables (book-to-market ratio, momentum, etc.); however, the portfolios comprise country indices and are structured identically as the evaluated strategies (i.e., equal- or value-weighted quintiles). This approach aims at assuring apple-to-apple comparisons; we want to ascertain that abnormal returns are solely driven by the return predicting signals and not by either asset universe or portfolio construction differences. For details of the country-level asset pricing factors, see Table A3 in the Online Appendix.

Panels A.2 and B.2 report the risk-adjusted returns. Overall, the multifactor models cannot explain the abnormal performance of the machine learning strategies. Their predictions go clearly beyond the simple asset pricing factors—such as value, size, or momentum. Like in the earlier test, particularly impressive alphas are recorded on the neural network and combination strategies; however, the abnormal returns are substantially positive in virtually all considered specifications.

Table 4, Panels A.3 and B.3, present the average turnover and breakeven trading costs on different machine learning techniques. We calculate the portfolio turnover for month t (PT_t) in line with Bollersev et al. (2018) and Koijen (2018), i.e., as the average share of a portfolio that needs to be replaced each month:

$$PT_t = \tfrac{1}{2}\textstyle\sum_{i=1}^{n}\left|w_{i,t-1} \times (1 + r_{i,t}) - w_{i,t}\right|, \tag{5}$$

where $w_{i,t-1}$ and $w_{i,t}$ are the weights of country i in the tested portfolio in two consecutive months, and $r_{i,t}$ is the country index return. Notably, in order to avoid double-counting the buys and sells, we calculate a one-sided (rather than two-sided) metric.

The portfolio turnover is generally high; however, it is not qualitatively more elevated than in stock-level machine learning strategies (Gu et al., 2020; Drobetz & Otto, 2021). In the case of the equal-weighted long-short portfolios, the average monthly turnover ratio ranges from 56.14% for PLS to 110.77% for RF. The elevated turnover has two major sources. First, the trading signals coming from the machine learnings techniques typically require dynamic portfolio rotation as they incorporate predictors—which may be short-term in nature. For example, the predictions in Gu et al. (2020) largely build on the short-term reversal effect—which is an anomaly that requires active portfolio reconstruction (Novy-Marx & Velikov, 2016). Likewise, our forecasts frequently incorporate the market breadth signal—which is also short-term in nature (Zaremba et al., 2021). Second, another contributing factor to the high turnover is the character of country portfolios. Our quintile portfolios, on average, comprise about 10 markets. Hence, replacing just one country in the portfolio automatically generates a turnover of approximately 10%. Finally, as the turnover derives mainly from changes in the composition—rather than rebalancing—the value-weighted portfolios reveal even higher portfolio rotation.

The breakeven costs for the equal-weighted long-short strategies range from 0.70% (RF) to 1.44% (FFN). The trading cost threshold for the combination strategy is 1.19%. The cost-efficiency of the machine learning strategies may be improved in at least two ways. First, by embracing long-only portfolios. Empirical evidence shows that the performance of long-only factor strategies does not linger far behind their long-short counterparts (Blitz et al., 2020). Furthermore, in our case, the long-only quintiles of the markets with the best return forecast do not fall vividly behind the spread strategies. For example, the equal-weighted long-only COMB strategy produces a mean return of 1.64% with a Sharpe ratio of 1.37; meanwhile, the long-only variant based on the top portfolios also yields 1.64% per month and with a Sharpe ratio of 1.05.

As seen in Table 4, pursuing the long-only strategies allows for the cutting of the portfolio turnover approximately by half. Consequently, the breakeven costs upsurge substantially, and their new range for the equal-weighted portfolios is 1.14% (RF) to 3.48% (FFN2). The new breakeven for the COMB portfolio is 2.57% per month.

Another simple yet popular option of coping with elevated transaction costs is extending the portfolio holding period (Novy-Marx & Velikov, 2019). Less frequent portfolio rebalancing leads to fewer trades and, in turn, lower costs. This, however, requires relatively persistent trading signals that predict returns further than just one month ahead. We investigate portfolios with extended holding periods in order to shed light on this point.

Table 5 reports the univariate portfolios that are formed on the machine learning forecasts using three-, six-, and 12-month holding periods. The portfolios are rebalanced monthly and, thus, incorporate an overlapping approach to holding periods. The overall results indicate that the machine learning profits are neither fragile nor short-term in nature. Although the magnitude of the abnormal returns declines along with the extension of the holding period, they remain robust and sizeable. Even if the portfolios are reformed only once in 12 months, the long-short strategies continue to produce significant abnormal returns.

[Insert Table 5 about here]

The alphas on the spread portfolios with the longest (12 month) holding period range from 0.36% to 0.84%. The best performing portfolio, in this case, is COMB. It exhibits a mean return of 0.73% (t-stat = 3.33) and the alphas equaling 0.84% (t-stat = 3.81). To sum up, despite the noticeable decline in profitability, the machine learning strategies survive—even in portfolios with 12-month holding periods.

5. The Sources of Return Predictability

Our evidence has, so far, demonstrated strong cross-sectional predictability of country equity premia around the world. We now explore the sources of this phenomenon. We confront two popular competing explanations: risk vs. mispricing. While neoclassical finance typically links return predictability with hidden risk premia, the behavioral view associates it with mispricing. Large-scale studies of *stock-level* anomalies seem to lean towards mispricing. For example, Engelberg et al. (2018) document that anomalies are incomparably stronger during earnings announcement days; they then link this observation with biased expectations and mispricing. Guo et al. (2020) reach similar conclusions—having studied the role of analysts' recommendations. Jiang et al. (2021) find anomalies more pronounced on high-attention days. Han (2021) decomposes anomaly returns into mispricing and risk constituents to demonstrate that only the first one plays a crucial role. Finally, Müller and Preissler (2021) also argue that risk cannot entirely explain anomaly returns. However, the evidence on predictability from other asset classes tends to be mixed. While Bartram et al. (2018) associate currency anomalies with mispricing, Choi and Kim (2018) and Bali et al. (2021)—who scrutinize corporate bonds—argue that risk-based explanations are more plausible.

What is the primary driving force behind the return predictability of country equity indices? To shed light on this issue, we replicate the tests from Bali et al. (2021). To begin with, we concentrate on the interactions between both risk changes and risk premia predictability. For example, Bali et al. (2020, 2021) document that swings in credit risk capture the uncertainty premium in asset prices; furthermore, Avramov et al. (2013) argue that variation in distress risk contributes to the occurrence of many anomalies. Bali et al. (2021) show that risk fluctuations contribute to return predictability of corporate bonds, but not individual stocks.

To picture various dimensions of country-specific risks, we use four different measures. First, we focus on numerical credit ratings—as in Bali et al. (2021).[9] We calculate an average rating from three major agencies; S&P, Moody's, and Fitch; and transform them into numeric scores—as in Avramov et al. (2013). We supplement the ratings with aggregate measures of a) financial; b) economic; and c) political risk from the International Country Risk Guide. With these four measures at hand, we first sort the markets into tertiles based on 24-month changes in risk estimates—as in Bali et al. (2021). Then, within each of the risk tertiles, the countries are sorted again based on the machine learning prediction. For the sake of brevity, we limit our presentation to the forecast combination (COMB); however, the results are qualitatively similar for individual prediction models—as well. The intersection produces nine double-sorted portfolios. Table 6 displays the results of this exercise.

[Insert Table 6 here]

The right-most columns of the table present the performance of long-short strategies that buy (sell) the markets with top (bottom) forecasts. First, the predictability is robust across all the risk-change tertiles. The mean returns and alphas are both positive and significant in all market segments.

The bottom rows of each panel present the difference-in-difference (diff-in-diff) test results, i.e., the spreads between the COMB strategy returns in the top and bottom risk-change subsamples. Overall, we observe no substantial influence of the risk changes on return predictability. The diff-in-diff returns and alphas are insignificant for three out of

[9] Avramov et al. (2013) shows that credit risk captured with sovereign ratings is priced in global equity markets.

four of our risk measures. The sole exception is an economic risk. To sum up, we do not find solid evidence to support risk-based roots of the return predictability. This conclusion aligns with Bali et al. (2021), who also do not observe such a link for equity markets.

We now continue the investigation with the influence of mispricing as a determinant of the return predictability in country equity indices. We expect the return predictability to be stronger in mispriced (overpriced or underpriced) markets. To explore this conjecture, we run two-way dependent sorts on mispricing (*MISP*) and expected returns.[10] We broadly follow our earlier approach from Table 6. Having initially grouped the countries into tertiles on *MISP*, we next sort them into tertiles on the machine learning predictions to obtain nine bivariate portfolios.

Table 7 reports the results of these tests. For conciseness, we only present the outcomes that pertain to the COMB model predictions. The other machine learning methods yield consistent results; therefore, we only briefly summarize them in Table A3 in the Online Appendix).

[Insert Table 7 here]

The markets with the highest predicted returns outperform those with the lowest predicted returns across all the MISP segments. The mean returns and alphas are both positive and significant in all three tertiles. Nonetheless, we can observe some heterogeneity across the subsets. The abnormal returns on the long-short portfolios that are formed on COMB forecasts are visibly stronger in the *Low MISP* and *High MISP* tertiles than in the *Medium MISP* one. The bottom section of Table 7 displays the difference-in-difference results, focusing on the spread between the extreme MISP tertiles and the middle one. The differences are significant for both overpriced and underpriced markets. This signifies that mispricing is a critical determinant of the country-level return predictability with machine learning models.

To conclude, among the two competing explanations—risk vs. mispricing—our evidence tends to lean towards mispricing. In this regard, our findings are entirely consistent with

---

[10] Bali et al. (2021) use the mispricing score (*MISP*) of Stambaugh, Yu, and Yuan (2015). This measure assesses the overall mispricing by aggregating 11 stock level anomaly variables. Because most of them do not have direct country-level counterparts, we compute an ad-hoc mispricing score based on established cross-sectional predictors of country index returns. Concretely, we use five variables: dividend yield (*DY*), momentum (*LtMom*), long-term reversal (*LtRev*), moving average (*MA*), and seasonality (*SEAS*) (e.g., Balvers et al., 2000; Asness et al., 2013; Keloharju et al., 2016; Zaremba et al., 2020; Baltussen et al., 2021; Ilmanen et al., 2021). We compute the average rank associated with these anomalies for each country, so that the higher (lower) value indicates a more overpriced (underpriced) market. The average ranks of the five predictors, rescaled to range between zero and 100, serves as the aggregate measure of mispricing (*MISP*). Countries with higher scores are deemed to be overpriced, and vice versa.

Bali et al. (2021). Their examination of machine learning models in the equity universe also favors mispricing-based versus risk explanation.

6. Global Variation in Return Predictability

Our considerations have, so far, concentrated on unconditional return predictability across the broad cross-section of markets. Nonetheless, the mispricing story yields testable implications on potential time-series and cross-sectional variation in return predictability. In this section, we explore these two issues further.

6.1. Does the Return Predictability Diminishes Over Time?

Asset pricing literature generally points out that equity anomalies weaken—or even disappear—over time. According to a popular narrative, investor learning, institutional trading activity, and improvements in market efficiency and liquidity drive the mispricing down (Schwer, 2003; Chordia et al., 2014; McLean & Pontiff, 2016; Calluzzo et al., 2019). Moreover, similar troubles may also plague the stock index anomalies (Zaremba et al., 2020). Hence, does the predictability of the cross-section of country risk premia weaken through time? Does the information content of country characteristics fade away?

Figure 5 illustrates the changes in the out-of-sample predictive R^2 coefficients through time. To reduce noise in the monthly values, we demonstrate rolling 10-year averages and report the values for both the traditional and rank-based R^2 measures. Our findings broadly match the view emerging from the stock-level anomaly literature. The predictability appears to gradually fade over time.

[Insert Figure 5 about here]

The R^2 coefficients that were relatively high in the 1990s and early 2000s then gradually declined through time. Whereas the magnitude of this decrease across various forecasting methods differs, the pattern is evident across all machine learning techniques. The precise timing of the decline is difficult to capture. Nevertheless, a brief overview—especially of the rank-based R^2 measures—suggests that the drop in predictability began following the Global Financial Crisis. Next, the R^2 measures reached a novel subdued plateau during the last decade.

The decline in predictability seems particularly detrimental when the traditional R^2 measure is examined (Figure 5, Panel A). In such a case, the R^2 coefficient has declined to approximately zero over the last decade. What this implies is that the return predictability essentially disappeared. The rank-based R^2 measure (Figure 5, Panel B), however, indeed decreased but remained substantially positive. The exact values in 2021 ranged between about 2% to 3.5%, meaning that market characteristics still contain valuable information about future returns. In other words, the machine learning strategies still separate market losers from winners; however, their efficiency is lower than 10 or 20 years earlier.

To better comprehend the economic importance of the drop in return predictability over time, we—again—turn to the portfolio analysis. Figure 6 plots the cumulative returns on long-short machine learning portfolios from Table 2 through time. Furthermore, Table 8 provides more formal insights by splitting the entire study period into halves.

[Insert Figure 6 about here]

[Insert Table 8 here]

These extra analyses confirm the diminishing efficiency of return forecasts. While the long-short machine learning strategies produce abnormal returns throughout the entire study period, their magnitude changes over time. The mean monthly returns on spread portfolios from 1995 to 2008 (Table 8, Panel 8) are between 1.63% and 2.32%, depending on the prediction technique. The best performing method, COMB, yields 2.32% per month (t-stat = 6.72). On the other hand, the average returns on the long-short strategies in the latter period (2008 to 2021) are lower—roughly by half. The average spread return ranges from 0.77% to 1.15%. The COMB strategy profits diminish to 0.95% monthly (t-stat = 3.32). To sum up, although market characteristics still predict future country equity returns, the strength of this relationship has noticeably weakened.

6.2. International Heterogeneity in Prediction Effectiveness

The behavioral narrative of equity anomalies argues that they are driven by investors' limited rationality, which cannot be easily arbitraged away (Pontiff, 1996; Shleifer & Vishny, 1997; Gromb & Vayanos, 2010). Hence, if the return predictability is mainly derived from mispricing, we would anticipate it to be boosted by high limits to arbitrage. Stock-level evidence tends to support this view, also in international markets (see, e.g., Watanabe et al., 2013; Hung et al., 2015; Azevado & Müller, 2020; Jacobs & Müller, 2020; Lam et al., 2020; Cakici & Zaremba, 2021). In order to explore this conjecture at the country level, we examine whether internationally varying limits to arbitrage affect the return predictability—as captured with machine learning models.

We employ four simple, yet common, proxies for limits to arbitrage: market size (*SIZE*), idiosyncratic risk (*IRISK*), liquidity (*LIQ*), and emerging market status (*EMER*).[11] We

[11] *IRISK* are binary variables taking a value of on when idiosyncratic volatility (*IVol*), as defined in Table A2 in the Online Appendix, take values higher than the cross-sectional median at t-1—and zero otherwise. *LIQ* is calculated identically using the Amihud illiquidity ratio (*Illiq*). *SIZE* is a dummy that takes a value of one (zero) if the market value (*MV*) at t-1 was lower (higher) than its cross-sectional median. Finally,

assume that limits of arbitrage are typically higher in small and emerging markets that are characterized by lower liquidity and higher idiosyncratic risk. Importantly, our simple measures tend to be positively correlated with more sophisticated metrics that capture market development: *de jure* and *de facto* indicators of financial openness, short-sale constraints, and other determinants of efficient capital movement across countries. Following the approach seen in Cosemans and Frehen (2021), we explore the impact of limits of arbitrage by interacting with the proxies above (*SIZE*, *IRISK*, *LIQ*, and *EMER*) in conjunction with the return forecast from machine learning models. We want to see whether stronger limits to arbitrage either improve or impair predictability.

As seen in Cosemans and Frehen (2021), we run Fama-MacBeth regressions with interaction terms. The dependent variable is the realized market return; furthermore, the independent variables include machine learning predictions and the interactions with the proxies for limits to arbitrage. Panel A shows the estimation of univariate regressions; Panels B to E focus on multivariate tests accounting for *SIZE*, *IRISK*, *LIQ*, and *EMER*.

[Insert Table 9 about here]

First of all, the machine learning forecasts are strongly associated with realized returns in all the specifications: both in the univariate and multivariate test. This means that they powerfully predict returns even after accounting for the role of market size, idiosyncratic risk, liquidity, and development. In other words, the predictability does not derive only from some dusty segment of small and illiquid global markets. However, this does not mean that limits to arbitrage do not play any sort of role. On the contrary, we observe strong interactions with each of the considered arbitrage constraint proxies for most (though not all) of the machine learning models. The influence of market size, liquidity, development, and idiosyncratic risk is evident for regularized regressions, dimension reduction techniques, and tree methods. This evidence indicates that the return predictability is, indeed, stronger across markets with higher limits to arbitrage. Furthermore, this complies with our findings in Section 6 that identify behavioral mispricing as the vital source of return predictability of country equity returns.

7. Conclusions

This paper employs machine learning methods to gain insights into an entirely new setting: the cross-section of country equity risk premia. To this end, we study data from 71 international stock markets from the years 1985 to 2021. We identify, classify, and reproduce 88 return predictors as inputs; with these variables at hand, we conduct an analysis using an array of different machine learning techniques: ordinary least squares,

---

EMER equals one if the market is classified as emerging in month *t*-1 by the International Monetary Fund—or zero otherwise (International Monetary Fund, 2020, 2021).

dimension reduction techniques, regularized regressions, support vector machines, regression trees, neural network, and forecast combinations.

Our findings demonstrate that machine learning methods can successfully predict returns in country equity indices. As at the stock level, nonlinearities and interactions play an essential role. In consequence, we find that neural networks produce highly accurate forecasts—outperforming the simply dimension reduction techniques or penalized regressions. Furthermore, a particularly effective method is forecast combination. This approach, suppressing individual model variance, produces return forecasts of superior accuracy. Even though none of the machine learning methods is perfect, they work very well when combined.

Importantly, when assessing the relative performance of different machine learning models, we supplement the traditional measures of global fit with the cross-sectional R^2. This metric assists in gauging how effectively a given technique may sort assets into portfolios. We find that even the simplest methods—with a seemingly low global fit—can still produce a decent cross-sectional fit. In consequence, they may still prove useful in practice despite the low or negative standard R^2 values.

A glimpse inside the black box of machine learning methods allows for determining principal drivers of the cross-section of market returns. Despite the growing factor zoo in asset pricing literature, a sparse set of variables can capture the variation in country equity returns. Nearly all models point to several simple predictors that really matter; these include long-term reversal, market value, earnings yield, market breadth, and long-term momentum. Numerous other seemingly relevant signals—such as credit, liquidity, or idiosyncratic risks—are of secondary importance.

All the machine learning techniques we consider can be forged into effective market allocation strategies. Portfolios from one-way sorts on the model predictions exhibit both economically and statistically significant abnormal returns; these cannot be explained by popular asset pricing factors. Interestingly, most alphas come from the long legs rather than short legs of the trading strategies, reducing the concerns of short selling limitations. Moreover, contrary to empirical findings from individual stocks, even the simple OLS method produces substantial alphas. The best performing strategy is neural networks. An equal-weighted quintile of top markets according to the FFN2 method outperforms their low-ranked counterparts by 1.69% per month. The associated long-short strategy displays a Sharpe ratio of 1.57.

An exploration of sources of return predictability links it with behavioral mispricing. The predictability is not affected by the swings in country-specific risks. On the other hand, it is measurably affected by the level of mispricing—being the most pronounced in both the overvalued and undervalued markets. Furthermore, in line with the mispricing narrative, it prevails in market segments with higher limits to arbitrage; these include smaller, riskier, and less liquid countries. Finally, similarly as for numerous stock-level anomalies, the return predictability of country index returns diminishes over time. In consequence, although it has not disappeared entirely, it was visibly weaker over the last 10 to 15 years than it was a decade before.

Future research on the topics in this paper could be extended to other asset classes. Machine learning methods have proven effective for the cross-section of equities, corporate bonds, and stock market indices. Do they work for international treasuries? Or currencies? Can they be applied across these asset classes?[12] These questions remain to be answered.

[12] An important research question is whether machine learning methods can be applied to international sovereign bonds as well as the role of cross-asset signals between stock and bond markets. Cakici and Zaremba (2022) pursue this line of research.

References

Albuquerque, R., Eichenbaum, M., Papanikolaou, D., & Rebelo, S. (2015). Long-run bulls and bears. *Journal of Monetary Economics*, 76, S21-S36.

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *Journal of Finance*, *68*(3), 929-985.

Asness, C.S., Liew, J.M., & Stevens, R.L. (1997). Parallels between the cross-sectional predictability of stock and country returns. *Journal of Portfolio Management*, 23(3), 79-87.

Avramov, D., Cheng, S., & Metzker, L. (2021). Machine learning versus economic restrictions: Evidence from stock return predictability. *Management Science*, in press.

