Alpha Go Everywhere: Machine Learning and International Stock Returns*

Darwin Choi
CUHK Business School

Wenxi Jiang
CUHK Business School

Chao Zhang
University of Oxford

March 2023

Abstract

We apply machine learning techniques to predict international stock returns using firm characteristics. Market-specific features are important, as neural network models (NNs) achieve stronger results when they are trained in each market separately than in a global model trained with U.S. data. NNs outperform linear models in predicting stock return rankings and forming profitable portfolios. In contrast, regression trees underperform linear models when the number of observations is low. We also show that adding foreign variables constructed from U.S. firm characteristics further enhances the return predictability of market-specific NNs, consistent with the notion that the markets are partially integrated.

JEL Codes: C52, G10, G12, G15

Keywords: International Asset Pricing, Cross-section of Stock Returns, Market Integration, Neural Networks, Regression Trees

*Darwin Choi, Department of Finance, CUHK Business School, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, dchoi@cuhk.edu.hk. Wenxi Jiang, Department of Finance, CUHK Business School, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, wenxijiang@baf.cuhk.edu.hk. Chao Zhang, Department of Statistics, University of Oxford, U.K., chao.zhang@stats.ox.ac.uk. We thank Arka Bandyopadhyay, Eugene Chow, Gavin Feng, Stefano Giglio, Jiantao Huang, Raymond Kan, Amy Kwan, Andrew Karolyi, Ravi Sastry, Bruno Solnik, Dacheng Xiu, and seminar participants at 2022 Asian Finance Association Annual Conference; CUHK Conference on Financial Technology; CUHK Derivatives and Quantitative Investing Conference; 2022 Hong Kong Conference for Fintech, AI, and Big Data in Business; University of Florida Research Conference on Machine Learning in Finance; Chinese University of Hong Kong; Chinese University of Hong Kong, Shenzhen; Hong Kong University of Science and Technology; Peking University HSBC Business School; and University of Melbourne for helpful comments. We thank Andrew Karolyi and Ying Wu for sharing their data on international factors. Hulai Zhang provided excellent research assistance. First draft: November 2019

1 Introduction

Variables related to firm characteristics, such as size, value, and momentum, predict international stock returns (Fama and French (1998, 2012, 2017); Hou, Karolyi, and Kho (2011); Rouwenhorst (1998)). The international finance literature also finds that global equity markets are partially, but not fully, integrated (see Karolyi and Stulz (2003) and Lewis (2011) for a review). The lack of full integration suggests that the predictive power of different firm characteristics varies across markets. At the same time, when the markets are not totally segmented, information obtained from the firm characteristics in a foreign market may be relevant for local stocks.

How do we form predictions in many different markets without pre-specifying the models? How can foreign firm characteristics improve the return predictability of local models? Our paper examines these issues using machine learning techniques and attempts to increase the economic profits earned by international investors. Constructing a market-specific asset pricing model typically requires the knowledge of institutional details (say, the price limit rules in China), but machine learning is capable of detecting non-monotonic relationships and complex interactions between returns and many characteristics even *without* such knowledge. Machine learning also allows us to explore various foreign variables and their interactions with local variables, which may not be ideal in a linear setting due to its inability to capture complex relationships.

Most studies that use machine learning to study the relationship between stock returns and firm-level variables focus on the U.S.[1] We first verify that the machine learning methods can be applied in a large number of international markets. Then we show that market-specific models, which are trained separately for each market and may capture market-specific return-characteristic relationships, outperform a global model trained using U.S. data. Based on a

[1] See, for example, Avramov, Cheng, and Metzker (2022); Feng, Polson, and Xu (2022); Feng and He (2021); Gu, Kelly, and Xiu (2020, 2021); Freyberger, Neuhierl, and Weber (2020); Rapach and Zhou (2020); Chen, Pelger, and Zhu (2019); Chinco, Clark-Joseph, and Ye (2019); Han, He, Rapach, and Zhou (2018). Karolyi and Van Nieuwerburgh (2020) highlight the risk of overfitting the data by these complex algorithms and that there is "only one out-of-sample sample" (from the U.S.).

machine-learning measure of similarity that we introduce, we find that the outperformance is larger when the international market is less similar to the U.S. We also find that market-specific machine learning models can be further enhanced by adding variables constructed from the distribution of U.S. firm characteristics.[2] The markets in which the U.S. variables are more relevant are also those classified as more integrated based on well-accepted measures.

Our analysis emphasizes non-linearity. As in Gu, Kelly, and Xiu (2020) (GKX), we compare the performance of linear and non-linear models. We examine linear OLS models and their variants: OLS with a Huber loss function that makes it less sensitive to outliers; LASSO, which selects a subset of predictors; and RIDGE, which restricts the magnitude of the regression coefficients. We study two classes of non-linear models—regression trees (RTs) and neural network (NN) models (with 1 to 5 hidden layers). Trees are a non-parametric method for classifications and regressions that predicts the value of a target variable by learning simple decision rules inferred from the data features. NN models aggregate and transform input signals into outputs, allowing for multiple layers of transformation and therefore complex interactions among the predictors. In each test, we set aside training and validation periods to train our models and select the hyperparameters, and then use the models to construct forecasts of one-month-ahead stock returns (denominated in U.S. dollars and in excess of the corresponding market return) in the testing period, which does not overlap the other two periods.