Avramov, D., Chordia, T., Jostova, G., & Philipov, A. (2012). The world price of credit risk. *Review of Asset Pricing Studies*, 2(2), 112-152.

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

Bali, T. G., & Cakici, N. (2010). World market risk, country-specific risk and expected returns in international stock markets. *Journal of Banking & Finance*, 34 (6), 1152–1165.

Bali, T. G., Subrahmanyam, A., & Wen, Q. (2021). Long-term reversals in the corporate bond market. *Journal of Financial Economics*, 139(2), 656-677.

Bali, T. G., Subrahmanyam, A., & Wen, Q. (2021). The macroeconomic uncertainty premium in the corporate bond market. *Journal of Financial and Quantitative Analysis,* 56(5), 1653-1678.

Bali, T., Goyal, A., Huang, D., Jiang, F., & Wen, Q. (2021). Different strokes: Return predictability across stocks and bonds with machine learning and big data. Georgetown McDonough School of Business Research Paper No. 3686164. Swiss Finance Institute Research Paper No. 20-110. Available at SSRN: https://ssrn.com/abstract=3686164.

Baltussen, G., Swinkels, L., & Van Vliet, P. (2021). Global factor premiums. *Journal of Financial Economics*, *142*(3), 1128-1154.

Balvers, R., Wu, Y., & Gilliland, E. (2000). Mean reversion across national stock markets and parametric contrarian investment strategies. *Journal of Finance*, 55(2), 745-772.

Bartram, S.M., Djuranovik, L., & Garratt, A. (2018). Currency anomalies. 31st Australasian Finance and Banking Conference 2018, Available at SSRN: https://ssrn.com/abstract=3222252 or http://dx.doi.org/10.2139/ssrn.3222252.

Bekaert, G., & Mehl, A. (2019). On the global financial market integration "swoosh" and the trilemma. *Journal of International Money and Finance*, 94, 227-245.

Bhojraj, S., & Swaminathan, B. (2006). Macromomentum: returns predictability in international equity indices. *Journal of Business*, 79(1), 429-451.

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

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

Bollerslev, T., Hood, B., Huss, J., & Pedersen, L. H. (2018). Risk everywhere: Modeling and managing volatility. *Review of Financial Studies*, 31(7), 2729-2773.

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

Brusa, F., Savor, P., & Wilson, M. (2020). One central bank to rule them all. *Review of Finance*, 24(2), 263-304.

Cagan, P. (1956). The monetary dynamics of hyperinflation. M. Friedman (Ed.), *Studies in the Quantity Theory of Money.* University of Chicago Press, Chicago, IL, pp. 25-117.

Cakici, N., & Zaremba, A. (2021). Salience theory and the cross-section of stock returns: International and further evidence. Journal of Financial Economics, in press.

Cakici, N., & Zaremba, A. (2022). Machine learning across asset classes: Return predictability in equity and government bond markets. Working paper.

Calluzzo, P., Moneta, F., & Topaloglu, S. (2019). When anomalies are publicized broadly, do institutions trade accordingly? *Management Science*, *65*(10), 4555-4574.

Chan, K., Hameed, A., & Tong, W. (2000). Profitability of momentum strategies in the international equity markets. *Journal of Financial and Quantitative Analysis*, 35(2), 153-172.

Chen, L., Pelger, M., & Zhu, J. (2020). Deep learning in asset pricing. Available at SSRN: https://ssrn.com/abstract=3350138.

Choi, D., Jiang, W., & Zhang, C. (2021). Alpha go everywhere: Machine learning and international stock returns. Available at SSRN 3489679.

Choi, J., & Kim, Y. (2018). Anomalies and market (dis)integration. *Journal of Monetary Economics*, 100, 16-34.

Chordia, T., Subrahmanyam, A., & Tong, Q. (2014). Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Journal of Accounting and Economics*, *58*(1), 41-58.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance*, *66*(4), 1047-1108.

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

Cortes, G. S., Taylor, B., & Weidenmier, M. D. (2021). Financial factors and the propagation of the Great Depression. *Journal of Financial Economics*, in press.

Cosemans, M., & Frehen, R. (2021). Salience theory and stock prices: Empirical evidence. *Journal of Financial Economics*, *140*(2), 460-483.

Cox, D.R. & Snell, E.J. (1989). *The Analysis of Binary Data*, 2nd ed. London: Chapman and Hall.

Danielsson, J., Valenzuela, M., & Zer, I. (2018). Learning from history: Volatility and financial crises. *Review of Financial Studies*, 31(7), 2774-2805.

Diamonte, R. L., Liew, J. M., & Stevens, R. L. (1996). Political risk in emerging and developed markets. *Financial Analysts Journal*, *52*(3), 71-76.

Diebold, F., and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253-263.

Drobetz, W., & Otto, T. (2021). Empirical asset pricing via machine learning: evidence from the European stock market. *Journal of Asset Management*, *22*(7), 507-538.

Drobetz, W., Haller, R., Jasperneite, C., and Otto, T. (2019). Predictability and the cross section of expected returns: Evidence from the European stock market. *Journal of Asset Management*, *20*(7), 508-533.

Engelberg, J.,R. McLean, D., & Pontiff, J. (2018). Anomalies and news. *Journal of Finance*, 73, 1971–2001.

Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1995). Country risk and global equity selection. *Journal of Portfolio Management*, *21*(2), 74-83.

Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1996). Political risk, economic risk, and financial risk. *Financial Analysts Journal*, *52*(6), 29-46.

Fama, E. F., & French, K. R. (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics*, 123(3), 441-463.

Ferreira, M. A., & Gama, P. M. (2007). Does sovereign debt ratings news spill over to international stock markets? *Journal of Banking & Finance*, 31(10), 3162-3182.

Filippou, I., Rapach, D., Taylor, M. P., & Zhou, G. (2020). Exchange rate prediction with machine learning and a smart carry portfolio. Available at SSRN 3455713.

Fisher, G. S., Shah, R., & Titman, S. (2017). Should you tilt your equity portfolio to smaller countries? *Journal of Portfolio Management*, *44*(1), 127-141.

Frazzini, A., & Pedersen, L.H. (2014). Betting against beta. *Journal of Financial Economics*, 111, 1-25.

French, K.R. (2022). U.S. Research Return Data. Data Library. Retrieved from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies*, *33*(5), 2326-2377.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, *5*, 1189-1232.

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *Annals of Statistics*, *28*(2), 337-407.

Geczy, C., & Samonov, M. (2017). Two centuries of multi-asset momentum (equities, bonds, currencies, commodities, sectors and stocks). Available at SSRN 2607730.

Gromb, D., & Vayanos, D. (2010). Limits of arbitrage. *Annual Review of Financial* Economcs, *2*(1), 251-275.

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

Gu, S., Kelly, B., & Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics*, *222*(1), 429-450.

Guo, L., Li, F. W., & Wei, K. J. (2020). Security analysts and capital market anomalies. *Journal of Financial Economics*, 137(1), 204-230.

Han, X. (2020). Risks versus mispricing: Decomposing asset pricing anomalies via classification. Available at SSRN: https://ssrn.com/abstract=3604970 or http://dx.doi.org/10.2139/ssrn.3604970.

Han, Y., He, A., Rapach, D., & Zhou, G. (2021). Expected stock returns and firm characteristics: E-LASSO, assessment, and implications. Available at SSRN: https://ssrn.com/abstract=3185335 or http://dx.doi.org/10.2139/ssrn.3185335.

Hjalmarsson, E. (2010). Predicting global stock returns. *Journal of Financial and Quantitative Analysis*, 45(1), 49-80.

Hung, M., Li, X., & Wang, S. (2015). Post-earnings-announcement drift in global markets: Evidence from an information shock. *Review of Financial Studies*, *28*(4), 1242-1283.

Ilmanen, A., Israel, R., Lee, R., Moskowitz, T. J., & Thapar, A. (2021). How do factor premia vary over time? A century of evidence. *Journal Of Investment Management*, *19*(4), 15-57.

International Monetary Fund (2020). Country Composition of WEO Groups. World Economic and Financial Surveys. World Economic Outlook. Available at https://www.imf.org/external/pubs/ft/weo/2020/01/weodata/groups.htm.

International Monetary Fund (2021). Changes to the Database. World Economic Outlook Database.. Available at https://www.imf.org/external/pubs/ft/weo/data/changes.htm.

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

Jensen, T. I., Kelly, B. T., & Pedersen, L. H. (2021). *Is there a replication crisis in finance?* NBER Working Paper No. w28432. National Bureau of Economic Research. Available at https://www.nber.org/papers/w28432.

Jiang, F., Tang, G., & Zhou, G. (2018). Firm characteristics and Chinese stocks. *Journal of Management Science and Engineering*, *3*(4), 259-283.

Jiang, L., Liu, J., Peng, L., & Wang, B. (2021). Investor attention and asset pricing anomalies. *Review of Finance*, in press.

Kelly, B. T., & Malamud, S. (2021). The virtue of complexity in machine learning portfolios. Swiss Finance Institute Research Paper No. 21-90. Available at SSRN: https://ssrn.com/abstract=3959708 or http://dx.doi.org/10.2139/ssrn.3959708.

Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3), 501-524.

Keloharju, M., Linnainmaa, J. T., & Nyberg, P. (2021). Are return seasonalities due to risk or mispricing? *Journal of Financial Economics*, *139*(1), 138-161.

Keloharju, M., Linnainmaa, J.T., & Nyberg, P. (2016). Return seasonalities. *Journal of Finance*, *71*(4), 1557-1589.

Koijen, R. S., Moskowitz, T. J., Pedersen, L. H., & Vrugt, E. B. (2018). Carry. *Journal of Financial Economics*, *127*(2), 197-225.

Lam, F. E. C., Li, Y., Prombutr, W., & Wei, K. J. (2020). Limits-to-arbitrage, investment frictions, and the investment effect: New evidence. *European Financial Management*, *26*(1), 3-43.

Leippold, M., Wang, Q., & Zhou, W. (2021). Machine learning in the Chinese stock market. *Journal of Financial Economics*, in press.

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

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

Miranda-Agrippino, S., & Rey, H. (2020). US monetary policy and the global financial cycle. *Review of Economic Studies*, *87*(6), 2754-2776.

Muir, T. (2017). Financial crises and risk premia. *Quarterly Journal of Economics*, 132(2), 765-809.

Müller, S., & Preissler, F. (2021). In good and in bad times? The relation between anomaly returns and market states. Available at SSRN: https://ssrn.com/abstract=3926059 or http://dx.doi.org/10.2139/ssrn.3926059.

Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, *28*(3), 777-787.

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

Novy-Marx, R., & Velikov, M. (2019). Comparing cost-mitigation techniques. *Financial Analysts Journal*, *75*(1), 85-102.

Pitkäjärvi, A., Suominen, M., & Vaittinen, L. (2020). Cross-asset signals and time series momentum. *Journal of Financial Economics*, *136*(1), 63-85.

Pontiff, J. (1996). Costly arbitrage: Evidence from closed-end funds. *Quarterly Journal of Economics*, 111(4), 1135-1151.

Pontiff, J., & Woodgate, A. (2008). Share issuance and cross-sectional returns. *Journal of Finance*, 63(2), 921-945.

Rad, H., Low, R. K. Y., Miffre, J., & Faff, R. W. (2021). The commodity risk premium and neural networks. Available at SSRN 3816170.

Rapach, D. E., Strauss, J. K., Tu, J., & Zhou, G. (2019). Industry return predictability: A machine learning approach. *Journal of Financial Data Science*, *1*(3), 9-28.

Rapach, D.E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, *23*(2), 821-862.

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

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. Journal of Finance, *52*(1), 35-55.

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

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

Struck, C., & Cheng, E. (2020). The cross section of commodity returns: A nonparametric approach. *Journal of Financial Data Science*, *2*(3), 86-103.

Thomson Reuters. (2008). Datastream Global Equity Indices: User Guide, Issue 5. Thomson Reuters Ltd.

Tobek, O., & Hronec, M. (2021). Does it pay to follow anomalies research? Machine learning approach with international evidence. *Journal of Financial Markets*, *56*, 100588.

Watanabe, A., Xu, Y., Yao, T., & Yu, T. (2013). The asset growth effect: Insights from international equity markets. *Journal of Financial Economics*, *108*(2), 529-563.

Zaremba, A., Szyszka, A., Karathanasopoulos, A., & Mikutowski, M. (2021). Herding for profits: Market breadth and the cross-section of global equity returns. *Economic Modelling*, *97*, 348-364.

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

Zhang, C. Y., & Jacobsen, B. (2013). Are monthly seasonals real? A three century perspective. *Review of Finance*, 17(5), 1743-1785.

Zhang, C. Y., & Jacobsen, B. (2021). The Halloween indicator, "Sell in May and Go Away": Everywhere and all the time. *Journal of International Money and Finance*, 110, 102268.

Figure 1. Research Sample Through Time

The figure exhibits the evolution of the research sample through time—the monthly number of markets covered and aggregate stock market capitalization in U.S. dollars.

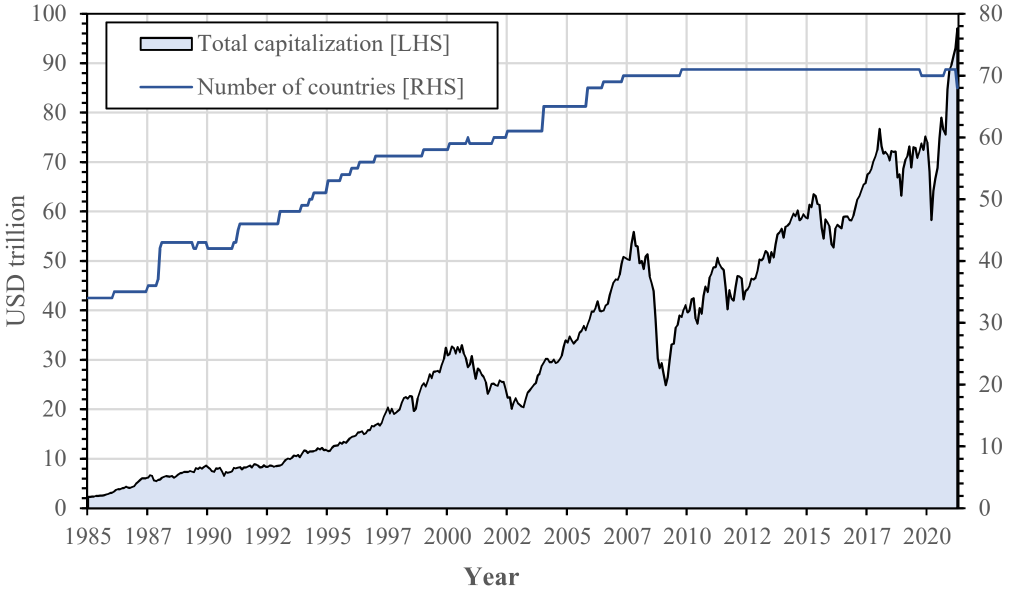


Figure 2. Characteristic Importance

The figure displays the rankings of 88 market characteristics employed in the study in terms of their overall model contribution. The color gradients indicate the rank of the variable importance; the dark blue (white) represent the most influential (least influential) predictors. The variables are ordered based on their average rank across all models. The sample comprises 71 country stock markets and the testing period is from January 1995 to April 2021.

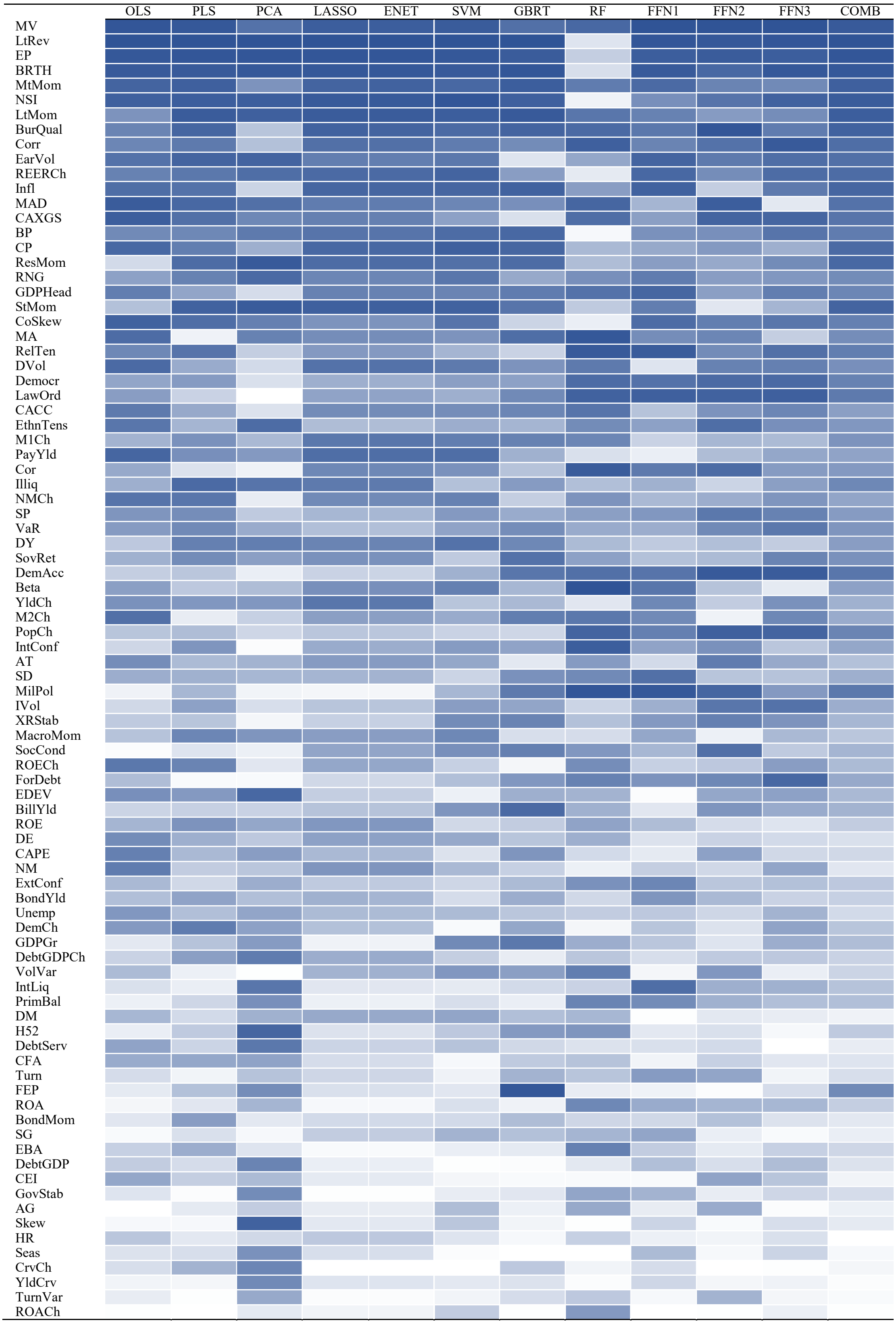


Figure 3. Relative Importance of Top Variables in Different Models

The figure presents the importance of the top 10 variables in the machine learning models examined in this study. The panels display the reduction in R^2 from setting all values of a given variable to zero in the training sample. The numbers are averaged across all the training samples and are rescaled to sum to 1. The sample comprises 71 country stock markets and the testing period is from January 1995 to April 2021.