While the literature identifies a long list of characteristics that seem to predict returns, data availability is lower internationally and we trim down the list of explanatory variables to 36.[3] These 36 variables include the most accessible stock characteristics such as past returns, market capitalization, trading volume, past returns of the industry, and accounting information. The first set of analysis, in which we train and validate our models using only past

[2]As Rapach, Strauss, and Zhou (2013) argue, the U.S. is a large trading partner for many markets, and its stock market is the world's largest and is relevant for other economies. They show that lagged U.S. market returns can predict the index returns in other markets.

[3]Harvey, Liu, and Zhu (2016) count that 316 factors have been proposed by 313 papers. McLean and Pontiff (2016) examine 97 characteristics in finance, accounting, and economics journals. Hou, Xue, and Zhang (2020) compile a list of 452 variables.

U.S. data, is a stringent test for the machine learning methods. We follow the methodologies of GKX and use their set of potential hyperparameter values. We use the hyperparameter and parameter values estimated from the U.S. to form the predictions in *all* 32 markets (31 international markets plus the U.S.).[4] Using 94 characteristics, 8 macroeconomic predictors, and 74 industry dummies (i.e., a total number of 920 ($= 94 \times (8 + 1) + 74$) covariates), GKX conclude that both RTs and NN models outperform linear models in terms of out-of-sample R^2 and long-short portfolio (top- minus bottom-decile of predicted returns) Sharpe Ratios. With 36 covariates, the R^2 and Sharpe Ratios of our RTs and NN models in the U.S. are comparable to those in GKX, consistent with recent evidence that a modest number of factors can explain cross-sectional U.S. stock returns (Feng, Giglio, and Xiu (2020); Freyberger et al. (2020); Kozak, Nagel, and Santosh (2020)).

We find that non-linear models, NN in particular, generate larger economic profits than linear models in most international markets. Compared with the best linear method, we find that the best NN model outperforms in equal- (value-) weighted Sharpe Ratios in 30 (27) of the 31 markets. The out-of-sample R^2 of RTs and NN models is, however, less impressive and is often similar to that of linear models, possibly due to extreme values of international stock returns and characteristics. Kelly, Malamud, and Zhou (2022) also point out that out-of-sample R^2 can be a poor measure of the economic value of prediction models. We use two alternative measures that compare the predicted and actual return ranks and deciles (and are hence less affected by outliers), and again show the dominance of machine learning models and allay concerns of overfitting in the U.S.-based analysis.

In the second set of tests, we examine whether the models are robust when trained with different data and environments. Here we train and validate each model separately for each market. GKX show that the highest equal-weighted Sharpe Ratio (2.45) and value-weighed

[4] Hyperparameters define the model structure and learning process. Typically, a few sets of hyperparameter values are specified manually, and the machine learning algorithm selects the best set. Parameters are estimated from the data automatically given the hyperparameter values. Although prior U.S. results are based on an out-of-sample period, the models' predictive power and estimated parameters could be highly sensitive to the choice of the hyperparameter values.

Sharpe Ratio (1.35) are achieved by a NN model in the U.S. In most international markets, we find that NN is also the most profitable model within the market and is able to generate annualized Sharpe Ratios that are close to or above 2 (in equal-weighted portfolios) or above 1 (value-weighted). In this analysis, RTs show signs of overfitting and produce poor predictions. The best tree model *underperforms* the best linear model in terms of Sharpe Ratio and our decile-based measures in 41–62% of the markets. The effectiveness of RTs seems to depend heavily on the number of observations, as the underperformance of RTs is more pronounced in markets where there are fewer stocks and a shorter time period.[5] Given this finding, we focus on NN models in the remaining analysis.

The Sharpe Ratios in the market-specific NN models are usually higher than those in our U.S.-trained NN models. The difference is larger when the two models are less similar, as defined by the centered kernel alignment (CKA) similarity index (Kornblith, Norouzi, Lee, and Hinton (2019)). To our best knowledge, we are the first paper in the finance literature that adopts the CKA index, which studies the last hidden layer and considers complex interactions to identify structural similarities between two NN models. The CKA index can therefore be roughly interpreted as a summary measure of how much an international market resembles the U.S. market in terms of return-characteristic relationships. Our results suggest that return predictability can be enhanced by building a different model that better incorporates the market-specific components; therefore, local models appear to dominate a global one (trained using U.S. data).

Among the 36 variables, we show that firm size, one-month return reversal, and daily return volatility are the most important predictors in the U.S., while in other large international markets some other predictors can dominate. For example, volatility of dollar trading volume and of share turnover are important predictors in China, consistent with Liu, Stambaugh, and Yuan (2019) and Leippold, Wang, and Zhou (2021), who show that turnover can

[5] Cong, Feng, He, and He (2022) point out that standard tree models assume data are i.i.d., ignoring the cross-sectional correlations and time-series information; and the algorithm focuses on local optimization and can be prone to overfitting. They introduce a class of interpretable tree-based models for analyzing unbalanced panel data.

capture the impact of speculative trading by retail investors in China.

Our evidence confirms that NN is powerful from the perspective of a U.S. investor who decides to invest in each of these markets separately. We run an additional test pooling all 32 markets together and adding 31 country dummies as model inputs, which allow return-characteristics relations to vary across markets within one model. This test also corresponds to the case that a U.S. investor invests globally (ignoring any frictions associated with short selling). NN models continue to yield the best predictions among all models. We construct an alpha relative to Fama and French (2015) five factors plus momentum factors for developed markets and emerging markets. The best NN model gives a monthly equal-weighted (value-weighted) alpha of 3.84% (2.12%) (the results are similar if we use alternative models developed by Hou et al. (2011) and Karolyi and Wu (2018), both of which target to explain global stock returns). NN models also face lower downside risk—the maximum drawdown and the maximum one-month loss are usually smaller than those of other models.

In the final analysis, we investigate whether information extracted from U.S. stocks can further enhance the return predictability of NN models in international markets. Although we do not specify an asset pricing model formally by estimating sensitivities to risk factors (betas), our test is motivated by Cohen, Polk, and Vuolteenaho (2003) and Huang (2022), who find that gaps in book-to-market and in past returns, respectively, can predict the corresponding factor's return premium. We show that U.S. characteristic gaps add considerable incremental power to NN in both the global model of non-U.S. stocks and the market-specific models. In the market-specific models, we find that the variable importance of the U.S. characteristic gaps increases with the market integration metrics constructed by Bekaert, Harvey, Lundblad, and Siegel (2011) and Akbari, Ng, and Solnik (2020). Therefore, we provide suggestive evidence that the cross-section of U.S. stocks contains information relevant for international stocks, above and beyond their own characteristics. Markets are partially integrated, and the U.S. variables are more useful in markets that are closer to full integration.

Our paper belongs to the burgeoning literature that predicts asset returns with machine learning.[6] Freyberger et al. (2020) propose an adaptive group LASSO procedure to select characteristics and find that many previously identified return predictors do not provide additional information. Kozak et al. (2020) construct a robust stochastic discount factor from a small number of principal components. Feng et al. (2020) develop a regularized two-pass cross-sectional regression approach and show that only a small number of factors remain significant over time. Rapach and Zhou (2020) extend the approach of Han et al. (2018), designed for forecasting cross-sectional stock returns, and use the combination elastic net to predict the market excess return. Bali, Goyal, Huang, Jiang, and Wen (2020) use U.S. stock and bond characteristics to examine cross-market return predictability and conclude that the stock and bond markets are somewhat disintegrated. Bianchi, Büchner, and Tamoni (2021) and Bianchi, Büchner, Hoogteijling, and Tamoni (2021) show that NN and RTs improve the predictions of U.S. Treasury bond returns over linear techniques. Although machine learning is powerful, our paper suggests that we should exercise caution when applying it to international markets, where the number of observations is lower than the U.S.[7]

In international studies, Griffin (2002) finds lower pricing errors when local versions of Fama and French's three-factor model are used, compared with a world factor model. Hou et al. (2011) and Bekaert, Hodrick, and Zhang (2009) show that stock returns can be explained by local and international factors built from firm characteristics such as size, book-

[6] While it is not the main focus of the paper, we show evidence in the Online Appendix that both the non-linearity in the return-characteristic relationships and the complex interactions among predictors are important for NN's superior performance. First, when we add non-linearity via spline functions of individual features, the performance of OLS and LASSO in predicting U.S. stock returns improves, but it is still behind NN's performance. Second, we introduce a new class of models, Multivariate Adaptive Regression Splines (MARS), which is similar to trees and NNs (see the Online Appendix for details). When MARS with two degrees of terms (MARS2) is used, it takes into account both non-linearity and interactions. MARS2 generates equal- and value-weighted Sharpe Ratios and R^2 in the U.S. market that are similar to those generated by NNs. MARS with one degree of terms (MARS1), which only allows non-linearity but not interactions, underperforms MARS2 and NNs.

[7] Two recent papers by Cakici and Zaremba (2022) and Cakici, Fieberg, Metko, and Zaremba (2022) also show that machine learning models are effective in predicting the index returns and stock returns globally, but they do not compare local and U.S.-trained models and do not examine the issues of the number of observations, foreign factors, and market integration.

to-market ratios, cash flow-to-price, and momentum.[8] Carrieri, Chaieb, and Errunza (2013), Hau (2011), Bekaert et al. (2009), and De Jong and De Roon (2005) observe that developed markets are integrated, but emerging markets are segmented. Bekaert et al. (2009) argue that a country's regulations, political risk, and stock market development are local segmentation factors and U.S. corporate credit spread is a global segmentation factor. While we show evidence that the return-generating process seems to vary across markets, international markets are not totally segmented. Our NN models identify U.S.-based variables that help explain the cross-section of international stock returns.

2 Data and Methodology

2.1 Data

We obtain data on stock returns, trading volume, market capitalization, and industry information from DataStream. We winsorize raw returns at the top and bottom 2.5% in each exchange in each month to correct for potential data errors. Following Hou et al. (2011) and Ince and Porter (2006), all monthly returns that are above 300% and reversed within 1 month, as well as zero monthly returns, are removed (DataStream repeats the last valid data point of the return index for delisted firms). We obtain firm accounting information from Factset. We follow Green, Hand, and Zhang (2017) and attempt to construct the 94 characteristics used in their paper, but due to low data availability of certain variables in some markets, we end up with 36 characteristics as our model input, listed in Appendix A. For the U.S. and China, we use the data with CRSP and CSMAR, respectively, because of better coverage. We download data for as many markets as possible and require each market

[8] Our paper does not examine whether the explanatory power arises from the firm-level characteristics or from the covariance structure of returns that is related to these characteristics. For evidence on these two views, see Daniel and Titman (1997); Davis, Fama, and French (2000); Daniel, Titman, and Wei (2001); Hou et al. (2011); see also Kelly, Pruitt, and Su (2019) and Gu et al. (2020), who use machine learning to analyze U.S. stocks. We leave the interesting question of why characteristics are priced internationally in machine learning models for future research.

to have at least 100 stocks with valid observations of return and the 36 characteristics for at least 3 years. As a result, 32 markets, including the U.S., are in the final sample. Our data range from 2.4 million stock-month observations in the U.S. to around 6,100 in Kuwait. Appendix B provides the details. We normalize all stock characteristics to zero mean and unit standard deviation by month and market before inputting them into the model.