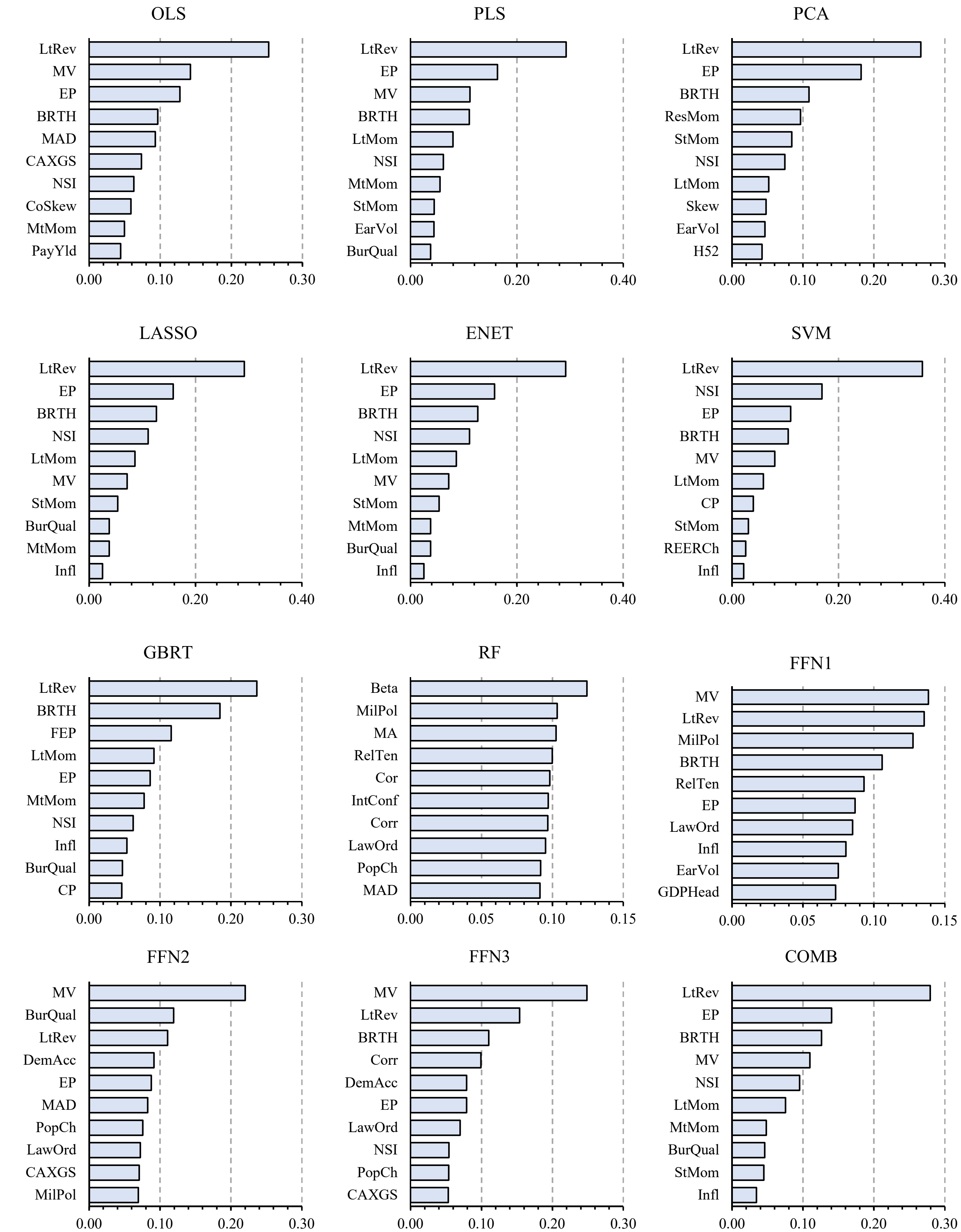


Figure 4. Characteristic Importance per Category

The figure displays the importance of 15 different categories of market characteristics, as classified in Table 1, in terms of their overall model contribution. The color gradients indicate the aggregate importance weight of individual characteristics summed within the categories. The dark blue (white) colors represent the most influential (least influential) groups. The variables are ordered based on their average rank across all the models. The sample comprises 71 country stock markets and the testing period is from January 1995 to April 2021.

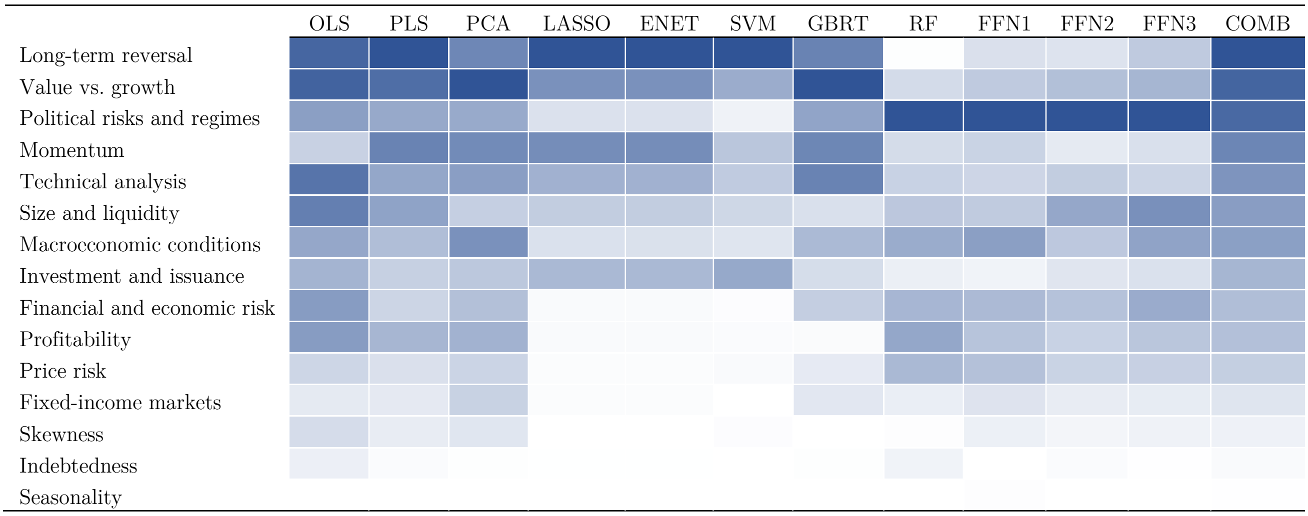
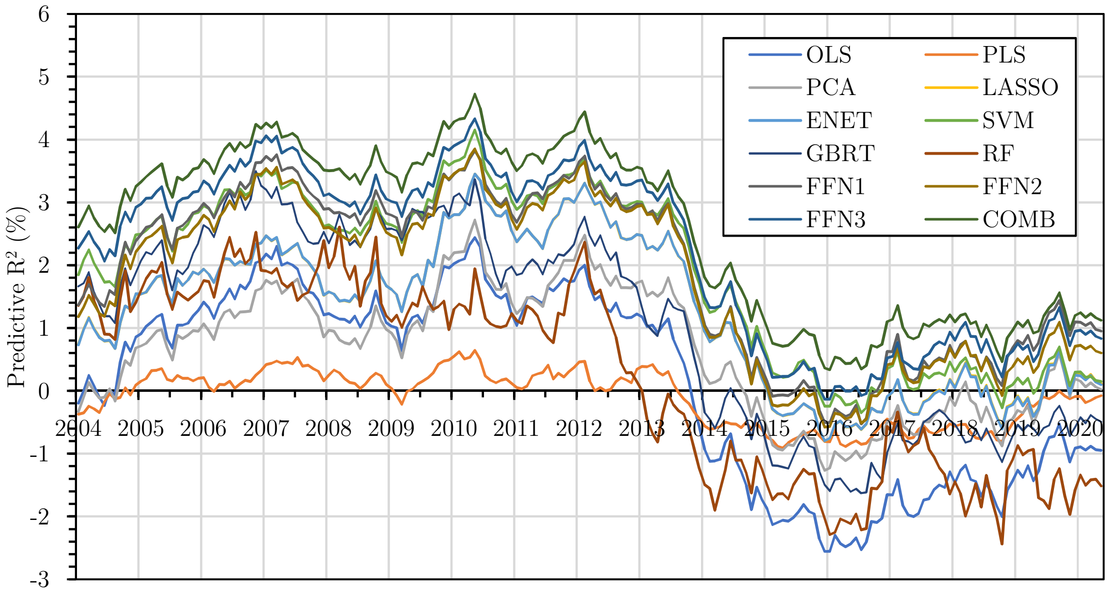


Figure 5. Predictive R^2 Coefficients from Machine Learning Models Through Time

The figure presents the predictive R^2 coefficients from different machine learning models through time. Each month, using our test samples at re-estimation dates, we run cross-sectional regressions of the realized excess returns on the respective predictions of different machine learning models. Panel A concerns the Predictive R^2; i.e., the out-of-sample adjusted R^2 (R^2_{OOS}) coefficient from the models. Panel B focuses on the Rank R^2 measure, which is obtained in a two-step procedure: first, we transform the predicted and realized returns into ranks and then map them into the [0,1] interval; second, we calculate the pseudo R^2 measure of Cox and Snell (1989). The exhibit below plots trailing 120-month averages of these estimates expressed in percentage terms. The sample comprises 71 country stock markets and the testing period is from January 1995 to April 2021.

Panel A: Predictive R^2



Panel B: Rank R^2

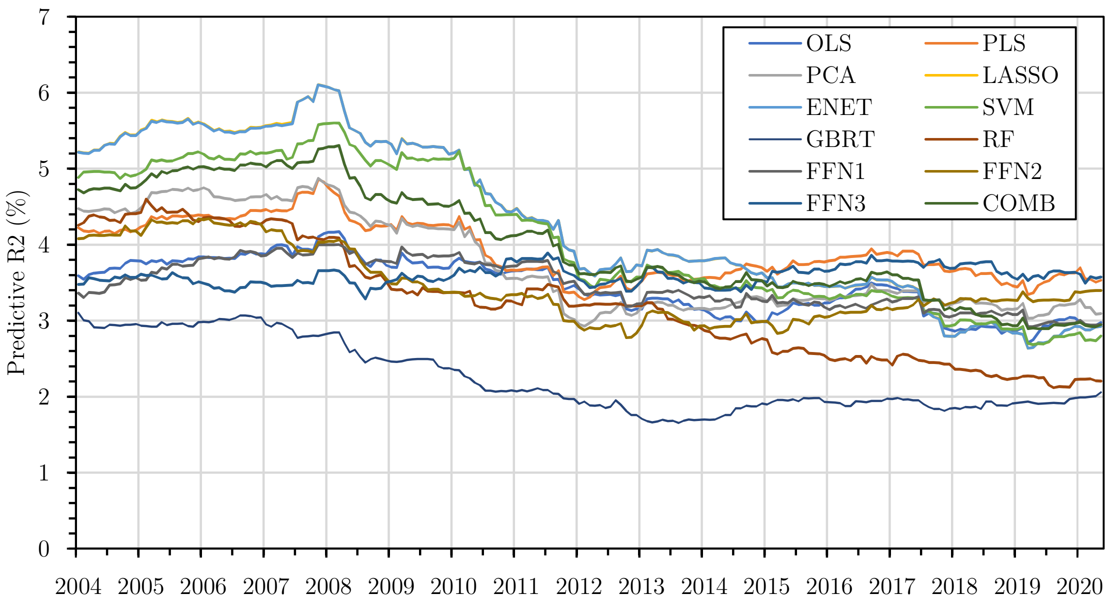


Figure 6. Cumulative Returns on Machine Learning Portfolios

The figure presents cumulative returns on long-short portfolios formed on the machine learning methods forecasts. The portfolios buy (sell) the quantile of markets with the highest (lowest) return predictions by the machine learning models. The portfolios are rebalanced monthly and equal-weighted. The returns are cumulated additively and are expressed in percentage terms. The reported period is from December 1994 to April 2021, and the sample comprises 71 country stock markets.

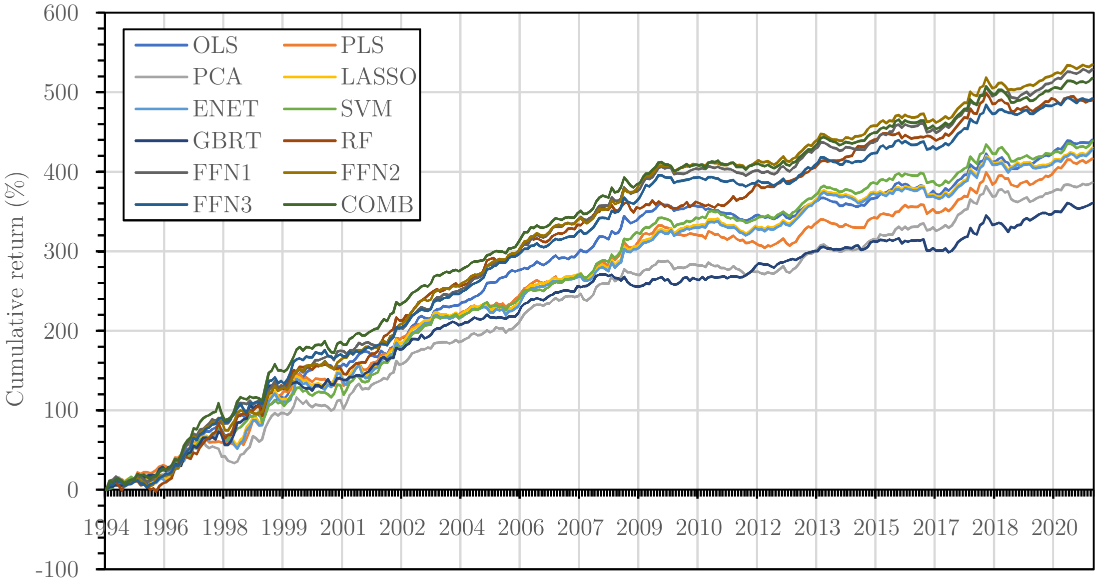


Table 1. Market Characteristics

The table summarizes the 88 market characteristics used in this study along with their *Symbols* used throughout the paper. We also report the variable averages and standard deviation. All the data in this table are winsorized at the 99% level. The sample comprises 71 country stock markets and the study period is from January 1985 to April 2021. Further details concerning the computations of the variables, along with relevant literature references, are provided in Table A1 in the Online Appendix.

Symbol	Variable	Average	Standard deviation
Panel A: Value vs. growth			
EP	Earnings yield	0.076	0.046
BM	Book-to-market ratio	0.652	0.313
CP	Cash flow-to-price ratio	0.162	0.087
SP	Sales-to-price ratio	0.667	0.415
EDEV	EBITDA-to-EV ratio	0.129	0.074
FEP	Forward earnings yield	0.089	0.042
DY	Dividend yield	3.082	1.716
CAPE	Cyclically adjusted P/E ratio	19.506	8.707
Panel B: Size and liquidity			
MV	Market value	11.022	2.191
Illiq	Amihud ratio	0.001	0.005
Turn	Turnover ratio	0.538	0.598
Dvol	Dollar volume	16.903	3.060
TurnVar	Turnover volatility	0.001	0.002
VolVar	Volume volatility	10.802	2.526
Panel C: Price risk			
Beta	Beta	0.899	0.502
Cor	Correlation	0.556	0.268
Vol	Total volatility	0.076	0.042
IVol	Idiosyncratic volatility	0.058	0.038
RNG	Price range	0.412	0.241
VAR	Value at risk	0.111	0.060
Panel D: Momentum			
LtMom	Long-term momentum	0.007	0.029
MtMom	Medium-term momentum	0.008	0.037
StMom	Short-term momentum	0.008	0.075
ResMom	Residual momentum	-0.036	0.287
Panel E: Seasonality			
Seas	Cross-sectional seasonality	0.010	0.027
Panel F: Profitability			
ROA	Return on asset	0.024	0.019
ROE	Return on equity	0.119	0.061
CFA	Cash profitability	0.041	0.027
EBA	EBIT-to-asset	0.044	0.030
NM	Net margin	0.089	0.063
SG	Sales growth	0.162	0.244
ROACh	ROA change	-0.001	0.013
ROECh	ROE change	-0.002	0.050
NMCh	Net margin change	-0.001	0.038
EarVol	Earnings volatility	0.008	0.010
AT	Asset turnover	0.317	0.206
Panel G: Indebtedness			
DE	Debt-to-equity ratio	0.931	1.002
DM	Debt-to-capitalization ratio	0.391	0.477
Panel H: Skewness			
SKEW	Total skewness	-0.396	1.134
COSKEW	Co-skewness	-0.377	7.163
Panel I: Long-term reversal			
LrRev	Long-run reversal	0.008	0.015
Panel J: Technical analysis			
MA	Moving average	1.044	0.165
Panel J cont.			
H52	52-week high effect	0.873	0.140
MAD	Moving average distance	1.031	0.132
BRTH	Market breadth	0.001	0.129
Panel K: Investment and issuance			
AG	Asset growth	0.183	0.266
CEI	Composite equity issuance	0.075	0.246
NSI	Net share issuance	0.017	0.224
HR	Hiring rate	0.092	0.191
PY	Payout ratio	0.431	0.196
Panel L: Macroeconomic conditions			
Unemp	Unemployment rate	0.075	0.045
Infl	Inflation rate	0.073	0.170
GDPGr	GDP growth	0.114	0.172
REERCh	REER change	0.000	0.003
DebtGDP	Debt-to-GDP ratio	0.539	0.329
PrimBal	Primary balance	0.009	0.034
M1Ch	M1 change	0.152	0.217
M2Ch	M2 change	0.153	0.314
PopCh	Population	0.113	0.128
DebtGDPCh	Debt-to-GDP ratio change	0.445	5.900
MacroMom	Macro momentum	0.000	0.483
Panel M: Fixed-income markets			
BillYld	Treasury bill yield	0.072	0.104
BondYld	Government bond yield	0.059	0.039
YldCrv	Yield curve slope	0.012	0.019
CrvCh	Yield curve change	0.000	0.016
YldCh	Yield change	-0.326	1.310
BondMom	Bond momentum	0.007	0.008
Panel N: Financial and economic risk			
ForDebt	Foreign debt (% GDP)	6.606	2.150
XRStab	Exchange rate stability	8.697	1.715
DebtServ	Foreign debt serv. (%export)	8.394	1.537
CAXGS	Current account (% exports)	11.964	1.264
IntLiq	Net international liquidity	2.302	1.398
GDPHead	GDP per head	3.102	1.444
CACC	Current account (% GDP)	10.832	2.422
SovRet	Sovereign risk	6.363	4.693
Panel P: Political risks and regimes			
GovStab	Government stability	7.831	1.720
SocCond	Socioeconomic conditions	7.117	1.968
IntConf	Internal conflict	9.640	1.964
ExtConf	External conflict	10.283	1.439
Corr	Corruption	3.643	1.322
MilPol	Military in politics	4.657	1.478
RelTen	Religious tensions	4.784	1.299
LawOrd	Law and order	4.438	1.319
EthnTens	Ethnic tensions	4.289	1.329
DemAcc	Democratic accountability	4.745	1.429
BurQual	Bureaucracy quality	2.979	0.903
Dem	Democracy index	0.589	0.267
DemCh	Democratization	0.001	0.016

Table 2. Predictive Performance of the Machine Learning Models

The table reports the predictive performance measures of the machine learning models that are examined in this study (see Section 2.3 for details). Panel A concerns the R^2 coefficients and slopes. Each month, using our test samples at re-estimation dates, we run cross-sectional regressions of the realized excess returns on the respective predictions of different machine learning models. *Predictive slope* indicates the time-series average of the slopes estimated in the monthly regressions, and *Predictive R^2* is the out-of-sample adjusted R^2 (R^2_{OOS}) coefficient. *Rank R^2 (%)* is obtained in a two-step procedure: first, we transform the predicted and realized returns into ranks and map them into the [0,1] interval; second, we calculate the pseudo R^2 measure of Cox and Snell (1989). Panel B presents the pairwise comparisons of the machine learning models using modified Diebold and Mariano (1995) tests (DM). The test statistic DM compares the mean squared forecast errors of a model in column a and row b: $DM_{a,b} = \frac{\bar{d}_{a,b}}{\hat{\sigma}_{\bar{d}_{a,b}}}$, where $d_{a,b,t+1} = MSFE^{(a)}_{t+1} - MSFE^{(b)}_{t+1}$ denotes the differences in the monthly mean squared forecast errors, $\bar{d}_{a,b} = \frac{d_{a,b,t+1}}{T}$ indicates the time-series average of these differences, and $\hat{\sigma}_{\bar{d}_{a,b}}$ is the standard error adjusted for heteroskedasticity and autocorrelation using the HAC estimator (Newey & West, 1987). The bold font denotes the values significant at the 5% level in standalone pairwise comparisons (|*t*-stat|>1.96), and the underline indicates the 5%-significance incorporating the Bonferroni adjustment for the multiple hypothesis framework ((|*t*-stat|>2.87). The sample comprises 71 countries and the testing period is from January 1995 to April 2021.