2.2 Model Estimation, Hyperparameter Tuning, and Out-of-sample Test

We focus on three categories of machine learning models: linear, regression trees (RTs), and neural network (NN), as in GKX. Linear models include OLS and its variants: OLS with a Huber loss function; LASSO, which selects a subset of predictors; and RIDGE, which restricts the magnitude of the regression coefficients. RTs are a non-parametric method for classifications and regressions. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. NN models aggregate and transform input signals into outputs, allowing for multiple layers of transformation and therefore complex interactions among the predictors. Both RT and NN models can capture non-linear and complex interaction effects. More technical details of the machine learning models are in the Online Appendix.

All models are set to predict the next month stock returns in U.S. Dollars in excess of the corresponding market return. This means that we focus on the return predictability in the cross section.[9] To train the model for each market, we separate the sample of the market into 3 non-overlapping parts, while maintaining their chronological order. Training data, which consist of the first 30% of the periods, are used to estimate the model subject to a particular set of hyperparameter values. Validation data, accounting for 20%, are deployed to construct forecasts and calculate objective functions based on the estimated

[9]The results are similar if we set to predict returns in excess of U.S. risk-free rate and if we use returns in the local currency instead of U.S. Dollars.

model from the training sample. During the validation process, we iteratively search for the best set of hyperparameters that optimizes the objective functions (and in each iteration we estimate the model again from the training data under the current hyperparameter values). Finally, testing data are the remaining 50%; they are "out-of-sample" in order to provide objective assessments of the models' performance after determining hyperparameters and normal parameters for the models.

Due to limited computational resources, as noted by GKX, models get retrained annually instead of monthly. Also, when we predict the returns in the next calendar year, the training data expands by one year whereas validation samples are maintained with the same size. For example, as shown in Appendix B, when predicting the cross-sectional stock returns in 1990 in the U.S., we set the training and validation samples as [1963, 1979] and [1980, 1989], respectively. When we predict the cross-sectional returns in 1991, the training and validation samples are [1963, 1980] and [1981, 1990], respectively.

We choose the same or the subset of the potential hyperparameter values of GKX, as shown in Appendix C. We first train and validate the model for the U.S. market following the above-mentioned procedure. Then, we apply the U.S.-estimated models to the corresponding years of other markets. This is our out-of-sample test using international data to investigate if the model overfits the U.S. data.

Our second test allows the machine learning model to be trained and validated using each market's data with the same set of potential hyperparameter values. Thus, the market-specific models, which can choose different hyperparameter values and vary across different countries/regions, are likely to be different from the U.S.-estimated models. If the machine learning model with its estimation and regularization techniques can truly capture the underlying data-generating process, which presumably varies across markets, market-specific models should outperform the U.S.-estimated model in non-U.S. markets.

2.3 Post-estimation Evaluation

We use a basket of measures to evaluate the overall performance of machine learning models and interpret the estimated models.

Sharpe Ratio. Our primary measure of model performance is the annualized Sharpe Ratio of long-short portfolio returns based on predicted returns (long stocks in the top decile of predicted returns and short stocks in the bottom decile). As a widely used measure of return predictability, our reported Sharpe Ratios can quantify the profitability when one exploits machine learning models for trading and be compared with other portfolios or trading strategies, such as the market portfolio or momentum.[10] Compared with other measures introduced later, Sharpe Ratios are economically meaningful from the perspective of investors.

Out-of-sample $\mathbf{R^2}$. To evaluate the predictability of each model, we report the out-of-sample R^2 (R^2_{oos}) based on Equation (1), which examines the model's forecast error (the sum of the squared differences between actual returns and predicted returns) and measures how the model's predictions fit the actual data. Following GKX, the denominator is the sum of squared excess returns *without demeaning*, as they argue that the alternative way of using the historical average will inflate the monthly out-of-sample R^2 by approximately 3%. We first calculate the R^2_{oos} for individual stocks, i.e.,

$$R^2_{oos} = 1 - \frac{\sum_{(i,t)\in \text{Test}} \left(r_{i,t} - \hat{r}_{i,t}\right)^2}{\sum_{(i,t)\in \text{Test}} r^2_{i,t}} \tag{1}$$

While it is an intuitive and widely used measure of prediction accuracy, as we show later, R^2_{oos} turns out to be sensitive to outliers (i.e., extreme prediction errors). This is particularly an issue of emerging markets, as realized returns and characteristics sometimes can have extreme values. In addition, Kelly et al. (2022) also point out that R^2_{oos} might

[10] The Sharpe Ratios are computed in the same manner as GKX. Kan, Wang, and Zheng (2022) note that high Sharpe Ratios are rarely delivered by professional fund managers. They show that out-of-sample Sharpe Ratios should be lower after taking into account the estimation risk of mean and co-variance of returns.

be a poor measure of the economic value of the forecast returns; for example, investors can generate potentially large economic profits even when R^2_{oos} is negative. To address those issues, we propose two alternative measures below.

Rank Correlation. We calculate the rank correlation between $r_{i,t}$ and $\hat{r}_{i,t}$, which measures the degree of similarity between the cross-sectional rankings of realized and predicted stock returns. In this paper, we choose Spearman's rank correlation coefficient, defined as the Pearson correlation coefficient between the rank variables. A higher rank correlation implies a more accurate model forecast.

Decile Score Distance. We sort stocks into deciles based on the model's predicted returns, and long (short) the top (bottom) decile. For each model, we calculate the actual return deciles of the long and the short portfolios in each market, and define a difference between the two as Decile Score Distance. If a model has zero predictive power, the actual return deciles would be 5.5 for both the long and the short portfolios on average, and Decile Score Distance would be zero. If a model has perfect predictive power, the actual return decile for the long (short) portfolio would be 10 (1), and the Decile Score Distance would be 9. Decile Score Distance measures the accuracy of model predictions in extreme deciles.

While machine learning models are regarded as "blackbox," the following measures are useful to interpret the return-characteristic relationship implied from the estimated models.

Relative Importance of Predictors. To identify significant predictors, we adopt the approach by Dimopoulos, Bourret, and Lek (1995) that the relative contribution of each input variable can be measured by computing the Sum of the Squares of the partial Derivatives (SSD). For the contribution of the j-th input variable to the function f that predicts excess stock returns, we calculate

$$\mathrm{SSD}_j = \sum_k \left(\left. \frac{\partial f}{\partial x_j} \right|_{x=x^k} \right)^2 \tag{2}$$

where x^k means the k-th observation. Then, we normalize all variables' SSD to sum of one, i.e., $\frac{\mathrm{SSD}_j}{\sum_i \mathrm{SSD}_i}$.[11]

[11] An alternative way to measure an input variable's importance (VI) is to calculate the decline in R^2

CKA Similarity Index. This measure compares the estimated structure of different machine learning models. In our context, we are interested in how a market's specific return-characteristic relationship is different from the U.S.-estimated one. While it is difficult to do so directly, we can nonetheless quantify the structural similarities of two estimated models. Specifically, we calculate a similarity index, centered kernel alignment (CKA) from Kornblith et al. (2019), which compares representations between different trained neural network models.

Specifically, let $X \in \mathbb{R}^{n \times p_1}$ denote a matrix of activations of p_1 neurons for n observations (e.g., the intermediate output of a specific hidden layer), and $Y \in \mathbb{R}^{n \times p_2}$ denote a matrix of activations of p_2 neurons for the same n observations. With respect to the choice of the hidden layer, we focus on the last hidden layer in each NN, because it is closest to the final model output. Then the linear version of CKA is obtained from

$$\mathrm{CKA}(X, Y) := \frac{\left\|Y^{\mathrm{T}} X\right\|_{\mathrm{F}}^{2}}{\left\|X^{\mathrm{T}} X\right\|_{\mathrm{F}}\left\|Y^{\mathrm{T}} Y\right\|_{\mathrm{F}}}, \tag{3}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm, an extension of the Euclidean norm on the space of all matrices.[12]

3 Predicting Stock Returns Using Machine Learning

3.1 U.S. Stock Returns

We first focus on the U.S. stock market and train the various machine learning models with the 36 stock characteristics (listed in Appendix A) to predict the cross section of monthly returns. Our main purpose is to verify whether the performance of our models are

when one sets all values of the input variable to zero. This is the approach used in GKX and Kelly et al. (2019). A negative VI value implies the increment of this input variable would lead to a decrease in output and vice-versa. The drawback of this measure is that it is hard to compare negative relative importance, especially across various markets.

[12] If $p_1 = p_2 = 1$, i.e., X and Y reduce to a n-dim vector, then $\mathrm{CKA}(X, Y) = \frac{\sum_{i=1}^{n} X_i Y_i}{\sqrt{\sum_{i=1}^{n} X_i^2}\sqrt{\sum_{i=1}^{n} Y_i^2}}$ is the cosine similarity between X and Y.

comparable with those in GKX, who input more than 900 features, before we apply our models to international markets.

We discuss the details of this test and present the results in the Online Appendix. Overall, with the 36 stock characteristics, our RTs and NN models appear to have similar return predictability to models in GKX using more than 900 inputs. One may be surprised by this finding, but it is consistent with some of the results in GKX and other studies. For example, GKX show that via dimension reduction, the ENET model selects only 20 to 40 features because the inputs and characteristics are partially redundant and fundamentally noisy signals (see Figure 3 of GKX). Furthermore, a few recent studies, such as Feng et al. (2020); Freyberger et al. (2020); Kozak et al. (2020), argue that a modest number of factors can explain cross-sectional U.S. stock returns. As we show later, using 36 characteristics seems to predict cross-sectional stock returns in international markets as well.

In the Online Appendix, we also show evidence that both non-linearities and the complex interactions among predictors contribute to the return predictability of machine learning models. The performance of OLS and LASSO improves after adding spline functions of individual features to capture non-linear terms; however, without interaction terms, these models still underperform RTs and NNs. Then we adopt a class of model called Multivariate Adaptive Regression Splines (MARS), which features a hyperparameter that specifies the maximum degree of terms. MARS1 with degree = 1 allows non-linearities, while MARS2 with degree = 2 allows both non-linearities and variable interactions. Only MARS2 generates performance that is close to that of NNs and RTs in terms of Sharpe Ratios and R^2_{oos}.