Panel A: Predictive R^2 and slopes

	OLS	PLS	PCA	LASSO	ENET	SVM	GBRT	RF	FFN1	FFN2	FFN3	COMB
Predictive R^2 (%)	-0.14	-0.23	0.28	0.91	0.90	1.47	0.95	-0.28	1.52	1.29	1.89	2.21
Predictive slopes	0.49	0.47	0.52	1.24	1.25	0.85	0.36	0.48	0.66	0.70	0.60	0.83
Rank R^2 (%)	3.33	3.78	3.68	4.03	4.03	3.84	2.87	3.90	3.56	3.44	3.61	3.95

Panel B: Diebold-Mariano (1995) tests

	PLS	PCA	LASSO	ENET	GBRT	SVM	RF	FFN1	FFN2	FFN3	COMB
OLS	-0.15	-0.36	0.97	0.95	**2.84**	1.40	0.89	$\underline{5.54}$	$\underline{5.90}$	$\underline{5.99}$	$\underline{5.16}$
PLS		-0.02	0.51	0.50	1.33	1.53	1.15	1.74	1.40	1.85	$\underline{2.93}$
PCA			**2.45**	**2.42**	**2.59**	1.22	0.87	**2.83**	$\underline{3.08}$	$\underline{4.24}$	$\underline{3.70}$
LASSO				-1.53	1.54	0.78	0.61	1.83	1.72	$\underline{3.00}$	$\underline{2.89}$
ENET					1.58	0.79	0.62	1.86	1.74	$\underline{3.04}$	$\underline{2.92}$
GBRT						0.36	0.38	1.30	0.60	**2.44**	$\underline{3.15}$
SVM							0.30	0.15	-0.16	0.27	1.09
RF								-0.16	-0.29	-0.09	0.16
FFN1									-1.41	0.34	1.62
FFN2										**2.01**	**2.49**
FFN3											1.37

Table 3. Portfolios from Univariate Sorts on Machine Learning Model Predictions

The table presents the monthly returns on quintile portfolios from univariate sorts on the predictions of different machine learning models from Section 2.3. *Low* (*High*) denotes the quintiles of markets with the lowest (highest) predicted return. *H-L* is the spread portfolio that assumes a long (short) position in the *High* (*Low*) quintiles. The portfolios are equal- or value-weighted (Panels A and B, respectively), and are reformed on a monthly basis. *Pred* and *Avg* indicate the average predicted and realized returns, respectively. *SD* is the standard deviation of monthly returns, *SR* is the annualized Sharpe ratio, and α is the average abnormal return from the global CAPM. *R*, *Avg*, *SD*, and α are expressed in percentages. The numbers in parentheses are t-statistics calculated using the HAC estimator (Newey & West, 1987). The sample comprises 71 country stock markets and the testing period is from January 1995 to April 2021.

	Panel A: Equal-weighted portfolios					Panel B: Value-weighted portfolios				
					OLS					
	Pred	Avg	SD	SR	α	Pred	Avg	SD	SR	α
Low (L)	-0.51	0.12	5.27	0.08	-0.47	-0.37	0.28	5.22	0.19	-0.34
2	0.43	0.48	5.07	0.33	-0.14	0.43	0.50	5.03	0.34	-0.13
3	0.98	0.69	4.91	0.49	0.09	0.98	0.62	5.16	0.42	-0.03
4	1.58	0.86	4.91	0.61	0.29	1.56	0.73	5.24	0.49	0.10
High (H)	2.96	1.51	5.27	0.99	0.97	2.68	1.15	6.57	0.60	0.41
H-L	3.47	1.39	3.60	1.34	1.44	3.06	0.87	4.25	0.71	0.75
t-stat		*(6.31)*			*(6.44)*		*(3.39)*			*(3.09)*
					PLS					
	Pred	Avg	SD	SR	α	Pred	Avg	SD	SR	α
Low (L)	-1.19	0.24	5.29	0.16	-0.37	-1.07	0.48	5.28	0.32	-0.17
2	-0.36	0.67	4.82	0.48	0.07	-0.36	0.71	4.81	0.51	0.10
3	0.19	0.40	4.71	0.30	-0.16	0.17	0.43	5.16	0.29	-0.22
4	0.79	0.79	5.14	0.53	0.20	0.76	0.58	5.69	0.36	-0.09
High (H)	2.11	1.56	5.51	0.98	1.01	1.85	1.15	6.92	0.58	0.40
H-L	3.30	1.32	3.99	1.14	1.38	2.92	0.67	4.68	0.50	0.56
t-stat		*(5.60)*			*(5.83)*		*(2.46)*			*(2.08)*
					PCA					
	Pred	Avg	SD	SR	α	Pred	Avg	SD	SR	α
Low (L)	-0.10	0.30	5.28	0.20	-0.31	0.01	0.45	5.17	0.33	-0.19
2	0.62	0.54	4.90	0.38	-0.05	0.62	0.71	4.94	0.49	0.08
3	1.08	0.50	4.72	0.37	-0.06	1.07	0.44	4.98	0.31	-0.17
4	1.62	0.79	5.23	0.52	0.19	1.60	0.50	5.70	0.31	-0.18
High (H)	2.72	1.53	5.44	0.97	0.98	2.52	1.14	6.77	0.60	0.41
H-L	2.82	1.22	3.95	1.07	1.30	2.51	0.69	4.65	0.51	0.60
t-stat		*(5.24)*			*(5.48)*		*(2.59)*			*(2.28)*
					LASSO					
	Pred	Avg	SD	SR	α	Pred	Avg	SD	SR	α
Low (L)	0.50	0.18	5.29	0.12	-0.43	0.57	0.33	5.27	0.25	-0.30
2	0.90	0.65	4.88	0.46	0.05	0.90	0.69	5.01	0.47	0.05
3	1.12	0.57	4.90	0.41	-0.02	1.12	0.59	5.03	0.41	-0.04
4	1.39	0.72	5.16	0.48	0.13	1.38	0.57	5.78	0.32	-0.11
High (H)	2.04	1.54	5.45	0.98	1.01	1.87	1.19	6.66	0.62	0.51
H-L	1.54	1.36	4.24	1.11	1.44	1.29	0.86	5.01	0.56	0.81
t-stat		*(5.53)*			*(5.77)*		*(2.89)*			*(2.79)*
					ENET					
	Pred	Avg	SD	SR	α	Pred	Avg	SD	SR	α
Low (L)	0.50	0.19	5.28	0.12	-0.42	0.57	0.34	5.23	0.26	-0.30
2	0.90	0.64	4.89	0.45	0.04	0.90	0.69	5.02	0.46	0.04
3	1.12	0.57	4.89	0.40	-0.02	1.12	0.59	5.02	0.41	-0.05

4	1.39	0.72	5.17	0.48	0.13	1.38	0.59	5.76	0.33	-0.09
High (H)	2.04	1.54	5.45	0.98	1.01	1.87	1.19	6.67	0.62	0.51
H-L	1.54	1.35	4.24	1.11	1.43	1.29	0.86	5.00	0.56	0.81
t-stat		*(5.50)*			*(5.74)*		*(2.86)*			*(2.76)*

SVM

	Pred	Avg	SD	SR	α	Pred	Avg	SD	SR	α
Low (L)	0.18	0.11	5.27	0.07	-0.49	0.27	0.30	5.08	0.21	-0.31
2	0.67	0.55	5.09	0.38	-0.07	0.66	0.62	4.96	0.43	-0.01
3	0.95	0.74	4.90	0.52	0.14	0.94	0.73	5.25	0.48	0.07
4	1.29	0.76	5.17	0.51	0.18	1.27	0.65	5.70	0.40	-0.04
High (H)	2.07	1.50	5.23	0.99	0.98	1.88	1.23	6.46	0.66	0.54
H-L	1.89	1.39	3.98	1.21	1.47	1.61	0.93	4.40	0.73	0.85
t-stat		*(6.45)*			*(6.60)*		*(3.97)*			*(3.75)*

GBRT

	Pred	Avg	SD	SR	α	Pred	Avg	SD	SR	α
Low (L)	-0.87	0.22	5.24	0.14	-0.38	-0.81	0.39	5.33	0.25	-0.25
2	0.16	0.43	4.83	0.31	-0.14	0.17	0.44	5.00	0.30	-0.19
3	0.83	0.66	4.90	0.46	0.08	0.83	0.51	5.25	0.34	-0.13
4	1.51	1.00	4.98	0.69	0.40	1.50	0.71	5.40	0.46	0.05
High (H)	2.71	1.36	5.43	0.86	0.77	2.57	0.99	5.94	0.58	0.29
H-L	3.57	1.14	3.74	1.06	1.15	3.37	0.60	3.77	0.55	0.54
t-stat		*(5.57)*			*(5.68)*		*(2.65)*			*(2.38)*

RF

	Pred	Avg	SD	SR	α	Pred	Avg	SD	SR	α
Low (L)	-0.98	0.06	5.40	0.04	-0.55	-0.83	0.18	5.77	0.11	-0.48
2	0.14	0.40	5.04	0.27	-0.19	0.15	0.47	5.20	0.31	-0.17
3	0.76	0.67	4.82	0.48	0.08	0.76	0.54	5.13	0.36	-0.11
4	1.39	0.91	4.95	0.64	0.34	1.37	0.79	5.17	0.53	0.15
High (H)	2.82	1.61	5.45	1.02	1.05	2.58	1.15	5.71	0.70	0.51
H-L	3.79	1.55	4.13	1.30	1.61	3.41	0.97	4.47	0.75	1.00
t-stat		*(6.50)*			*(6.79)*		*(3.59)*			*(3.77)*

FFN1

	Pred	Avg	SD	SR	α	Pred	Avg	SD	SR	α
Low (L)	-0.33	0.01	5.21	0.01	-0.57	-0.18	0.29	5.22	0.20	-0.32
2	0.48	0.50	5.02	0.34	-0.12	0.48	0.56	4.99	0.39	-0.07
3	0.89	0.67	4.99	0.46	0.06	0.87	0.62	5.13	0.42	-0.03
4	1.36	0.79	4.97	0.55	0.21	1.33	0.73	5.48	0.46	0.07
High (H)	2.59	1.68	5.25	1.11	1.15	2.38	1.33	6.76	0.68	0.58
H-L	2.92	1.67	3.73	1.55	1.72	2.56	1.04	4.51	0.80	0.90
t-stat		*(7.74)*			*(7.88)*		*(3.89)*			*(3.51)*

FFN2

	Pred	Avg	SD	SR	α	Pred	Avg	SD	SR	α
Low (L)	-0.22	0.04	5.26	0.03	-0.56	-0.12	0.31	5.12	0.21	-0.30
2	0.50	0.59	4.92	0.41	-0.01	0.50	0.65	4.91	0.46	0.03
3	0.93	0.71	4.93	0.50	0.11	0.92	0.56	5.17	0.38	-0.08
4	1.42	0.59	4.95	0.41	0.01	1.41	0.56	5.41	0.36	-0.10
High (H)	2.61	1.74	5.36	1.12	1.20	2.39	1.44	6.76	0.74	0.69
H-L	2.83	1.69	3.73	1.57	1.75	2.51	1.13	4.54	0.86	1.00
t-stat		*(7.57)*			*(7.95)*		*(4.47)*			*(4.14)*

FFN3

	Pred	Avg	SD	SR	α	Pred	Avg	SD	SR	α
Low (L)	0.04	0.04	5.18	0.03	-0.54	0.10	0.33	5.28	0.22	-0.28
2	0.58	0.53	5.01	0.37	-0.08	0.58	0.56	4.97	0.39	-0.07
3	0.92	0.74	4.85	0.53	0.16	0.91	0.67	5.20	0.44	0.02

4	1.33	0.73	4.77	0.53	0.17	1.31	0.56	5.33	0.36	-0.08
High (H)	2.35	1.60	5.57	0.99	1.03	2.19	1.25	7.00	0.62	0.48
H-L	2.31	1.56	3.72	1.45	1.57	2.10	0.92	4.69	0.68	0.77
t-stat		*(7.07)*			*(7.12)*		*(3.51)*			*(3.08)*
					COMB					
	Pred	Avg	SD	SR	α	Pred	Avg	SD	SR	α
Low (L)	-0.02	0.00	5.36	0.00	-0.61	-0.17	0.25	5.36	0.16	-0.38
2	0.54	0.59	5.02	0.41	-0.02	0.46	0.64	4.99	0.44	0.00
3	0.88	0.67	4.92	0.47	0.07	0.88	0.58	5.13	0.39	-0.06
4	1.27	0.76	4.84	0.54	0.20	1.35	0.75	5.41	0.48	0.09
High (H)	2.23	1.64	5.39	1.05	1.10	2.25	1.23	6.83	0.63	0.49
H-L	2.25	1.64	4.13	1.37	1.71	2.42	0.98	5.06	0.67	0.87
t-stat		*(6.75)*			*(7.07)*		*(3.26)*			*(3.04)*

Table 4. Practical Properties of the Machine Learning Portfolios

The table presents the drawdown, turnover statistics, and multifactor alphas of quintile portfolios formed on using the machine learning models. The equal- and value-weighted strategies (Panels A and B, respectively) are based on predictions from different machine learning models that are indicated in the first row (see Section 2.3 for details). Panels A.1 and B.1 concern portfolio turnover. The long-only (*LO*) portfolios buy the quintile of markets with the highest predictions; the long-short strategies (*LS*) additionally sell the quintile of markets with the lowest predictions. The portfolios are rebalanced monthly. *Turnover* denotes the average monthly one-sided turnover, calculated following Koijen et al. (2018), as a share of portfolio that needs to be replaced each month. *Breakeven* indicates the associated breakeven transaction costs. Panes A.2 and B.2 display the maximum monthly loss on both the long-short strategy (*Max 1M loss*) and maximum drawdown (*Max DD*). Panels A.3 and B.3 report the results of spanning tests of the long-short machine learning strategies with the six-factor model of Fama and French (2018). α_{FF6} denotes the monthly alpha and R^2_{FF6} is the adjusted coefficient of determination. The subscript "S" denotes global stock-level factors from French (2022). The subscript "C" indicates the factors formed of country indices that mimic the portfolio structure as the evaluated strategies (equal- or value-weighted quintiles) and are derived from identical asset universe. All values are reported in percentages. The numbers in parentheses are Newey-West (1987) adjusted *t*-statistics. The sample comprises 71 country stock markets and the testing period is from January 1995 to April 2021.

Panel A: Equal-weighted portfolios

	OLS	PLS	PCA	LASSO	ENET	SVM	GBRT	RF	FFN1	FFN2	FFN3	COMB
						Panel A.1: Loss statistics						
Max 1M loss (%)	12.19	10.90	10.63	11.80	11.80	13.35	13.63	12.87	10.63	11.18	9.82	13.01
Max DD (%)	26.73	25.45	25.00	27.18	27.18	25.44	30.72	25.85	24.37	24.81	25.43	27.93
						Panel A.2: Spanning tests						
α_{FF6-S}	1.37	1.35	1.19	1.21	1.21	1.24	0.96	1.42	1.62	1.73	1.60	1.62
	(4.89)	*(4.51)*	*(4.52)*	*(4.14)*	*(4.11)*	*(4.79)*	*(3.95)*	*(5.79)*	*(6.28)*	*(6.40)*	*(5.76)*	*(5.85)*
R^2_{FF6-S}	2.55	0.87	2.56	4.23	4.15	4.80	2.93	9.41	3.52	2.72	2.90	4.23
α_{FF6-C}	1.00	1.00	0.79	0.95	0.94	0.89	0.68	1.07	1.22	1.29	1.11	1.12
	(4.77)	*(4.11)*	*(3.73)*	*(3.83)*	*(3.80)*	*(4.00)*	*(3.21)*	*(5.19)*	*(6.33)*	*(5.97)*	*(5.25)*	*(4.97)*
R^2_{FF6-C}	21.74	22.10	25.68	24.75	24.67	26.09	18.28	23.82	21.89	21.59	20.68	27.66
						Panel A.3: Portfolio turnover						
LO Turnover (%)	28.40	25.69	26.00	32.03	31.95	31.98	44.96	53.89	27.35	26.31	27.51	30.36
LO Breakeven (%)	2.66	3.03	2.94	2.40	2.41	2.34	1.51	1.49	3.08	3.30	2.91	2.69
LS Turnover (%)	63.11	56.14	57.47	71.13	71.08	72.66	95.21	110.77	58.65	58.84	58.21	68.95
LS Breakeven (%)	1.10	1.17	1.06	0.96	0.95	0.95	0.60	0.70	1.43	1.44	1.34	1.19

Panel B: Value-weighted portfolios

	OLS	PLS	PCA	LASSO	ENET	SVM	GBRT	RF	FFN1	FFN2	FFN3	COMB
						Panel B.1: Loss statistics						
Max 1M loss (%)	11.75	13.35	12.39	12.76	12.76	16.21	12.54	14.58	10.47	18.91	16.27	15.24
Max DD (%)	26.39	32.37	32.07	31.20	31.20	29.34	25.97	28.01	29.81	36.39	34.84	36.00
						Panel B.2: Spanning tests						
α_{FF6-S}	0.76	0.57	0.60	0.64	0.63	0.72	0.48	0.94	0.80	0.92	0.66	0.79
	(2.82)	*(2.06)*	*(2.29)*	*(2.13)*	*(2.10)*	*(2.88)*	*(1.82)*	*(3.47)*	*(3.09)*	*(3.45)*	*(2.49)*	*(2.69)*
R^2_{FF6-S}	6.72	2.50	2.05	1.63	1.74	4.81	3.60	8.18	8.64	6.02	6.62	9.45
α_{FF6-C}	0.58	0.31	0.32	0.52	0.51	0.57	0.34	0.79	0.62	0.75	0.56	0.57
	(2.40)	*(1.37)*	*(1.44)*	*(2.07)*	*(2.08)*	*(2.72)*	*(1.69)*	*(3.66)*	*(2.95)*	*(3.49)*	*(2.59)*	*(2.44)*
R^2_{FF6-C}	11.92	25.47	25.16	20.52	20.91	21.23	15.99	15.00	19.88	19.82	19.40	24.15
						Panel B.3: Portfolio turnover						
LO Turnover (%)	47.94	41.40	40.70	53.47	53.20	52.08	58.06	68.23	41.96	40.10	43.60	45.53
LO Breakeven (%)	1.20	1.39	1.45	1.11	1.12	1.18	0.85	0.84	1.59	1.79	1.44	1.35
LS Turnover (%)	89.99	77.87	79.50	101.95	101.79	99.05	117.29	136.67	78.49	79.42	78.86	94.45
LS Breakeven (%)	0.48	0.43	0.43	0.39	0.40	0.47	0.26	0.35	0.66	0.71	0.58	0.52

Table 5. Performance of Machine Learning Portfolios with Extended Holding Periods

The table presents the monthly returns on quintile portfolios from univariate sorts on the predictions of different machine learning models from Section 2.3. *Low* (*High*) denotes the quintiles of markets with the lowest (highest) predicted return. The portfolios are equal-weighted and are reformed once in three (Panel A), six (Panel B), or 12 (Panel C) months. The table also reports the average return (*H-L R*) and alpha from the global CAPM (*H-L* α) on a long-short strategy buying (selling) the long (short) quintile. The returns and alphas are reported in percentages. The numbers in parentheses are t-statistics calculated using the HAC estimator (Newey & West, 1987). The sample comprises 71 country stock markets and the testing period is from January 1995 to April 2021.