3.2 International Stock Returns with the U.S.-Estimated Models

Now we run a stringent test: applying the U.S.-estimated model to each of the 31 international markets individually. According to the machine learning literature, the regularization techniques we apply are known to prevent model overfit effectively. Nonetheless, examining model performance in some real-world, out-of-sample data is still meaningful. Assuming that the return-characteristic relationship is (at least partially) in common across countries, international markets are ideal out-of-sample data relative to the U.S. market and allow us to test overfitting. Furthermore, tuning the hyperparameter values is critical to achieving desirable model performance. When the model is heavily tuned over one data sample (i.e., the U.S. market), the possibility of overfitting is an important concern. Thus, we only use U.S. data to tune the model, making our tests below truly out of the sample.

We follow the definition of training, validation, and testing periods for the U.S. market specified in Section 2.2. Specifically, to predict stock returns in an international market in a particular year, we train and validate the machine learning models using past U.S. data only. Then for the following year, the training data expands by a year and the validation period maintains the same size.

Panel A of Table 1 reports equal- and value-weighted Sharpe Ratios of long-short portfolio returns, along with the Sharpe Ratio of the market portfolio during the sample period. We list the markets based on the descending order of the number of observations and highlight the method that gives the highest Sharpe Ratio in each market.

Starting with the equal-weighted portfolios on the left, we make two observations. First, in every market, machine learning-based models outperform traditional models (i.e., OLS-3 and OLS) or the passive market portfolio. Second, models taking into account nonlinear and complex interaction effects (i.e., RTs and NN models) outperform linear machine learning models (LASSO and RIDGE). The patterns are similar but slightly weaker for value-weighted Sharpe Ratios. Furthermore, the predictive power of NN models is economically sizable: using the best NN model in each market, the average equal-weighted (value-weighted) Sharpe Ratio of the 31 markets is 1.94 (1.07); 19 markets have an equal-weighted Sharpe Ratio above 1.5 and 26 markets above one, and 15 markets have a value-weighted Sharpe Ratio greater than one and 26 larger than 0.75.

We systematically compare the models' performance in Panel C. Specifically, we pick the best-performing model in each of the three categories (i.e., linear, trees, and NN) and calculate the difference in Sharpe Ratios and other measures. We find that on average the best performing tree model can generate an equal-weighted Sharpe Ratio that is 0.41 higher than the best linear model across the 31 markets, and among them 26 (or 84%) markets have a positive difference. Comparing NN with linear models, the average difference is even higher, at 0.65, with 30 (or 97%) markets being positive. The best NN model outperforms the best tree by 0.25 in the Sharpe Ratio on average, and 25 (or 81%) out of 31 markets have a positive difference. For value-weighted Sharpe Ratios, RTs do not appear to significantly outperform linear models: only 15 markets (48%) have a positive difference, while NN models still significantly outperform linear and tree models. In sum, based on Sharpe Ratios, NN models generate stronger return predictability than trees and linear models in the international markets, which is consistent with GKX's conclusion in the U.S. market.

However, the out-of-sample R^2 reported in Panel B.1 shows a different picture. While NN models are still the best model in more than half (17) of the markets, OLS-3 stands out in 11 markets. Regression trees do not give the best prediction in any market, and in many markets they generate negative out-of-sample R^2.[13] Panel C also shows that, in terms of R^2_{oos}, neither NN nor tree model outperforms linear models.

The results based on R^2_{oos} contradict those based on Sharpe Ratios. This pattern echoes Kelly et al. (2022), who point out that R^2_{oos} may be a poor measure as investors can generate potentially large economic profits even when R^2_{oos} is negative. In particular, R^2_{oos} can be sensitive to outliers (i.e., extreme prediction errors). This is particularly an issue of emerging markets, as realized returns and characteristics sometimes can have extreme values.

To investigate this possibility, we use two alternative measures, Rank Correlation and Decile Score Distance, defined in Section 2.3. These measures are based on relative ranks of returns and are thus less affected by extreme realized returns. As reported in Panels B.2 and B.3, the performance of non-linear machine learning models appears to be better than the results reported in Panel B.1. Panel C compares RT with linear models and finds the average

[13] A negative value of R^2_{oos} means that the model underperforms a naive model that always predicts zero expected return.

difference in rank correlation is 1.69%, with 26 (or 84%) markets being positive. The best NN model outperforms the best linear by 1.75%, and 29 (or 94%) out of 31 markets have a positive difference. The performance between the best RT and NN models is very close. The results are generally similar when Decile Score Distance is used. The only difference is that NN outperforms RT models in 22 (or 71%) of the 31 markets, suggesting that NNs are better at predicting extreme returns. This is aligned with the finding that NN models generate higher Sharpe Ratios than RT models.

Overall, the findings allay the concern of overfitting in the U.S.-based analysis when the effect of outliers is minimized. In the following analysis, we drop R^2_{oos} and focus on Rank Correlation and Decile Score Distance.

3.3 International Stock Returns with Market-Specific Models

Here we let each market train and validate its own model. Compared with the U.S. data, international data on stock return and characteristics appear to exhibit wider variation and more extreme observations, contain more frequent data errors or missing values, and have smaller sample sizes (both a smaller cross-section and a shorter time period). Those data limitations can possibly make the estimation of model parameters less consistent and efficient. Machine learning models feature a large number of parameters to be estimated. The heterogeneity of data quality and sample size across countries allows us to understand the robustness of various machine learning models. Our analysis sheds light on the application of machine learning models to return predictability.

We use the same procedure to split the samples, as described in Section 2.2, and the same set of hyperparameter values, listed in Appendix C. Table 2 summarizes the models' performance in Sharpe Ratios (Panel A) and Rank Correlation and Decile Score Distance (Panel B). The markets are sorted in descending order of the number of available observations. Panel C compares the model performance by categories.

Similar to what we find with U.S.-estimated models, NN models exhibit the strongest return predictability in most of the markets in terms of Sharpe Ratios. For the equal-weighted (value-weighted) Sharpe Ratio, NN models outperform linear and tree models by 0.60 (0.41) and 0.44 (0.61) on average or in 81% (78%) and 78% (88%) of the markets, respectively.[14] Economically, the best NN model achieves an equal-weighted (value-weighted) Sharpe Ratio above 1.5 (1) in 21 (20) of the 32 markets. Also, the patterns based on Rank Correlation and Decile Score Distance are similar. For example, the best-performing NN model's Rank Correlation outperforms by 1.08% and 1.12% on average or in 75% and 72% of the markets, compared to the best of linear and RT models, respectively.

By comparison, market-specific tree models do not seem to dominate linear models. In Panel C, relative to linear models, the average equal-weighted Sharpe Ratio of RTs is higher by 0.17, while the average value-weighted Sharpe Ratio of RTs is *lower* by 0.21. The average Rank Correlation and Decile Score Distance of RTs are similar to that of linear models.[15]

RTs may perform relatively poorly because of the high degrees of freedom in their structure and overfitting in-sample, despite the various regularization techniques we apply. Panel C of Table 2 shows that trees' performance is especially poor in markets where the number of observations is low: in the top half of markets with more observations, tree models' average equal- and value-weighted Sharpe Ratio is higher than that of linear models in 81% and 50% of the markets, respectively; but in the bottom half, these numbers fall to 38% and 25%. Using Rank Correlation and Decile Score Distance, in the top half of markets with more observations, RTs outperform linear models in 88% and 69% of the markets, respectively; the corresponding numbers are 25% and 38% in the bottom half. This suggests that RT models need more data to converge to a stable parameter estimation.

Note that, despite its complex structure as well, NN models appear to be more robust

[14] The Sharpe Ratios of the market portfolios in this table are different from those in Table 1 because the sample periods are shorter. In Table 2, the market portfolio generates the highest Sharpe Ratio in several markets, particularly for value-weighted Sharpe Ratios and in markets with fewer observations.

[15] In the Online Appendix, we compare the model performance using the equal- and value-weighted Sharpe Ratios of long-short portfolios formed using 9th minus 2nd decile portfolios (reported in Table A4). Similar to our main tests, market-specific tree models do not outperform linear models. Also, NN models' 9th-minus-2nd Sharpe Ratios are closer to those generated by linear models, suggesting that NNs are better in ranking stocks with more extreme returns.

to sample size. While Panel C documents a corresponding drop from the top half to the bottom half, the drop is smaller (94% and 81% vs. 69% and 75%), for equal- and value-weighted Sharpe Ratios. Linear models are also robust in estimation due to their simpler model structure, which in turn, however, limits their ability to capture complex return-characteristics relationships.

To better illustrate how the performance of RTs and NN models varies with sample size, we plot the performance improvement of RTs or NNs over the linear model (y-axis) against the log of the number of observations of the market (x-axis) in Figure 2. We also plot a fitted line and the 95% confidence intervals. We consider four performance measures, i.e., equal- and value-weighted Sharpe Ratio, Rank Correlation, and Decile Score Distance. The dashed line indicates the value of zero on the y-axis. First, one can see that for tree models, in markets with fewer observations, the scatter dots often fall under zero. Second, while the performance of NNs also increases in sample size, the fitted line is significantly above zero for the whole range of the x-axis.

While there is no clear theoretical explanation that RTs are more vulnerable to overfitting, our tests confirm that, at least for this type of financial data, the structure and the regularization settings of NN can fit and learn in a more robust manner. This is consistent with the evidence in the machine learning literature that random forests can be inconsistent (Tang, Garreau, and von Luxburg (2018)) and that NN models with multiple layers do not overfit the training data (Caruana, Lawrence, and Giles (2000); Kelly et al. (2022)). In sum, we conclude that NN models exhibit strong and more robust return predictability than trees or linear models. In the following tests, we focus on NN1–NN5 models.

4 Applications to International Asset Pricing

The results in the previous section suggest that NN models can capture and learn the true return-characteristics relationship in various markets. In this section, we exploit NN models as the tool to examine two long-standing questions in the international asset pricing literature—common versus market-specific return structure and cross-market integration. Studies using traditional methods build on strong assumptions about the function form between expected return and stock characteristics. Possible mis-specification of the function form makes it difficult to interpret non-results. NN models can mitigate the issue, because of their non-parametric nature of the model structure that can potentially capture all possible non-linear and complex interaction effects of stock characteristics. Using machine learning techniques, we provide new evidence for the questions.