	OLS	PLS	PCA	LASSO	ENET	SVM	GBRT	RF	FFN1	FFN2	FFN3	COMP
	Panel A: Three-month holding period											
Low (L)	0.29	0.28	0.35	0.18	0.17	0.13	0.30	0.28	0.18	0.27	0.19	0.18
2	0.39	0.63	0.52	0.61	0.62	0.60	0.44	0.51	0.59	0.46	0.48	0.46
3	0.75	0.45	0.53	0.76	0.74	0.63	0.77	0.70	0.63	0.65	0.76	0.70
4	0.83	0.87	0.94	0.79	0.80	0.89	0.93	0.91	0.83	0.77	0.74	0.77
High (H)	1.38	1.41	1.33	1.32	1.32	1.40	1.22	1.25	1.41	1.50	1.47	1.54
H-L R	1.09	1.13	0.98	1.15	1.15	1.27	0.91	0.97	1.22	1.23	1.28	1.36
	(5.17)	*(5.00)*	*(4.44)*	*(4.85)*	*(4.84)*	*(5.73)*	*(4.95)*	*(4.52)*	*(5.74)*	*(6.09)*	*(6.60)*	*(6.41)*
H-L α	1.14	1.21	1.05	1.23	1.23	1.36	0.88	1.02	1.29	1.30	1.33	1.46
	(5.38)	*(5.34)*	*(4.78)*	*(5.12)*	*(5.13)*	*(5.98)*	*(4.26)*	*(4.54)*	*(5.93)*	*(6.05)*	*(6.12)*	*(6.58)*
	Panel B: Six-month holding period											
Low (L)	0.48	0.44	0.45	0.31	0.30	0.27	0.48	0.37	0.45	0.41	0.38	0.32
2	0.46	0.61	0.64	0.66	0.67	0.77	0.47	0.70	0.60	0.60	0.61	0.53
3	0.74	0.61	0.62	0.71	0.69	0.63	0.75	0.60	0.54	0.47	0.57	0.80
4	0.66	0.77	0.70	0.86	0.86	0.79	1.01	0.80	0.76	0.82	0.68	0.67
High (H)	1.29	1.21	1.24	1.11	1.11	1.17	0.93	1.17	1.29	1.35	1.38	1.33
H-L R	0.81	0.77	0.79	0.80	0.80	0.90	0.45	0.80	0.84	0.93	1.00	1.01
	(3.63)	*(3.54)*	*(3.76)*	*(3.32)*	*(3.33)*	*(3.77)*	*(2.31)*	*(3.97)*	*(3.78)*	*(4.34)*	*(4.76)*	*(4.44)*
H-L α	0.85	0.86	0.85	0.88	0.89	0.98	0.41	0.87	0.93	0.98	1.05	1.11
	(3.99)	*(3.91)*	*(3.90)*	*(3.63)*	*(3.65)*	*(4.17)*	*(2.04)*	*(3.93)*	*(4.20)*	*(4.57)*	*(4.90)*	*(4.73)*
	Panel C: Twelve-month holding period											
Low (L)	0.56	0.54	0.49	0.49	0.49	0.42	0.55	0.44	0.53	0.51	0.52	0.50
2	0.44	0.54	0.71	0.59	0.61	0.81	0.61	0.76	0.69	0.53	0.50	0.46
3	0.72	0.63	0.58	0.69	0.67	0.52	0.74	0.74	0.58	0.66	0.53	0.84
4	0.74	0.84	0.80	0.88	0.88	0.76	0.83	0.68	0.73	0.72	0.85	0.62
High (H)	1.18	1.08	1.07	0.99	0.99	1.13	0.93	1.00	1.11	1.24	1.23	1.23
H-L R	0.62	0.55	0.58	0.49	0.50	0.71	0.38	0.56	0.59	0.73	0.71	0.73
	(2.88)	*(2.53)*	*(2.87)*	*(2.07)*	*(2.09)*	*(2.87)*	*(2.05)*	*(2.79)*	*(2.70)*	*(3.39)*	*(3.50)*	*(3.33)*
H-L α	0.70	0.62	0.68	0.59	0.60	0.81	0.36	0.55	0.69	0.78	0.75	0.84
	(3.39)	*(2.92)*	*(3.15)*	*(2.43)*	*(2.44)*	*(3.44)*	*(1.81)*	*(2.77)*	*(3.19)*	*(3.59)*	*(3.59)*	*(3.81)*

Table 6. Bivariate Portfolio Sorts on Risk Changes and Predicted Returns

The table presents the monthly returns on portfolios from bivariate sorts on risk changes and return predictions from the forecast combination (COMB) machine learning model. In the first step, we sort the markets into tertiles based on 24-month changes in the measures of sovereign risk (Panel A), financial risk (Panel B), economic risk (Panel C), and political risk (Panel D). Subsequently—within each of these subsets—we sort portfolios into *Low*, *Medium*, and *High* tertiles (as indicated in the top row) based on the COMB predictions. Furthermore, we calculate a spread *H-L* portfolio that buys (sells) the markets with the *High* (*Low*) return predictions. All portfolios are equally weighted and rebalanced monthly. *H-L R* is the average monthly return on this portfolio and *H-L* α is the associated alpha from the global CAPM. The last row of each panel reports the differences in returns on the *H-L* portfolios between the tertiles of high and low risk changes. The numbers in parentheses are bootstrap (for returns) and Newey-West (1987) adjusted (for alphas) *t*-statistics. Both the returns and alphas are reported in percentages. The sample comprises 71 country stock markets and the testing period is from January 1995 to April 2021.

	Low (L)	Medium	High (H)	H-L R	t-stat$_R$	H-L α	t-stat$_\alpha$
			Pane A: Changes in sovereign risk				
High Δ sovereign risk	0.29	0.72	1.27	0.99	*(4.74)*	1.06	*(5.34)*
Medium Δ sovereign risk	0.38	0.72	1.04	0.66	*(4.28)*	0.75	*(4.65)*
Low Δ sovereign risk	0.27	0.74	1.09	0.82	*(4.17)*	0.89	*(4.03)*
High - Low Δ sovereign risk				0.16	*(0.66)*	0.17	*(0.70)*
			Pane B: Changes in financial risk				
High Δ financial risk	0.12	0.54	1.45	1.33	*(4.55)*	1.39	*(4.64)*
Medium Δ financial risk	0.40	0.72	1.12	0.72	*(3.60)*	0.77	*(3.05)*
Low Δ financial risk	0.28	0.89	1.03	0.75	*(3.25)*	0.80	*(2.80)*
High - Low Δ financial risk				0.58	*(1.75)*	0.59	*(1.55)*
			Pane C: Changes in economic risk				
High Δ economic risk	0.21	0.70	1.51	1.30	*(4.50)*	1.34	*(5.44)*
Medium Δ economic risk	0.17	0.52	0.87	0.70	*(3.33)*	0.80	*(3.63)*
Low Δ economic risk	0.49	0.88	1.16	0.67	*(2.98)*	0.66	*(2.89)*
High - Low Δ economic risk				0.63	*(1.86)*	0.68	*(2.11)*
			Pane D: Changes in political risk				
High Δ political risk	0.49	0.55	1.38	0.89	*(3.73)*	0.92	*(3.96)*
Medium Δ political risk	0.27	0.45	1.19	0.91	*(4.10)*	0.99	*(4.16)*
Low Δ political risk	0.24	0.70	1.24	0.99	*(3.80)*	1.04	*(4.14)*
High - Low Δ political risk				-0.11	*(-0.33)*	-0.12	*(-0.39)*

Table 7. Bivariate Portfolio Sorts on Mispricing and Predicted Returns

The table presents the monthly returns on portfolios from bivariate sorts on the mispricing score (*MISP*) and return predictions from the forecast combination (COMB) machine learning model. In the first step, we sort the markets into tertiles based on *MISP*. Subsequently—within each of these subsets—we sort portfolios into *Low*, *Medium*, and *High* tertiles (as is indicated in the top row) based on the COMB predictions. Furthermore, we calculate a spread *H-L* portfolio that buys (sells) the markets with the *High* (*Low*) return predictions. All portfolios are equally weighted and rebalanced monthly. *H-L R* is the average monthly return on this portfolio and *H-L* α is the associated alpha from the global CAPM. The bottom rows report the differences in returns on the *H-L* portfolios between the *Low* and *High MISP* tertiles and the middle one. The numbers in parentheses are bootstrap (for returns) and Newey-West (1987) adjusted (for alphas) *t*-statistics. The returns and alphas are reported in percentages. The sample comprises 71 country stock markets and the testing period is from January 1995 to April 2021.

	Low (L)	Medium	High (H)	H-L R	H-L α
Low MISP	0.58	0.94	1.78	1.20	1.21
	(2.08)	*(3.63)*	*(5.38)*	*(5.08)*	*(5.09)*
Medium MISP	0.33	0.71	0.69	0.36	0.40
	(1.14)	*(2.54)*	*(2.47)*	*(2.11)*	*(2.41)*
High MISP	-0.08	0.46	1.08	1.16	1.15
	(-0.23)	*(1.40)*	*(3.14)*	*(4.12)*	*(4.09)*
Low-Medium MISP				0.84	0.81
				(2.98)	*(2.82)*
High-Medium MISP				0.80	0.75
				(2.43)	*(2.46)*

Table 8. Performance of Machine Learning Portfolios in Subperiods

The table presents the monthly returns on quintile portfolios from univariate sorts on the predictions of different machine learning models from Section 2.3. *Low* (*High*) denotes the quintiles of markets with the lowest (highest) predicted return. The table also reports the average return (*H-L R*) and alpha from the global CAPM (*H-L* α) on a long-short strategy buying (selling) the long (short) quintile. Both the returns and alphas are reported in percentages. The numbers in parentheses are t-statistics that are calculated using the HAC estimator (Newey & West, 1987). The sample comprises 71 country stock markets. The results are reported for two subperiods: January 1995 to February 2008 (Panel A) and March 2008 to April 2021 (Panel B).

	OLS	PLS	PCA	LASSO	ENET	SVM	GBRT	RF	FFN1	FFN2	FFN3	COMB
	Panel A: First half (January 1995 - February 2008)											
Low (L)	0.11	0.35	0.50	0.23	0.24	0.17	0.23	0.13	0.04	0.13	0.04	-0.04
2	0.59	0.96	0.69	0.85	0.85	0.64	0.59	0.85	0.76	0.85	0.63	0.78
3	1.03	0.65	0.85	0.90	0.89	1.11	0.98	0.96	0.83	0.96	1.16	1.05
4	1.43	1.16	1.10	1.28	1.30	1.35	1.53	0.97	1.33	0.97	1.21	1.18
High (H)	2.12	2.15	2.13	2.01	2.00	1.98	1.94	2.37	2.30	2.37	2.21	2.29
H-L R	2.01	1.80	1.63	1.78	1.76	1.81	1.71	2.23	2.26	2.23	2.16	2.32
	(7.03)	*(5.36)*	*(4.54)*	*(4.77)*	*(4.70)*	*(6.65)*	*(6.03)*	*(6.96)*	*(8.20)*	*(6.96)*	*(7.33)*	*(6.72)*
H-L α	2.06	1.89	1.75	1.90	1.88	1.90	1.75	2.29	2.33	2.29	2.14	2.39
	(7.12)	*(5.63)*	*(4.87)*	*(5.14)*	*(5.06)*	*(6.74)*	*(6.53)*	*(7.07)*	*(8.36)*	*(7.07)*	*(7.11)*	*(6.91)*
	Panel B: Second half (March 2008 - April 2021)											
Low (L)	0.13	0.13	0.11	0.13	0.14	0.05	0.20	-0.04	-0.02	-0.04	0.04	0.03
2	0.37	0.38	0.39	0.44	0.43	0.46	0.27	0.32	0.23	0.32	0.44	0.40
3	0.35	0.16	0.15	0.25	0.25	0.36	0.34	0.45	0.51	0.45	0.33	0.28
4	0.29	0.43	0.48	0.16	0.15	0.17	0.46	0.20	0.24	0.20	0.24	0.34
High (H)	0.90	0.96	0.92	1.07	1.08	1.01	0.77	1.11	1.07	1.11	0.99	0.99
H-L R	0.77	0.83	0.81	0.94	0.95	0.96	0.57	1.15	1.09	1.15	0.95	0.95
	(2.61)	*(2.72)*	*(2.94)*	*(3.11)*	*(3.13)*	*(3.08)*	*(2.24)*	*(4.14)*	*(3.67)*	*(4.14)*	*(3.33)*	*(3.32)*
H-L α	0.82	0.88	0.85	0.99	1.00	1.04	0.56	1.22	1.12	1.22	1.00	1.02
	(2.75)	*(2.87)*	*(3.11)*	*(3.17)*	*(3.19)*	*(3.25)*	*(2.32)*	*(4.65)*	*(3.78)*	*(4.65)*	*(3.63)*	*(3.65)*

Table 9. Machine Learning Predictions and International Variation in Limits to Arbitrage

The table reports the average slope coefficients from cross-sectional regressions of monthly country equity returns on the predictions from machine learning models, proxies for limits to arbitrage, and interaction terms. We interact the model predictions (*PRED*) with four binary variables associated with limits to arbitrage and market development: *SIZE*, *IRISK*, *LIQ*, and *EMER*. *SIZE* takes the value of one if market capitalization at time *t*-1 is lower than a cross-sectional median, and zero otherwise. *IRISK* takes the value of one if idiosyncratic risk at *t*-1 is higher than a cross-sectional median, and zero otherwise. LIQ takes the value of one if Amihud's (2002) illiquidity ratio at *t*-1 is higher than a cross-sectional median, and zero otherwise. Finally, *EMER* takes the value of one for emerging and developing markets—and zero otherwise. The monthly predictions of country index returns come from 11 different models that were described in Section 2.3. The numbers in parentheses are t-statistics that are calculated using the HAC estimator (Newey & West, 1987). The coefficients for *SIZE*, *IRISK*, *LIQ,* and *EMER* are multiplied by 100. $\overline{R^2}$ is the average cross-sectional adjusted coefficient of determination (expressed in percentage terms). The sample comprises 71 country stock markets, and the testing period is from January 1995 to April 2021.

	OLS	PLS	PCA	LASSO	ENET	GBRT	RF	FFN1	FFN2	FFN3	COMP
Panel A: Univariate regressions											
PRED	0.49	0.47	0.52	1.24	1.25	0.36	0.48	0.66	0.70	0.60	0.83
	(4.91)	*(4.70)*	*(4.81)*	*(4.37)*	*(4.37)*	*(5.46)*	*(5.59)*	*(5.80)*	*(5.70)*	*(4.23)*	*(5.41)*
$\overline{R^2}$	4.53	4.44	4.55	5.90	5.91	2.43	6.69	6.44	5.61	4.64	6.07
Panel B: Controlling for market size											
PRED	0.23	0.20	0.20	0.58	0.59	0.18	0.37	0.64	0.60	0.38	0.72
	(2.75)	*(2.36)*	*(2.08)*	*(2.17)*	*(2.20)*	*(3.59)*	*(4.76)*	*(5.85)*	*(5.23)*	*(2.60)*	*(5.04)*
SIZE	-0.25	-0.38	-0.41	-0.45	-0.45	-0.39	-0.28	-0.13	-0.17	-0.27	-0.15
	(-1.15)	*(-1.89)*	*(-2.03)*	*(-2.46)*	*(-2.46)*	*(-2.11)*	*(-1.56)*	*(-0.73)*	*(-0.90)*	*(-1.31)*	*(-0.80)*
PRED*SIZE	0.34	0.40	0.45	0.39	0.39	0.49	0.38	0.04	0.13	0.28	0.13
	(2.02)	*(2.60)*	*(2.96)*	*(3.21)*	*(3.22)*	*(3.75)*	*(3.70)*	*(0.36)*	*(0.97)*	*(1.67)*	*(0.97)*
$\overline{R^2}$	7.75	9.24	9.15	10.14	10.15	7.93	10.82	9.66	9.21	8.99	9.55
Panel C: Controlling for market idiosyncratic risk											
PRED	0.36	0.27	0.26	0.75	0.75	0.22	0.38	0.76	0.72	0.46	0.82
	(4.85)	*(2.84)*	*(2.29)*	*(2.67)*	*(2.69)*	*(4.19)*	*(4.99)*	*(5.82)*	*(6.07)*	*(3.15)*	*(5.66)*
IRISK	-0.21	-0.48	-0.57	-0.49	-0.50	-0.56	-0.49	-0.02	-0.13	-0.27	-0.21
	(-1.01)	*(-2.25)*	*(-2.71)*	*(-2.64)*	*(-2.65)*	*(-2.80)*	*(-2.82)*	*(-0.11)*	*(-0.62)*	*(-1.22)*	*(-1.06)*
PRED*IRISK	0.15	0.29	0.38	0.29	0.29	0.42	0.31	-0.15	-0.04	0.16	0.00
	(1.06)	*(1.86)*	*(2.34)*	*(2.85)*	*(2.87)*	*(3.44)*	*(3.49)*	*(-1.42)*	*(-0.37)*	*(0.99)*	*(0.01)*
$\overline{R^2}$	6.51	8.02	8.24	8.98	9.00	6.79	10.04	8.64	8.09	7.60	8.46
Panel D: Controlling for market liquidity											
PRED	0.38	0.30	0.30	0.71	0.72	0.23	0.40	0.73	0.68	0.49	0.84
	(4.58)	*(3.31)*	*(2.78)*	*(2.51)*	*(2.42)*	*(4.53)*	*(5.27)*	*(6.89)*	*(6.01)*	*(3.30)*	*(5.87)*
LIQ	-0.09	-0.33	-0.40	-0.44	-0.44	-0.33	-0.24	0.00	-0.05	-0.17	-0.09
	(-0.51)	*(-1.94)*	*(-2.42)*	*(-2.93)*	*(-2.93)*	*(-2.02)*	*(-1.70)*	*(-0.03)*	*(-0.31)*	*(-0.95)*	*(-0.60)*
PRED*LIQ	0.14	0.27	0.34	0.30	0.31	0.42	0.30	-0.09	0.01	0.16	0.01
	(0.80)	*(1.71)*	*(2.21)*	*(2.58)*	*(2.59)*	*(3.22)*	*(3.01)*	*(-0.77)*	*(0.08)*	*(0.92)*	*(0.06)*
$\overline{R^2}$	6.79	8.26	8.26	8.84	8.86	6.62	9.76	8.72	8.21	7.94	8.70
Panel E: Controlling for market development											
PRED	0.27	0.23	0.26	0.67	0.68	0.16	0.36	0.71	0.68	0.49	0.84
	(3.13)	*(2.45)*	*(2.34)*	*(2.71)*	*(2.74)*	*(3.33)*	*(4.77)*	*(5.68)*	*(5.50)*	*(3.12)*	*(5.38)*
EMER	-0.22	-0.36	-0.40	-0.39	-0.39	-0.38	-0.31	-0.03	-0.09	-0.21	-0.11
	(-1.02)	*(-1.54)*	*(-1.76)*	*(-1.87)*	*(-1.87)*	*(-1.86)*	*(-1.60)*	*(-0.14)*	*(-0.44)*	*(-0.96)*	*(-0.55)*
PRED*EMER	0.25	0.33	0.36	0.34	0.34	0.46	0.31	-0.10	0.01	0.17	0.00
	(1.59)	*(2.06)*	*(2.26)*	*(2.75)*	*(2.76)*	*(3.39)*	*(3.10)*	*(-0.85)*	*(0.05)*	*(1.01)*	*(-0.01)*
$\overline{R^2}$	7.65	9.45	9.33	10.29	10.30	8.15	10.90	9.85	9.32	9.23	9.65

Online Appendix for "Empirical Asset Pricing via Machine Learning: The Global Edition"

[FOR ONLINE PUBLICATION ONLY]

Abstract

Section A provides additional tables and figures from the study. Table A1 presents basic statistics of returns on country stock markets that are included in the sample. Table A2 details the market characteristics that are covered in the study. Table A3 displays the statistical properties of country-level asset pricing factors. Table A4 reports the results of the bivariate portfolio sorts mispricing and return predictions from the different models. Section B contains the description of the machine learning methods that are used in this study: linear regressions (B.1), dimension reduction techniques (B.2), penalized linear regressions (B.3), support vector machine(B.4), tree models (B.5), neural networks (B.6), and forecast combination (B.7).

A. Additional Tables and Figures From the Study

Table A1. Country Stock Markets Covered in the Study

The table presents the list of country stock markets considered in this study along with the essential statistical properties of index excess returns: average, standard deviation, skewness, kurtosis, minimum, and maximum. All the return data is in percentages. *No.* is the running number. *Start date* indicates the first available monthly return. *#Obs.* is the number of observations in the sample. The last column, *Market value*, displays the average monthly aggregate market capitalization—expressed in U.S. dollars.

No.	Country	Average	Standard deviation	Skewness	Kurtosis	Minimum	Maximum	Start date	#Obs.	Market value
1	Argentina	1.07	14.30	2.07	15.71	-54.53	118.14	Jan 1985	414	28.83
2	Australia	0.87	6.54	-1.08	5.85	-43.86	18.09	Jan 1985	436	610.74
3	Austria	0.95	7.17	0.18	4.37	-34.33	36.57	Jan 1985	436	65.97
4	Bahrain	0.33	3.56	-0.39	3.04	-15.56	11.29	Jan 2004	208	15.58
5	Belgium	0.80	5.67	-0.51	4.05	-32.41	23.90	Jan 1985	436	189.11
6	Brazil	1.74	14.49	0.85	5.17	-56.67	89.07	Jan 1985	421	428.34
7	Bulgaria	1.42	9.46	0.38	3.31	-36.67	38.56	Nov 2000	246	2.26
8	Canada	0.67	5.31	-0.79	3.63	-26.58	20.33	Jan 1985	436	910.28
9	Chechia	0.95	8.32	1.37	13.10	-26.96	68.45	Dec 1993	329	27.41
10	Chile	1.31	7.70	0.96	9.56	-32.16	62.36	Jan 1985	436	104.03
11	China	1.21	12.04	6.40	78.76	-28.08	153.22	Jun 1994	323	1729.92
12	Colombia	0.85	8.39	0.49	3.95	-35.39	48.68	Jan 1988	400	60.17
13	Croatia	0.52	6.94	0.13	5.40	-31.15	32.34	Nov 2005	186	13.90
14	Cyprus	0.13	11.51	0.81	10.06	-65.03	69.54	Jan 1993	340	5.53
15	Denmark	1.03	5.47	-0.39	2.06	-26.47	20.36	Jan 1985	436	151.04
16	Egypt	0.67	8.21	0.11	2.72	-33.84	39.67	Jan 1995	316	25.68
17	Estonia	1.42	9.22	0.26	3.65	-41.35	42.25	Aug 1995	309	1.62
18	Finland	1.08	7.68	0.03	1.45	-29.08	29.44	Jan 1985	436	137.15
19	France	0.94	5.96	-0.27	1.01	-21.61	21.43	Jan 1985	436	1206.26
20	Germany	0.77	6.04	-0.39	1.10	-20.79	19.23	Jan 1985	436	1027.39
21	Greece	1.01	10.72	0.79	3.99	-33.68	57.84	Jan 1985	436	51.26
22	Hong Kong	1.04	7.21	-0.57	5.48	-45.99	28.93	Jan 1985	436	945.28
23	Hungary	0.85	9.45	0.45	5.96	-39.22	59.52	Feb 1991	363	18.54
24	Iceland	0.66	8.95	-3.14	23.35	-75.07	23.70	Jul 2002	225	8.22
25	India	0.98	9.66	0.51	3.38	-32.14	53.62	Jan 1985	436	556.92
26	Indonesia	1.15	12.12	2.04	15.56	-41.15	93.68	Jan 1988	400	126.28
27	Ireland	0.96	6.50	-0.37	2.34	-25.66	26.28	Jan 1985	436	57.83
28	Israel	0.80	6.14	-0.36	0.76	-19.84	16.85	Jan 1985	436	69.66
29	Italy	0.70	7.05	0.11	0.80	-23.19	26.56	Jan 1985	436	456.96
30	Japan	0.41	5.87	0.26	1.17	-18.19	26.42	Jan 1985	436	3473.27
31	Jordan	0.32	5.55	-0.07	6.01	-31.05	29.32	Dec 1987	398	13.64
32	Korea	1.05	9.73	1.02	6.54	-32.48	70.35	Jan 1985	436	436.05
33	Kuwait	0.52	5.27	-0.31	2.14	-18.10	17.42	Jan 2004	208	50.60
34	Latvia	1.08	9.21	1.49	11.97	-35.49	63.52	May 1996	299	1.42
35	Lithuania	0.93	9.11	3.22	27.03	-32.98	81.10	Jan 1996	304	2.39
36	Luxembourg	0.76	5.73	-0.41	3.26	-27.51	22.64	Jan 1985	436	22.46
37	Malaysia	0.63	7.73	0.32	5.95	-33.72	45.76	Jan 1985	436	174.99
38	Malta	0.36	4.95	-0.13	1.22	-18.11	14.02	Feb 2000	255	3.63
39	Mauritius	0.72	5.33	-0.22	4.02	-24.59	20.00	Aug 1989	380	4.40
40	Mexico	1.50	9.49	-0.99	6.31	-60.89	35.57	Jan 1985	436	199.86
41	Morocco	0.82	4.93	-0.02	3.82	-24.36	23.23	Apr 1994	325	26.87

42	Netherlands	0.85	5.37	-1.00	3.95	-30.99	16.27	Jan 1985	436	444.64
43	New Zealand	0.86	6.54	-0.25	2.87	-34.96	29.21	Jan 1985	436	35.07
44	Nigeria	0.26	7.37	-0.54	1.20	-25.45	16.02	Oct 2009	139	21.77
45	Norway	1.03	7.42	-0.54	1.88	-30.70	23.82	Jan 1985	436	128.85
46	Oman	0.27	4.36	-0.60	3.67	-20.76	15.47	Nov 2005	186	10.44
47	Pakistan	0.79	9.06	0.09	2.94	-38.34	34.96	Jan 1988	400	21.71
48	Peru	0.83	6.61	-0.05	4.10	-29.48	30.96	Jan 1993	340	34.04
49	Philippines	1.29	8.98	0.92	6.45	-34.12	55.90	Jan 1985	436	79.21
50	Poland	1.22	11.99	1.82	13.68	-33.79	100.65	May 1991	360	77.80
51	Portugal	0.22	6.16	-0.27	1.36	-28.08	21.62	Feb 1988	399	47.95
52	Qatar	1.02	8.04	0.66	5.06	-24.39	44.88	Jan 2004	208	96.62
53	Romania	1.27	12.51	1.07	8.20	-43.97	84.53	Jan 1997	292	13.25
54	Russia	1.74	12.95	0.03	2.50	-57.04	48.06	Jan 1995	316	380.59
55	Saudi Arabia	0.18	7.19	-0.33	1.23	-23.90	21.17	Nov 2005	186	254.90
56	Singapore	0.65	6.67	-0.44	4.21	-37.62	25.97	Jan 1985	436	238.44
57	Slovakia	0.47	4.33	-0.50	2.37	-18.86	14.55	Apr 2006	181	4.31
58	Slovenia	0.47	5.97	-0.33	1.99	-23.40	19.77	Jan 1999	256	5.71
59	South Africa	0.92	7.61	-0.61	1.78	-35.78	19.45	Jan 1985	436	220.83
60	Spain	0.92	6.79	0.02	1.77	-24.69	28.18	Jan 1985	436	423.18
61	Sri Lanka	0.70	8.13	0.74	2.58	-24.31	37.72	Jul 1987	406	5.50
62	Sweden	1.09	6.89	-0.29	1.16	-26.20	22.42	Jan 1985	436	302.00
63	Switzerland	0.92	4.76	-0.44	1.16	-18.85	15.23	Jan 1985	436	768.40
64	Taiwan	0.94	9.63	0.73	4.08	-33.87	56.73	Jan 1988	400	339.97
65	Thailand	1.11	9.42	0.13	2.68	-32.61	40.57	Jan 1985	436	126.92
66	Turkey	1.77	15.83	1.54	8.69	-41.25	119.58	Feb 1986	423	93.35
67	UAE	1.12	7.56	0.59	3.61	-22.83	33.89	Jan 2004	208	124.94
68	UK	0.68	5.16	-0.33	1.55	-21.83	16.35	Jan 1985	436	2100.75
69	USA	0.82	4.33	-0.73	2.36	-20.87	13.48	Jan 1985	436	13193.37
70	Venezuela	3.31	23.43	1.82	14.62	-95.54	172.47	Jan 1988	388	10.34
71	Vietnam	0.47	8.09	-0.41	1.40	-25.27	23.29	May 2007	168	50.50

Table A2. Market Characteristics

The table details the market characteristics that are considered in the study. *No.* is the running number. *Symbol* denotes the abbreviation of the anomaly that is used in the study. *Panels A* to *L* contain a replication of anomalies from the stock level, so both original references and their country-level replications are provided. The signals in *Panels M* to *P* do not have their firm-level parallels. The data sources in the last column are indicated in the order of priority. If the data from the first source is unavailable, it is spliced and backfilled with the data from the second source.

No.	Abbr.	Anomaly	Key Original References	Key Country-Level References	Implementation Details	Data source(s)
Panel A: Value vs. Growth						
1	EP	Earnings yield	Basu (1977)	Macedo (1995), Heckman et al. (1996), Asness, Liew et al. (1997), Kim (2012), Asness et al. (2013), Angelidis and Tessaromatis (2017), Lawrenz and Zorn (2017), Zaremba et al. (2020), Baltussen et al. (2021), Radha (2021)	Trailing 12-month net profit at t-5 to the market value of equity at t-1.	Datastream, Global Financial Data
2	BM	Book-to-market ratio	Rosenberg et al. (1985)		Book value of equity at t-5 to market value of equity at t-1.	Datastream
3	CP	Cash flow-to-price ratio	Lakonishok et al. (1994)		Trailing 12-month cash flow at t-5 to the market value of equity at t-1.	Datastream
4	SP	Sales-to-price ratio	Barbee et al. (1996)		Trailing 12-month sales at t-5 to the market value of equity at t-1.	Datastream
5	EDEV	EBITDA-to-EV ratio	Loughran and Wellman (2011)		Trailing 12-month earnings before interest, taxes, depreciation, and amortization (EBITDA) at t-5 to enterprise value (EV) at t-1.	Datastream
6	FEP	Forward earnings yield	Elgers et al. (2001)		I/B/E/S estimates of forward 12-month earnings to the market value of equity at t-1.	Datastream
7	DY	Dividend yield	Litzenberger and Ramaswamy (1979)		Trailing 12-month dividend yield at t-1.	Datastream, Global Financial Data
8	CAPE	Cyclically adjusted price-to-earnings ratio	Cambell and Shiller (1998, 2001), Bunn et al. (2014), Siegel (2016)		The current market value of an index portfolio divided by the average annual earnings during the past 10 years that has been adjusted for the inflation rate—all recorded at t-5. Where net earnings for the country were negative, the ratio of 100 has been used.	Global Financial Data
Panel B: Size and Liquidity						
9	MV	Market value	Banz (1981)	Keppler and Traub (1993), Lee (2011), Liang and Wei (2012), Fisher et al. (2017), Chen et al. (2018)	The natural logarithm of market value of equity in USD at t-1 (multiplied by -1).	Datastream, Global Financial Data
10	Illiq	Amihud ratio	Amihud (2002)		A reciprocal of the annualized average ratio of the dollar trading volume-to-return ratio over the last 260 trading days (≈one year).	Datastream

11	Turn	Turnover ratio	Datar et al. (1998)		The average ratio of the dollar trading volume to the market value of equity over the last 780 trading days (≈one year). The final value is annualized and multiplied by -1.	Datastram
12	Dvol	Dollar volume	Brennan et al. (1998)		The natural logarithm of the annualized value of the average daily trading value over the last 22 trading days (multiplied by -1).	Datastream
13	TurnVar	Turnover volatility	Chordia et al. (2001)		The standard deviation of the daily turnover ratio over the last 260 days (≈one year).	Datastream
14	VolVar	Volume volatility	Chordia et al. (2001)		The natural logarithm of the standard deviation of the daily dollar volume over the last 260 days (≈one year).	Datastream
Panel C: Price Risk						
15	Beta	Beta	Fama and MacBeth (1973)	Macedo (1995), Bali and Cakici (2010), Frazzini and Pedersen (2014), Umutlu (2015), Baghdadabad and Mallik (2018), Atilgan et al. (2019), Gao et al. (2019), Hollstein et al. (2019), Zaremba et al. (2020), Liang and John Wei (2020), Baltussen et al. (2021)	The slope coefficient from the regression of index excess returns on the global market factor (MKT), estimated over a trailing 36-month period.	Datastream, Global Financial Data
16	Cor	Correlation	Asness et al. (2020)		Pearson's product-moment correlation coefficient between the index excess returns and the global market factor (MKT), estimated over a trailing 36-month period (multiplied by -1).	Datastream, Global Financial Data
17	Vol	Total volatility	Ang et al. (2006), Baker, Bradley, and Wurgler (2011)		The standard deviation of the excess returns, estimated over a trailing 36-month period.	Datastream, Global Financial Data
18	IVol	Idiosyncratic volatility	Ang et al. (2006)		The volatility of the residuals from a regression of index excess returns on the global market factor (MKT), estimated over a trailing 36-month period.	Datastream, Global Financial Data
19	RNG	Price range	Blau and Whitby (2017)		The difference between the natural logarithms of the maximum and minimum index values over the last 260 trading days (≈one year).	Datastream
20	VAR	Value at risk	Bali and Cakici (2004)		The 5th percentile of monthly returns over the last 60 months (multiplied by -1).	Datastream, Global Financial Data
Panel D: Momentum						
21	LTMom	Long-term momentum	Fama and French (1996)	Asness et al. (1997), Chan et al. (2000), Kortas et al. (2005), Balvers and Wu (2006), Bhojraj and Swaminathan (2006), Asness et al. (2013), Clare et al. (2016), Geczy and	The average monthly log-return in months t-12 to t-2.	Datastream, Global Financial Data
22	MtMom	Medium-term momentum	Jegadeesh and Titman (1993)		The average monthly log-return in months t-7 to t-2.	Datastream, Global Financial Data
23	StMom	Short-term momentum	Medhat and Schmelling (2021)		The log-return in the last month (t-1).	Datastream, Global Financial Data

24	ResMom	Residual momentum	Blitz et al. (2011), Blitz et al. (2020)	Samonov (2016), Zaremba et al. (2020), Baltussen et al. (2019, 2021)	The average residual from the regression of index excess returns on the global market factor (MKT) in months *t*-12 to *t*-2. The regression model is estimated for months *t*-36 to *t*-1. Following Blitz, Hanauer, and Vidojevic (2020), the residuals are scaled by their standard deviation.	Datastream, Global Financial Data
Panel E: Seasonality						
25	Seas	Cross-sectional seasonality	Heston and Sadka (2008)	Keloharju et al. (2016, 2021), Baltussen et al.(2021)	The average same-calendar month log-return over trailing 20 years (as available).	Datastream, Global Financial Data
Panel F: Profitability						
26	ROA	Return on asset	Balakrishnan et al. (2010), Kogan and Papanikolaou (2013)	Calice and Lin (2021), Zaremba and Andreu (2018)	Trailing 12-month net profit to total assets at *t*-5.	Datastream
27	ROE	Return on equity	Haugen and Baker (1996)		Trailing 12-month net profit to shareholder equity at *t*-5.	Datastream
28	CFA	Cash profitability	Ball et al. (2016)		Trailing 12-month cash flow to total assets at *t*-5.	Datastream
29	EBA	EBIT-to-asset	Cakici et al. (2021)		Trailing 12-month earnings before interest and taxes (EBIT) to total assets at *t*-5.	Datastream
30	NM	Net margin	Soliman (2008)		Trailing 12-month net profit to total sales at *t*-5.	Datastream
31	SG	Sales growth	Lakonishok et al. (1994)		The 12-month change in the natural logarithms of trailing 12-month total sales recorded at month *t*-5.	Datastream
32	ROACh	ROA change	Balakrishnan et al. (2010)		The difference between the return on assets ROA at month t-5 and its value 12-months earlier.	Datastream
33	ROECh	ROE change	Balakrishnan et al. (2010)		The difference between the return on assets ROE at month t-5 and its value 12-months earlier.	Datastream
34	NMCh	Net margin change	Soliman (2008)		The difference between the net margin at month t-5 and its value 12-months earlier.	Datastream
35	EarVol	Earnings volatility	Francis et al. (2004)		The standard deviation of the return on assets (ROA) over the last 16 quarters.	Datastream
36	AT	Asset turnover	Soliman (2008)		The ratio of trailing 12-month sales to total assets at *t*-5.	Datastream
Panel G: Indebtedness						
37	DE	Debt-to-equity ratio	Fama and French (1992), Barbee et al. (1996)	Calice and Lin (2021), Zaremba and Andreu (2018)	The ratio of net debt to equity at month *t*-5.	Datastream
38	DM	Debt-to-capitalization ratio	Bhandari (1998), Penman et al. (2007)		The ratio of net debt at *t*-5 to market value of equity at month *t*-1.	Datastream

Panel H: Skewness						
39	SKEW	Total skewness	Amaya et al. (2015), Bali et al. (2016)	Harvey (2000), Baltas and Salinas (2019)	The moment measure of skewness of daily returns over the last 780 trading days (≈three years). If the daily data is not available, then the skewness estimated over the last 36 monthly returns is used.	Datastream, Global Financial Data
40	COSKEW	Co-skewness	Harvey, and Siddique (2000)		The co-skewness that is calculated following the method of Harvey and Siddique (2000); i.e., as a slope coefficient on the squared market factor return, estimated over the last 36 months.	Datastream, Global Financial Data
Panel I: Long-Term Reversal						
41	LrRev	Long-run reversal	DeBondt and Thaler (1985)	Richard (1997), Balvers et al. (2000), Balvers and Wu (2006), Spierdijk et al. (2012), Zaremba et al. (2019)	The average log-return in months t-60 to t-13 (multiplied by -1).	Datastream, Global Financial Data
Panel J: Technical Analysis						
42	MA	Moving average	Brock et al. (1992), Sullivan et al. (1999), Han et al. (2013)	Du (2008), Malin and Bornholt (2010), Hsu et al. (2010), Neely et al. (2015), Clare et al. (2016), Zaremba et al. (2020, 2021a), Baltussen et al. (2021), Sermpinis et al. (2021)	The ratio of the most recent index value to the average value over the last 250 days.	Datastream
43	H52	52-week high effect	George and Hwang (2004)		The ratio of the most recent index value to the maximum value over the previous 260 days (≈one year).	Datastream
44	MAD	Moving average distance	Avramov et al. (2021)		The ratio of the average index value over the last 21 days to the average value over the last 200 days.	Datastream
45	BRTH	Market breadth	Qi and Zhao (2008), Fang et al. (2014)		The difference in the numbers of raising and falling stocks within the index portfolio over the last month divided by their sum.	Datastream
Panel K: Investment and Issuance						
46	AG	Asset growth	Cooper et al. (2008)	Baker and Wurgler (2000), Boudoukh et al. (2007), Zaremba and Andreu (2018), Wen (2019), Calice and Lin (2021)	The 12-month change in the natural logarithms of total assets at t-5 (multiplied by -1).	Datastream
47	CEI	Composite equity issuance	Daniel and Titman (2006)		The 36-month change in the natural logarithms of market value minus the 36-month total log-return.	Datastream, Global Financial Data
48	NSI	Net share issuance	Pontiff and Woodgate (2008)		The 12-month change in the aggregate number of shares outstanding in each country (from t-13 to t-1). The shares outstanding are estimated through the use of the Share Value Index by Global Financial Data.	Global Financial Data

49	HR	Hiring rate	Belo et al. (2014)	The 12-month change in the natural logarithms of the number of employees at t-5 (multiplied by -1).	Datastream
50	PY	Payout ratio	Lamont (1998)	The ratio of 12-month trailing dividend to earnings ratio at t-5.	Datastream, Global Financial Data
Panel L: Macroeconomic Conditions					
51	Unemp	Unemployment rate	Erb et al. (1995b), Flannery and Protopapadakis (2002), Rapach et al. (2005, 2010), Campbell and Thompson (2008), Welch and Goyal (2008), Rapach and Zhou (2013), Møller and Rangvid (2015), Baetje and Menkhoff (2016), Hollstein et al. (2020), Atanasov (2021), Goyal et al. (2021),	The unemployment rate at t-5.	Global Financial Data
52	Infl	Inflation rate		The 12-month consumer inflation rate at t-5.	Global Financial Data
53	GDPGr	GDP growth		The annual nominal gross domestic product (GDP) growth rate at t-5.	Global Financial Data
54	REERCh	Real effective exchange rate change		The average monthly log-change on the real effective exchange rate in months t-60 to t-1.	Global Financial Data
55	DebtGDP	Debt-to-GDP ratio		The government debt-to-gross domestic product (GDP) ratio at t-5.	Global Financial Data
56	PrimBal	Primary balance		The difference between government revenues and expenditures scaled by the gross domestic product (GDP) at t-5.	Global Financial Data
57	M1Ch	M1 change		The 12-month log-change in the M1 measure of money supply (i.e., from t-13 to t-1).	Global Financial Data
58	M2Ch	M2 change		The 12-month log-change in the M2 measure of money supply (i.e., from t-13 to t-1).	Global Financial Data
59	PopCh	Population	Geanakoplos et al. (2004), Goyal (2004), Ang and Maddaloni (2005), Brunetti and Torricelli (2010), Cornell (2012), Arnott and Chaves (2012)	The change in the country's total population over the last 10 years (i.e., from t-121 to t-1).	Global Financial Data
60	DebtGDPCh	Change in the debt-to-GDP ratio	Wisniewski and Jackson (2021)	The 12-month change in the government debt-to-gross domestic product (GDP) ratio recorded at t-5 (i.e. from t-17 to t-5).	Global Financial Data
61	MacroMom	Macro momentum	Brooks (2017)	An average of three monthly z-scores that are associated with the 12-month changes in the following macroeconomic variables: a) Annual gross domestic product (GDP) growth rate at t-5; b) unemployment rate at t-5; and c) 12-month consumer inflation rate at t-5. The z-scores for components b and c are multiplied by -1.	Global Financial Data
Panel M: Fixed-Income Markets					
62	BillYld	Treasury bill yield	Chen et al. (1986), Campbell (1987), Fama and French (1989), Welch (2008), Rapach et al. (2005), Hjalmarsson	The annualized yield to redemption of the three-month treasury bills at t-1.	Global Financial Data

63	BondYld	Government bond yield	(2010), Rapach and Zhou (2013), Pettenuzzo et al (2014), Baetje and Menkhoff (2016), Andrade and Chhaochharia (2018), Goyal et al. (2021)	The yield to maturity (expressed on the annual basis) of the 10-year government bonds at t-1.	Datastream, Global Financial Data
64	YldCrv	Yield curve slope		The difference between the annual yields to redemption of 10-year government bonds and three-month treasury bills at *t*-1.	Datastream, Global Financial Data
65	CrvCh	Yield curve change		The 12-month change in the yield curve slope; where the yield curve slope is defined as the difference between the annual yields to redemption of 10-year government bonds and three-month treasury bills at t-1.	Datastream, Global Financial Data
66	YldCh	Yield change	Zaremba et al. (2021b)	The 12-month change in the yield to maturity of the 10-year government bonds (from *t*-13 to *t*-1).	Datastream, Global Financial Data
67	BondMom	Bond momentum	Pitkäjärvi et al. (2020)	The average monthly log-return on 10-year government bonds over the last 12 months (*t*-12 to *t*-1), expressed in local currency.	Datastream, Global Financial Data
Panel N: Financial and Economic Risk					
68	ForDebt	Foreign debt as a percentage of GDP	Erb et al. (1995a, 1996a, 1996b, 1997), Ferson and Harvey (1994), Bekaert et al. (1997), Harvey (2004), Aggarwal and Goodell (2008), Harvey and Ferson (2008), Suleman et al. (2017)	The estimated gross foreign debt in a given year, converted into U.S. dollars at the average exchange rate for that year, is expressed as a percentage of the gross domestic product converted into U.S. dollars at the average exchange rate for that year (ICRG risk rating).	PRS Group
69	XRStab	Exchange rate stability		The appreciation or depreciation of a currency against the U.S. dollar (against the German mark /euro in the case of the USA) over a calendar year or the most recent 12-month period is calculated as a percentage change (ICRG risk rating).	PRS Group
70	DebtServ	Foreign debt service as a percentage of exports of goods and services		The estimated foreign debt service for a given year (converted into U.S. dollars at the average exchange rate for that year) is expressed as a percentage of the sum of the estimated total exports of goods and services for that year. It is converted into U.S. dollars at the average exchange rate for that year (ICRG risk rating).	PRS Group
71	CAXGS	Current account as a percentage of exports of goods and services		The balance of the current account of the balance of payments for a given year (converted into U.S. dollars at the average exchange rate for that year) is expressed as a percentage of the sum of the estimated total exports of goods and services for that year. It is converted into U.S. dollars at the average exchange rate for that year (ICRG risk rating).	PRS Group

72	IntLiq	Net international liquidity as months of import cover		The total estimated official reserves for a given year (converted into U.S. dollars at the average exchange rate for that year), including official holdings of gold (converted into U.S. dollars at the free market price for the period), but excluding the use of IMF credits and the foreign liabilities of the monetary authorities. It is divided by the average monthly merchandise import cost, which is converted into U.S. dollars at the average exchange rate for the period (ICRG risk rating).	PRS Group
73	GDPHead	Gross domestic product per head		The estimated GDP per head for a given year (converted into U.S. dollars at the average exchange rate for that year) is expressed as a percentage of the average of the estimated total GDP of all the countries covered by ICRG (ICRG risk rating).	PRS Group
74	CACC	Current account as a percentage of GDP		The estimated balance on the current account of the balance of payments for a given year (converted into U.S. dollars at the average exchange rate for that year) is expressed as a percentage of the estimated GDP of the country concerned, which is converted into U.S. dollars at the average rate of exchange for the period covered (ICRG risk rating).	PRS Group
75	SovRet	Sovereign risk	Erb et al. (1995a, 1996a), Avramov et al. (2012)	We closely follow Avramov et al. (2012) and transform sovereign ratings from three agencies; S&P, Fitch, and Moody's; into numerical values from 1 to 24—increasing in credit risk. The final score is the average numerical rating of the available agencies.	Bloomberg
Panel P: Political Risks and Regimes					
76	GovStab	Government stability	Diamonte et al. (1996), Erb et al. (1996b), Bilson et al. (2002), Lehkonen and Heimonen (2015), Vortelinos and Saha (2016), Dimic et al. (2015)	An assessment of both the government's ability to carry out its declared program(s), as well as its ability to stay in office (ICRG risk rating).	PRS Group
77	SocCond	Socioeconomic conditions		An assessment of the socioeconomic pressures at work in society that could either constrain government action or fuel social dissatisfaction (ICRG risk rating).	PRS Group
78	IntConf	Internal conflict		An assessment of political violence in the country and its actual (or potential) impact on governance (ICRG risk rating).	PRS Group
79	ExtConf	External conflict		An assessment of both the risk to the incumbent government from foreign action; this ranges from non-violent external pressure (diplomatic pressures, withholding of aid, trade restrictions, territorial disputes, sanctions, etc.) to violent external pressure (cross-border conflicts to all-out war) (ICRG risk rating).	PRS Group

80	Corr	Corruption		An assessment of corruption within the political system (ICRG risk rating).	PRS Group
81	MilPol	Military in politics		An assessment of military involvement within politics (ICRG risk rating).	PRS Group
82	RelTen	Religious tensions		An assessment of religious tensions within a society (ICRG risk rating).	PRS Group
83	LawOrd	Law and order		A joint assessment of two components: the "Law" element, expressing the strength and impartiality of the legal system; and the "Order" element, reflecting the popular observance of the law (ICRG risk rating).	PRS Group
84	EthnTens	Ethnic tensions		An assessment of the degree of tension within a country—being attributable to racial, nationality, or language divisions (ICRG risk rating).	PRS Group
85	DemAcc	Democratic accountability		A measure of how responsive government is to its people; on the basis that the less responsive it is, the more likely it is that the government will fall either peacefully in a democratic society or potentially violently in a non-democratic one (ICRG risk rating).	PRS Group
86	BurQual	Bureaucracy quality		The institutional strength and quality of the bureaucracy, which helps to absorb shocks and minimize revisions of policy when governments change (ICRG risk rating).	PRS Group
87	Dem	Democracy index	Lehkonen and Heimonen (2015), Lei and Wisniewski (2018), Burnie (2021)	The Liberal Democracy Index by V-Dem, indicating to what extent the ideal of liberal democracy is achieved a t-1.	V-Dem
88	DemCh	Democratization	Miller (2021)	The 12-month change in the Liberal Democracy Index by V-Dem. The index indicates to what extent the ideal of liberal democracy is achieved a t-1.	V-Dem

Table A3. Counry-Level Asset Pricing Factors

The table presents the basic statistical properties of monthly returns on country-level asset pricing factors: market excess returns (MKT), small minus big (SMB), high minus low (HML), momentum (MOM), robust minus weak (RMW), and conservative minus aggressive (CMA). The cross-sectional facors SMB, HML, MOM, RMV, and CMA are based on country sorts on MV, BM, LtMom, ROE, and AG (see Table A2 for variable definitions). The long-short factor portflios take positions in extreme quintiles and use an equal- or value-weighting scheme (Panels A and B, respectively). Average, standard deviation, minimum, and maximum are all reported in percentages. The sample comprises 71 country stock markets and the study period is January 1995 to April 2021.

	MKT	SMB	HML	MOM	RMW	CMA
	Panel A: Equal-weighted factor portfolios					
Average	0.60	0.34	0.50	0.86	0.28	-0.26
St. deviation	4.48	3.94	3.64	4.64	3.68	3.58
Skewness	-0.76	0.12	0.29	-0.13	-0.40	0.25
Kurtosiss	2.08	1.53	0.26	0.57	1.60	0.95
Minimum	-20.75	-15.78	-9.05	-15.56	-14.02	-11.25
Maximum	12.53	15.33	11.04	13.27	10.48	13.61
	Panel B: Value-weighted factor portfolios					
Average	0.60	0.20	0.52	0.45	0.18	0.20
St. deviation	4.48	3.82	4.13	5.54	3.71	3.76
Skewness	-0.76	0.24	0.68	-0.36	-0.41	-0.09
Kurtosiss	2.08	1.19	1.75	1.02	1.49	0.29
Minimum	-20.75	-12.79	-10.55	-16.88	-15.77	-13.18
Maximum	12.53	15.29	19.43	17.59	13.25	9.99

Table A4. Bivariate Portfolio Sorts Mispricing and Return Predictions from Different Models

The table presents the monthly returns on portfolios from bivariate sorts on the mispricing score (*MISP*) and return predictions from different machine learning models described in Section 2.3. In the first step, we sort the markets into tertiles based on *MISP*. Subsequently—within each of these subsets—we sort portfolios into low, middle, and high tertiles (as inticated in the top row) based on the return predictions from models indicated in the top row. Panel A presents the returns on zero-investment strategies that buy (sell) the tertile of markets with the highest (lowest) predicted returns aross different *MISP* tertile. Panel B reports the differences in returns on the long-short strategies between the *Low* and *High MISP* tertiles and the middle one. All portfolios are equally-weighted and are rebalanced monthly. The returns and alphas are reported in percentages. The numbers in parentheses bootstrap t-statistics. The sample comprises 71 country stock markets and the testing period is from January 1995 to April 2021.

	OLS	PCA	PLS	LASSO	ENET	SVM	GBRT	RF	FFN1	FFN2	FFN3
	Panel A: Average returns on long-short machine learning portfolios										
Low MISP	1.27	0.90	1.19	0.89	0.88	0.87	0.83	1.19	1.36	1.21	1.18
	(5.33)	*(3.90)*	*(5.34)*	*(3.77)*	*(3.74)*	*(3.69)*	*(3.59)*	*(5.01)*	*(5.54)*	*(4.96)*	*(4.91)*
Mid MISP	0.33	0.16	0.11	0.32	0.33	0.35	0.51	0.69	0.50	0.27	0.23
	(1.89)	*(0.98)*	*(0.66)*	*(1.67)*	*(1.73)*	*(2.13)*	*(3.09)*	*(3.93)*	*(3.00)*	*(1.72)*	*(1.41)*
High MISP	0.90	1.14	0.96	0.95	0.94	0.89	0.73	0.93	1.00	1.13	1.13
	(3.35)	*(4.44)*	*(3.54)*	*(3.25)*	*(3.23)*	*(3.35)*	*(2.94)*	*(3.54)*	*(3.72)*	*(4.38)*	*(4.38)*
	Panel B: Differences-in-differences										
Low-Mid MISP	0.94	0.74	1.08	0.57	0.56	0.52	0.33	0.50	0.87	0.94	0.94
	(3.17)	*(2.68)*	*(4.17)*	*(1.92)*	*(1.87)*	*(1.84)*	*(1.22)*	*(1.90)*	*(2.98)*	*(3.18)*	*(3.47)*
High-Mid MISP	0.57	0.98	0.85	0.63	0.61	0.53	0.22	0.24	0.50	0.85	0.90
	(1.84)	*(3.41)*	*(2.77)*	*(1.91)*	*(1.87)*	*(1.79)*	*(0.72)*	*(0.78)*	*(1.64)*	*(2.89)*	*(3.04)*

B. Machine Learning Methods

In all the models described below, we define the excess return on a country equity index i at time t+1 as:

$$r_{i,t+1} = E_t(r_{i,t+1}) + \varepsilon_{i,t+1}, \tag{B1}$$

where $r_{i,t+1}$ denotes the excess return on index $i = 1, \dots, N_T$ in month $t = 1, \dots, T$. The expected excess returns are calculated using a constant function of predictor variables $z_{i,t}$ available at period t:

$$E_t(r_{i,t+1}) = g(z_{i,t}), \tag{B2}$$

where the P-dimensional vector $z_{i,t}$ contains market characteristics used to predict returns. The function $g(.)$ is flexible and changes across different machine learning models.

B.1. Linear Regression: OLS

The linear regression that is estimated using the ordinary least squares (OLS) is one of the simplest, yet prevalent, machine learning methods used in finance. The model assumes that the conditional expectation of returns on security i at time t can be approximated by the following linear function:

$$g(z_{i,t}; \theta) = z'_{i,t}\theta. \tag{B3}$$

The model parameters' vector θ can be conveniently estimated using OLS by minimizing the loss function:

$$L(\theta) = \tfrac{1}{TN} \textstyle\sum_{i=1}^{N} \sum_{t=1}^{T} [r_{i,t+1} - g(z_{i,t}; \theta)]^2. \tag{B4}$$

Importantly, the cross-sectional OLS regressions do not rely on any hyperparameters and—thus—do not require the sample splitting into training and validation periods. As indicated in Wooldridge (2001) and Gu, Kelly, and Xiu (2020), the parameter estimates in (B4) are efficient and unbiased if the number of predictors is small relative to the number of time observations. Nonetheless, in real-life machine learning problems, the number of covariates is substantial; this leads to the overfitting, invalidating efficiency, or consistency of OLS estimates. The subsequent models described in this appendix represent literature solutions to cope with these problems.

B.2. Dimension Reduction Techniques: PCA, PLS

The number of covariates in the vector $z_{i,t}$ in Equation (B2) is typically high, leading to a risk of overfitting. A potentially effective solution to this problem may be reducing the number of market characteristics to a smaller quantity of factors. In this regard, we consider two popular techniques: the principal component analysis (PCA) and the partial least squares (PLS).

To begin with the PCA, this technique assumes the transformation of a set of return predictors into a smaller number of orthogonal principal components. These new decorrelated predictors are designed to have maximum possible variance and, hence, explanatory power over the initial set of predictors. Once the optimal number of the few leading components is identified using a validation procedure, they are used as new variables in order to predict the cross-section of index returns.

A potential deficiency of the PCA method is that the leading principal components aim to maximize the common variation across the characteristics, disregarding their association with future returns. Consequently, while this approach effectively reduces the number of dimensions and handles overfitting, there is no guarantee that the newly created variables will contain substantial information about stock performance. Theoretically, the components may be dominated by covariates that efficiently explain the set of considered features but have minor predictive abilities.

In contrast, the PLS attenuates the drawbacks of PCA by directly extracting the strongest signals based on their links with stock returns. The covariation between the predictors and asset returns is exploited via a model-averaging approach. Specifically, PLS seeks to find linear combinations of the considered predictors; this is so that the newly created components maximize their correlation with future cross-sectional returns.

In practice, the first PLS component is created by running univariate regressions of realized returns on individual market characteristics. The resulting coefficients can be regarded as measures of "partial" sensitivity of equity index returns to each variable. Then, the first component is formed by weighting the predictors based on their coefficients; this is so that the higher (lower) weights are associated with the stronger (weaker) predictors. The subsequent components are formed using a similar procedure; however, the predictors are initially orthogonalized with respect to the already created component(s). Similarly, as the PLS, the PCA method has only one tuning parameter optimized via validation: the number of components employed in the predictive regressions.

B.3. Penalized Linear Regressions: LASSO, ENET

The regularization of linear regression is a common approach to coping with overfitting problems. A standard method is to include a penalty term to the objective function. Popular regularized models include the least absolute shrinkage and selection operator (LASSO), as well as the Ridge regressions. Furthermore, the elastic net (ENET), employed in finance firstly by Rapach et al. (2013), is a convex combination of these two methods. In our study, we follow Leippold, Wang, and Zhou (2021) and include LASSO and ENET among the considered prediction models. Both methods have an identical specification as the simple OLS regression. The main difference lies in the structure of the loss function, which includes an additional penalty term $\phi(\theta;.)$. The precise functional form of this penalty term may differ; furthermore, the elements of the coefficients vector θ may be shrunk towards zero and regularized.

The penalty function for LASSO takes the following form:

$$\phi^{LASSO}(\theta;.) = \lambda \sum_{j=i}^{P} |\theta_j|, \tag{B5}$$

where $\lambda>0$ is the hyperparameter determining the magnitude of shrinkage; i.e., the size of the penalty. We employ a standard regularization approach relying on a geometric series of λ values with the largest on giving a null model (all coefficients are zero). Subsequently, we select and use the λ parameter that generates the lowest MSE in the validation sample.

The penalty function for ENET, in turn, is given by:

$$\phi^{ENET}(\theta;.) = (1-\alpha)\lambda \sum_{j=i}^{P} |\theta_j| + \frac{1}{2}\alpha\lambda \sum_{j=i}^{P} \theta_j^2, \tag{B6}$$

where $\lambda>0$ plays an identical role as in Equation (B5), and α determines the relative weight between the two penalty components. The λ hyperparameter is determined following the same approach as for LASSO, and $\alpha=0.5$. One advantage of ENET, relative to LASSO, is that it copes more effectively with the correlation between covariates (see Zou and Hastie [2005], as well as Diebold and Shin [2019], for details). Following the convention in the literature, we do not shrink the intercept in both models.

B.4. Support Vector Machine: SVM

Support Vector Machine (SVM) seeks hyperplanes to territorially split the multidimensional vector space into groups belonging to similar classes. In the asset pricing context, the vector space comprises the stock-level return predictors. The hyperplanes are located in areas of closely neighboring vectors. The SVM algorithm typically concentrates on the immediate neighbors of the potential hyperplanes, which are called “support vectors.” This operation aims at increasing the computation speed of the algorithm. In its optimization procedure, SVM targets to specify the optimal hyperplanes by means of minimizing the number of misclassified support vectors and maximizing the distance between the correctly classified ones.

SVM may be used for both binary and multi-class problems. In the latter case, the optimal fit is searched via pairwise class comparisons. SVM is strongly regularized to avoid overfitting. In a high-dimensional space, SVM may be used both as a classification and regression method—though the last method attracts less attention in finance. In such a case, it is sometimes called Support Vector Regression and may be directly employed to predict cross-sectional returns. We estimate the model parameters using the average stochastic gradient descent (ASGD) algorithm of Xu (2011).

B.5 Tree Models: RF, GBRT

Tree models are both flexible and non-parametric machine learning techniques to handle both classification and regression problems. We employ two methods from this domain: random forests (RF) and gradient boosted regression trees (GBRT). Both techniques can be regarded as ensemble methods, as they build on a number of individual “trees.”

The tree methods partition observation into multiple subgroups, typically named as “leaves.” A tree is built in a sequence of steps; furthermore, its structure is determined by the decision nodes and splitting variables. A splitting variable produces two disjoint branches at each split point. The tree grows subsequent sets of branches until the terminal nodes “leaves” are reached. In asset pricing practice, the final product is returns clustered by predictors.

Formally, a simple tree with a depth of L and K leaves can be described by the following equation:

$$g(z_{i,t};\theta,K,L) = \sum_{k}^{K} \theta_K 1_{\{z_{i,t}\in C_K(L)\}}, \tag{B6}$$

where L indicates the depth measured with the largest number of nodes in a complete branch, $C_K(L)$ denotes the k-th partition of the variables, θ_K is the sample average of the outcomes within the partition, and $1_{\{.\}}$ is the indicator function. If an index i with a set of return predictive variables zi,t is clustered into the k-th leaf, then θ_K will indicate the return prediction. The tree models offer substantial flexibility in terms of both split points and variable selection (see James et al. [2013] for review). Simple trees are prone to overfitting, so they are required to be heavily regularized. The RF and GBRT used in this study belong to the most popular regularization techniques.

RF relies on the bootstrap aggregation algorithm of Breiman et al. (2001), usually named "bagging." This method builds on average multiple trees to reduce the forecast variation. To be specific, the model uses bootstrapped samples of the original data to train a certain number of trees and uses random subsets of variables to grow the branches. The averaged outcome of these de-correlated trees reduces the overfit and, therefore, results in more stable predictions. Our models assume 30 trees with a minimum leaf size of five. The number of features equals 30.

The GBRT algorithm has a different structure and aims at producing a "strong learner" from a combination of weak learners. Assume a GBRT model with only two trees. The first of them is formed to fit equity returns to market characteristics. Subsequently, the second tree (of identical depth) is constructed to fit the residuals from the first tree. The ensemble forecast of this simple GBRT model is calculated as the prediction of the first tree plus the second tree's prediction multiplied by the learning rate (0,1). The subsequent trees can be formed using the same procedure: fitting the residuals from the already grown trees and multiplying them times the learning rate. We fit the GBRT model using the least-squares boosting algorithm (Breinan, 2001; Hastie, Tibshirani, & Friedman, 2008). We assume up to 100 learning cycles.

B.6. Neural Networks: FFN1, FFN2, FFN3

Neural networks can effectively approximate nonlinear functions, as well as account for interactions between predictors. Therefore, they attracted much attention in different fields—not just limited to finance. In our study, we employ feed-forward neural networks. A typical structure of such a network comprises an "input layer" with the input variables (return predictors); several "hidden" layers, which contain activation functions and transform the predictors; and an "output layer," transforming the outcomes from hidden layers into the final return predictions. The more hidden layers are included in the model, the more flexible it becomes. The information flows from the input layers through the hidden layers to be aggregated into forecasts through the output layers.

We consider three different neural networks with one, two, or three hidden layers; these are denoted as FFN1, FFN2, and FFN3—respectively. The respective layers include eight, four, and two neurons—similar to Gu, Kelly, and Xiu (2020) or Leippold, Wang, and Zhou (2021). Each of them takes the result from the previous layer and forges it into output.

Our implementation of neural networks generally follows Gu et al. (2020). The neurons may include many different activation functions and we rely on a rectified linear unit, which is defined as $\sigma(x) - \max(0, x)$. To train the model, we follow Da Nard, Hediger, and Leippold (2020) and employ the Adam optimization algorithm of Kingma and Ba

(2014); with default parameters with the maximum number of epochs amounting to 1000, learning rate equals 0.01, and its increase of 1.05.

B.7. Forecast Combination: COMB

The forecasts combination assumes merging multiple predictions from different models. The underlying reasoning is associated with the concept that forecasts from individual models may have high variance. Hence, combining them may reduce the overall variance and—thus—decrease the prediction error. The overall effect tends to improve the accuracy of return predictability (Rapach et al., 2010; Chen et al., 2019). In our study, we follow Bali et al. (2021) and calculate the COMB forecasts as the simple equal-weighted average of all 11 individual models that are considered: OLS, PCA, PLS, LASSO, ENET, SVM, RF, GBRT, FFN1, FFN2, and FFN3.

References

Aggarwal, R., & Goodell, J. W. (2008). Equity premia in emerging markets: National characteristics as determinants. *Journal of Multinational Financial Management*, *18* (4), 389-404.

Amaya, D., Christoffersen, P., Jacobs, K., & Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, 118(1), 135- 167.

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time series effects. *Journal of Financial Markets*, 5(1), 31–56.

Andrade, S. C., & Chhaochharia, V. (2018). The costs of sovereign default: Evidence from the stock market. *The Review of Financial Studies*, *31*(5), 1707-1751.

Ang, A., & Maddaloni, A. (2005). Do demographic changes affect risk premiums? Evidence from international data. *Journal of Business*, *78*(1), 341-380.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, *61*(1), 259-299.

Angelidis, T., & Tessaromatis, N. (2017). Global equity country allocation: An application of factor investing. *Financial Analysts Journal*, *73*(4), 55-73.

Arnott, R. D., & Chaves, D. B. (2012). Demographic changes, financial markets, and the economy. *Financial Analysts Journal*, *68*(1), 23-46.

Asness, C. S., Liew, J. M., & Stevens, R. L. (1997). Parallels between the cross-sectional predictability of stock and country returns. *Journal of Portfolio Management*, *23*(3), 79.

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, *68*(3), 929-985.

Asness, C., Frazzini, A., Gormsen, N. J., & Pedersen, L. H. (2020). Betting against correlation: Testing theories of the low-risk effect. *Journal of Financial Economics*, *135*(3), 629-652.

Atanasov, V. (2021). Unemployment and aggregate stock returns. *Journal of Banking & Finance*, *129*, 106159.

Atilgan, Y., Bali, T. G., Demirtas, K. O., & Gunaydin, A. D. (2019). Global downside risk and equity returns. *Journal of International Money and Finance*, 98, 102065.

Avramov, D., Chordia, T., Jostova, G., & Philipov, A. (2012). The world price of credit risk. *Review of Asset Pricing Studies*, 2(2), 112-152.

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

Baetje, F., & Menkhoff, L. (2016). Equity premium prediction: Are economic and technical indicators unstable? *International Journal of Forecasting*, *32*(4), 1193-1207.

Baghdadabad, M. T., & Mallik, G. (2018). Global idiosyncratic risk moments. *Empirical Economics*, *55*(2), 731-764.

Baker, M., & Wurgler, J. (2000). The equity share in new issues and aggregate stock returns. *Journal of Finance*, *55*(5), 2219-2257.

Baker, M., Bradley, B., & Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analyst Journal*, 67(1), 40-54.

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

Bali, T.G., & Cakici, N. (2010). World market risk, country-specific risk and expected returns in international stock markets. *Journal of Banking & Finance*, *34*(6), 1152-1165.

Bali, T., Goyal, A., Huang, D., Jiang, F., & Wen, Q. (2021). Different strokes: Return predictability across stocks and bonds with machine learning and big data. Georgetown McDonough School of Business Research Paper No. 3686164. Swiss Finance Institute Research Paper No. 20-110. Available at SSRN: https://ssrn.com/abstract=3686164.

Bali, T.G. & Cakici, N. (2004). Value at risk and expected stock returns. *Financial Analyst Journal*, 60(2), 57-73.

Bali, T.G., Engle, R.F., & Murray, S. (2016). *Empirical Asset Pricing: The Cross Section of Stock Returns.* Wiley, Hoboken.

Ball, R., Gerakos, J., Linnainmaa, J. T., & Nikolaev, V. (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics*, *121*(1), 28-45.

Baltas, N., & Salinas, G. (2019). Cross-asset skew. Available at SSRN 3505422.

Baltussen, G., Swinkels, L., & Van Vliet, P. (2021). Global factor premiums. *Journal of Financial Economics*, *142*(3), 1128-1154.

Baltussen, G., van Bekkum, S., & Da, Z. (2019). Indexing and stock market serial dependence around the world. *Journal of Financial Economics*, *132*(1), 26-48.

Balvers, R. J., & Wu, Y. (2006). Momentum and mean reversion across national equity markets. *Journal of Empirical Finance*, *13*(1), 24-48.

Balvers, R., Wu, Y., & Gilliland, E. (2000). Mean reversion across national stock markets and parametric contrarian investment strategies. *Journal of Finance*, *55*(2), 745-772.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9, 3–18.

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., Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1997). The cross-sectional determinants of emerging equity market returns. *Quantitative Investing for the Global Markets*, Ch. 9, 221-272.

Belo, F., Lin, X., & Bazdresch, S. (2014). Labor hiring, investment, and stock return predictability in the cross section. *Journal of Political Economy*, 122, 129–177.

Bhandari, L.C. (1998). Debt/equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance*, 43(2), 507–528.

Bhojraj, S., & Swaminathan, B. (2006). Macromomentum: Returns predictability in international equity indices. *Journal of Business*, *79*(1), 429-451.

Bilson, C. M., Brailsford, T. J., & Hooper, V. C. (2002). The explanatory power of political risk in emerging markets. *International Review of Financial Analysis*, *11*(1), 1-27.

Blau, B. M., & Whitby, R. J. (2017). Range-based volatility, expected stock returns, and the low volatility anomaly. *Plos one*, *12*(11), e0188517.

Blitz, D., Hanauer, M. X., & Vidojevic, M. (2020). The idiosyncratic momentum anomaly. *International Review of Economics & Finance*, *69*, 932-957.

Blitz, D., Huij, J., & Martens, M. (2011). Residual momentum. *Journal of Empirical Finance*, *18*(3), 506-521.

Boudoukh, J., Michaely, R., Richardson, M., & Roberts, M. R. (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. *Journal of Finance*, 62(2), 877-915.

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

Brennan, M. J., Chordia, T., & Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics*, 49(3), 345-373.

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

Brooks, J. (2017). A Half Century of Macro Momentum. *White paper, AQR*. Available at http://www.smallake.kr/wp-content/uploads/2017/10/A-Half-Century-of-Macro-Momentum.pdf.

Brunetti, M., & Torricelli, C. (2010). Demographics and asset returns: does the dynamics of population ageing matter?. *Annals of Finance*, *6*(2), 193-219.

Bunn, O., Staal, A., Zhuang, J., Lazanas, A., Ural, C., & Shiller, R. (2014). Es-cape-ing from overvalued sectors: Sector selection based on the cyclically adjusted price-earnings (CAPE) ratio. *Journal of Portfolio Management*, 41(1), 16-33.

Burnie, D. A. (2021). Democracy, dictatorship, and economic freedom signals in stock market. *International Journal of Finance & Economics*, *26*(1), 375-390.

Cakici, N., Chatterjee, S., Tang, Y., & Tong, L. (2021). Alternative profitability measures and cross-section of expected stock returns: international evidence. *Review of Quantitative Finance and Accounting*, 56, 369-391.

Calice, G., & Lin, M. T. (2021). Exploring risk premium factors for country equity returns. *Journal of Empirical Finance*, *63*, 294-322.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of financial economics*, *18*(2), 373-399.

Campbell, J. Y., & Shiller, R.J. (1998). Valuation ratios and the long-run stock market outlook. *Journal of Portfolio Management*, 24(6), 11-26.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, *21*(4), 1509-1531.

Chan, K., Hameed, A., & Tong, W. (2000). Profitability of momentum strategies in the international equity markets. *Journal of Financial and Quantitative Analysis*, *35*(2), 153-172.

Chen, L., Pelger, M., & Zhu, J. (2020). Deep learning in asset pricing. Available at SSRN: https://ssrn.com/abstract=3350138.

Chen, N. F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, 383-403.

Chen, Y., Eaton, G. W., & Paye, B. S. (2018). Micro (structure) before macro? The predictive power of aggregate illiquidity for stock returns and economic activity. *Journal of Financial Economics*, *130*(1), 48-73.

Chordia, T., A. Subrahmanyam, R. Anshuman. (2001). Trading activity and expected stock returns. *Journal of Financial Economics* 59, 3–32.

Clare, A., Seaton, J., Smith, P. N., & Thomas, S. (2016). The trend is our friend: Risk parity, momentum and trend following in global asset allocation. *Journal of Behavioral and Experimental Finance*, *9*, 63-80.

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

Cornell, B. (2012). Demographics, GDP, and future stock returns: The implications of some basic principles. *Journal of Portfolio Management*, *38*(4), 96-99.

Daniel, K. & Titman, S. (2006). Market reactions to tangible and intangible information. *Journal of Finance*, 61, 1605-1643.

Datar, V., Naik, N., & Radcliffe, R. (1998). Liquidity and stock returns: an alternative test. *Journal of Financial Markets*, 1, 203 − 220.

De Nard, G., Hediger, S., & Leippold, M. (2020). Subsampled factor models for asset pricing: The rise of VASA. Available at SSRN: https://ssrn.com/abstract=3557957 or http://dx.doi.org/10.2139/ssrn.3557957.

DeBondt, W.F.M. & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, *40*(3), 793-805.

Diamonte, R. L., Liew, J. M., & Stevens, R. L. (1996). Political risk in emerging and developed markets. *Financial Analysts Journal*, *52*(3), 71-76.

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

Dimic, N., Orlov, V., & Piljak, V. (2015). The political risk factor in emerging, frontier, and developed stock markets. *Finance Research Letters*, *15*, 239-245.

Du, D. (2008). The 52-week high and momentum investing in international stock indexes. *Quarterly Review of Economics and Finance*, *48*(1), 61-77.

Elgers, P. T., Lo, M. H., & Pfeiffer Jr, R. J. (2001). Delayed security price adjustments to financial analysts' forecasts of annual earnings. *Accounting Review*, *76*(4), 613-632.

Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1995a). Country risk and global equity selection. *Journal of Portfolio Management*, *21*(2), 74-83.

Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1995b). Inflation and world equity selection. *Financial Analysts Journal*, *51*(6), 28-42.

Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1996a). Expected returns and volatility in 135 countries. *Journal of Portfolio Management*, 2, 46-58.

Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1996b). Political risk, economic risk, and financial risk. *Financial Analysts Journal*, *52*(6), 29-46.

Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1997). *Country risk in global financial management*. Cfa Inst.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of financial economics*, *25*(1), 23-49.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607-636.

Fama, E.F., & French, K.R. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*(2), 427-465.

Fama, E.F., & French, K.R. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, 51(1), 55-84.

Fang, J., Qin, Y., & Jacobsen, B. (2014). Technical market indicators: An overview. *Journal of Behavioral and Experimental Finance*, *4*, 25-56.

Ferson, W. E., & Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of Political Economy*, *99*(2), 385-415.

Ferson, W. E., & Harvey, C. R. (1994). Sources of risk and expected returns in global equity markets. *Journal of Banking & Finance*, *18*(4), 775-803.

Fisher, G. S., Shah, R., & Titman, S. (2017). Should you tilt your equity portfolio to smaller countries? *Journal of Portfolio Management*, *44*(1), 127-141.

Flannery, M. J., & Protopapadakis, A. A. (2002). Macroeconomic factors do influence aggregate stock returns. *Review of Financial Studies*, *15*(3), 751-782.

Francis, J., LaFond, R., Olsson, P.M., & Schipper, K. (2004). Costs of equity and earnings attributes. *Accounting Review*, 79(4), 967-1010.

Frazzini, A., & Pedersen, L.H. (2014). Betting against beta. *Journal of Financial Economics*, 111, 1-25.

Gao, G. P., Lu, X., & Song, Z. (2019). Tail risk concerns everywhere. *Management Science*, 65(7), 3111-3130.

Geanakoplos, J., Magill, M., & Quinzii, M. (2004). Demography and the long-run predictability of the stock market. *Brookings Papers on Economic Activity*, *2004*(1), 241-325.

Geczy, C. C., & Samonov, M. (2016). Two centuries of price-return momentum. *Financial Analysts Journal*, *72*(5), 32-56.

George, T. J., & Hwang, C. Y. (2004). The 52-week high and momentum investing. *Journal of Finance*, *59*(5), 2145-2176.

Goyal, A. (2004). Demographics, stock market flows, and stock returns. *Journal of Financial and Quantitative Analysis*, *39*(1), 115-142.

Goyal, A., Welch, I., & Zafirov, A. (2021). A Comprehensive Look at the Empirical Performance of equity Premium Prediction II. *Available at SSRN 3929119*.

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

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

Harvey, C. R. (2004). Country risk components, the cost of capital, and returns in emerging markets. *Available at SSRN 620710*.

Harvey, C. R., & Ferson, W. (2008). *2. An Exploratory Investigation of the Fundamental Determinants of National Equity Market Returns* (pp. 59-148). University of Chicago Press.

Harvey, C.R. & Siddique, A. (2000). Conditional skewness in asset pricing tests. *Journal of Finance*, 55(3), 1263-1295.

Harvey, C.R. (2000). The drivers of expected returns in international markets. Available at SSRN 795385.

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

Haugen, R. A., & Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, 41(3), 401-439.

Heckman, L., Mullin, J. J., & Sze, H. (1996). Valuation ratios and cross-country equity allocation. *Journal of Investing*, *5*(2), 54-63.

Heston, S.L., & Sadka, R. (2008). Seasonality in the cross-section of stock returns. *Journal of Financial Economics*, 87, 418-445.

Hjalmarsson, E. (2010). Predicting global stock returns. *Journal of Financial and Quantitative Analysis*, *45*(1), 49-80.

Hollstein, F., Nguyen, D. B. B., Prokopczuk, M., & Simen, C. W. (2019). International tail risk and world fear. *Journal of International Money and Finance*, 93, 244-259.

Hollstein, F., Prokopczuk, M., Tharann, B., & Wese Simen, C. (2020). Predicting the equity premium around the globe: Comprehensive evidence from a large sample. *Available at SSRN 3567622*.

Hsu, P. H., Hsu, Y. C., & Kuan, C. M. (2010). Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias. *Journal of Empirical Finance*, 17(3), 471-484.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48, 65–91.

Keloharju, M., Linnainmaa, J. T., & Nyberg, P. (2021). Are return seasonalities due to risk or mispricing? *Journal of Financial Economics*, *139*(1), 138-161.

Keloharju, M., Linnainmaa, J.T., & Nyberg, P. (2016). Return seasonalities. *Journal of* Finance, 71(4), 1557-1589.

Keppler, A. M., & Traub, H. D. (1993). The small-country effect: Small markets beat large markets. *Journal of Investing*, *2*(3), 17-24.

Kim, D. (2012). Value premium across countries. *Journal of Portfolio Management*, *38*(4), 75-86.

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

Kogan, L., & Papanikolaou, D. (2013). Firm characteristics and stock returns: the role of investment-specific shocks. *Review of Financial Studies*, 25, 2718-2759.

Lakonishok, J., Shleifer, A., & Vishny, R.W. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, 49, 1541-1578.

Lamont, O. (1998). Earnings and expected returns. *Journal of Finance*, 53(5), 1563-1587.

Lawrenz, J., & Zorn, J. (2017). Predicting international stock returns with conditional price-to-fundamental ratios. *Journal of Empirical Finance*, *43*, 159-184.

Lee, K. H. (2011). The world price of liquidity risk. *Journal of Financial Economics*, *99*(1), 136-161.

Lehkonen, H., & Heimonen, K. (2015). Democracy, political risks and stock market performance. *Journal of International Money and Finance*, *59*, 77-99.

Lei, X., & Wisniewski, T. P. (2018). Democracy and stock market returns. *Available at SSRN 3198561*.

Leippold, M., Wang, Q., & Zhou, W. (2021). Machine learning in the Chinese stock market. *Journal of Financial Economics*, in press.

Liang, S. X., & John Wei, K. C. (2020). Market volatility risk and stock returns around the world: Implication for multinational corporations. *International Review of Finance*, *20*(4), 923-959.

Liang, S. X., & Wei, J. K. (2012). Liquidity risk and stock returns around the world. *Journal of Banking & Finance*, *36*(12), 3274-3288.

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(2), 163-195.

Loughran, T., & Wellman, J.W. (2011). New evidence on the relation between the enterprise multiple and average stock returns. *Journal of Financial and Quantitative Analysis*, 46, 1629-1650.

Macedo, R. (1995). Value, relative strength, and volatility in global equity country selection. *Financial Analysts Journal*, *51*(2), 70-78.

Malin, M., & Bornholt, G. (2010). Predictability of future index returns based on the 52-week high strategy. *The Quarterly Review of Economics and Finance*, *50*(4), 501-508.

Medhat, M., & Schmeling, M. (2021). Short-term Momentum. *Review of Financial Studies*, in press.

Miller, M. (2021). Democratization, inequality, and risk premia. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*. Available at SSRN: https://ssrn.com/abstract=3488208 or http://dx.doi.org/10.2139/ssrn.3488208.

Møller, S. V., & Rangvid, J. (2015). End-of-the-year economic growth and time-varying expected returns. *Journal of Financial Economics*, *115*(1), 136-154.

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

Penman, S. H., Richardson, S. A., & Tuna, I. (2007). The book-to-price effect in stock returns: accounting for leverage. *Journal of Accounting Research*, 45(2), 427-467.

Pitkäjärvi, A., Suominen, M., & Vaittinen, L. (2020). Cross-asset signals and time series momentum. *Journal of Financial Economics*, *136*(1), 63-85.

Pontiff, J., & Woodgate, A. (2008). Share issuance and cross-sectional returns. *Journal of Finance*, 63(2), 921-945.

Qi, M., & Zhao, X. (2008). Market Breadth, Trin Statistic, and Market Returns. *Journal of Investing*, *17*(1), 65-73.

Radha, S. S. (2020). Using CAPE to forecast country returns for designing an international country rotation portfolio. *Journal of Portfolio Management*, *46*(7), 101-117.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2013). International stock return predictability: What is the role of the United States? *Journal of Finance*, 68(4), 1633-1662.

Rapach, D. E., Wohar, M. E., & Rangvid, J. (2005). Macro variables and international stock return predictability. *International Journal of Forecasting*, *21*(1), 137-166.

Rapach, D., & Zhou, G. (2013). Forecasting stock returns. In *Handbook of economic forecasting* (Vol. 2, pp. 328-383). Elsevier.

Rapach, D.E., Strauss, J.K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, *23*(2), 821-862.

Richards, A. J. (1997). Winner-loser reversals in national stock market indices: Can they be explained? *Journal of Finance*, *52*(5), 2129-2144.

Rosenberg, B., Reid., K, & Lanstein, R. (1985). Persuasive evidence of market inefficiency. *Journal of Portfolio Management*, *11*, 9-17.

Sermpinis, G., Hassanniakalager, A., Stasinakis, C., & Psaradellis, I. (2021). Technical analysis profitability and Persistence: A discrete false discovery approach on MSCI indices. *Journal of International Financial Markets, Institutions and Money*, *73*, 101353.

Siegel, J. J. (2016). The Shiller CAPE ratio: A new look. *Financial Analysts Journal*, 72(3), 41-50.

Soliman, M.T. (2008). The use of DuPont analysis by market participants. *Accounting Review*, 83(3), 823–853.

Spierdijk, L., Bikker, J. A., & Van den Hoek, P. (2012). Mean reversion in international stock markets: An empirical analysis of the 20th century. *Journal of International Money and Finance*, *31*(2), 228-249.

Suleman, T., Gupta, R., & Balcilar, M. (2017). Does country risks predict stock returns and volatility? Evidence from a nonparametric approach. *Research in International Business and Finance*, *42*, 1173-1195.

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

Umutlu, M. (2015). Idiosyncratic volatility and expected returns at the global level. *Financial Analysts Journal*, *71*(6), 58-71.

Vortelinos, D. I., & Saha, S. (2016). The impact of political risk on return, volatility and discontinuity: Evidence from the international stock and foreign exchange markets. *Finance Research Letters*, *17*, 222-226.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455-1508.

Wen, Q. (2019). Asset growth and stock market returns: A time-series analysis. *Review of Finance*, *23*(3), 599-628.

Wisniewski, T. P., & Jackson, P. M. (2021). Government debt expansion and stock returns. *International Journal of Finance & Economics*, *26*(4), 5017-5030.

Wooldridge, J.M. (2001). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Xu, W. (2011). Towards optimal one pass large scale learning with averaged stochastic gradient descent. arXiv preprint arXiv:1107.2490.

Zaremba, A., & Andreu, L. (2018). Paper profits or real money? Trading costs and stock market anomalies in country ETFs. *International Review of Financial Analysis*, *56*, 181-192.

Zaremba, A., Cakici, N., Bianchi, R.J., and Long, H. (2021b). Yield curve shifts and the cross-section of global equity returns. Available at SSRN: https://ssrn.com/abstract=3756047 or http://dx.doi.org/10.2139/ssrn.3756047.

Zaremba, A., Szyszka, A., Karathanasopoulos, A., & Mikutowski, M. (2021a). Herding for profits: Market breadth and the cross-section of global equity returns. *Economic Modelling*, *97*, 348-364.

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

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