4.1 Return-Characteristics Relationships: Common or Market-Specific?

Is the return-characteristics relationship generally common across different markets or dominated by market-specific features? On the one hand, under the rational framework, stock return should only depend on the stock's risk. In that sense, the return-generating function should be common across different countries. On the other hand, voluminous studies show that institutional frictions and investor behavior can influence asset returns. Since different countries may have distinct institutional settings or investor cultures, the return-characteristics relationship should be, at least to some extent, market-specific.

To shed some light on this question, we first analyze the relative importance of the 36 characteristics for each market's best performing NN model, based on the model estimations in Table 2. Results for the top 25 markets based on the number of observations are shown in Figure 3 (other markets are omitted for brevity). We observe some similar variables: e.g., log market capitalization ($mvel1$) and reversal (mom_1) are strong predictors for many markets. However, some market-specific features show up. For example, volatility of dollar trading volume (*stddolvol*) and of share turnover (*stdturn*) are important predictors in Japan and China, but not in other markets. These differences in variable importance suggest that the return structure may not be the same across countries.

Second, we compare the performance of market-specific models and their U.S.-estimated counterparts. To answer this question, we cannot simply compare the Sharpe Ratios in Tables 1 and 2. This is because the two models are not trained by the same amount of data (i.e., U.S. data sample is much larger and longer than any other markets), and the sample size for training and validating the model can influence the accuracy of model estimation (although it is less of a concern for NNs). Therefore, to make the comparison sensible, we require the U.S. model to be estimated only using the data over the same sample years that the market-specific model uses.

For example, China's data are available from 1999 to 2017, with 1999–2004 as the training period and 2005–2007 as the validating period. To compare the China-specific model with the U.S.-estimated counterpart, we train and validate with the U.S. data in 1999–2004 and 2005–2007, respectively.[16] Then, for each machine learning method, we compare the return predictions from the U.S-estimated model with those from the market-specific model, based on Sharpe Ratios. We repeat this procedure for each of the 31 international markets in our sample and summarize the differences across all markets.

Panel A of Table 3 presents the results. We find that market-specific models generally outperform their U.S. estimated counterparts. For example, market-specific models improve equal-weighted Sharpe Ratios by 0.69 to 0.77 and value-weighted Sharpe Ratios by 0.40 to 0.52 on average across the 31 markets. The improvement is pervasive: 74–87% of the markets experience an increase in Sharpe Ratio.

Two natural questions that follow are to what extent a country's specific model differs from the U.S. estimated one, and whether the difference, which presumably captures some useful market-specific return-characteristics relationship, is related to the improvement in return predictability. To address the first question, while it is difficult to directly show or interpret what market-specific relationship is really captured, we can nonetheless obtain some clue by comparing the structural similarities between the U.S.-estimated and market-specific

[16]A stricter approach is to further require the number of stocks to be the same in each cross-section. Given that the U.S. market has more stocks than most of the markets in our sample, our current approach gives market-specific models a disadvantage.

models. We adopt a similarity index, centered kernel alignment (CKA), from Kornblith et al. (2019). CKA similarity index compares representations between different trained neural network models.

For each market, we first compute the CKA similarities between representations from U.S.-estimated models and market-specific models. Specifically, given a dataset, we extract the intermediate output of the same hidden layer from U.S.-estimated models and market-specific models, and then compute the CKA according to Equation (3). Then, we examine the relationship between the CKA values and the Sharpe Ratio improvements from U.S.-estimated models to market-specific models across markets.[17]

In Panel B of Table 3, we split the markets in our sample equally into two groups based on its model's CKA, i.e., high versus low, and calculate the average improvements in Sharpe Ratio from U.S.-estimated to market-specific models. We notice that across the five models (NN1 to NN5), low CKA similarities are associated with more improvement in both equal- and value-weighted Sharpe Ratios. Also, such improvement is economically greater for NN5 than for NN1. For example, for NN5 model, low CKA markets exhibit an improvement of 0.83 in value-weighted Sharpe Ratio, while the number is 0.21 for high CKA countries; and the difference of improvement between high and low CKA markets is smaller for NN1 models. This is consistent with the conjecture that a more complex network structure can potentially better incorporate market-specific components and enhance return predictability. It is also clear from Figure 4, which shows the significant and negative relation between CKA similarity and Sharpe Ratio improvement across markets.

A global model: pooling all stocks

Then, based on the previous results, we pool all stocks in our global sample to train and validate a unified model to predict expected returns. This is to leverage the advantage

[17] If the variables that have high variable importance (in terms of SSDs) are different in two markets, we may also consider the return-characteristic relationships in the two markets to be different. However, this ignores the complex interactions among variables, and it is difficult to choose the number of important variables we should examine. Comparing the CKA similarity is a more systematic approach.

of machine learning models to take into account both common and market-specific, complex return-characteristics relationship. Also, with more data and larger space for portfolio selection, NN models can be better trained and have stronger predictive power.

Besides the 36 stock characteristics (listed in Appendix A), we add 31 dummies to indicate the 31 non-U.S. markets as the input of the global model. These market dummies allow NNs to learn possible country-specific structures, through, for example, the interaction between the country dummy and certain stock characteristics. NN models are set to predict the future stock returns in excess of the global average stock return. While the sample starts from 1963, our testing period is from January 1990 to December 2017 due to the availability of risk factors (more details below). For brevity, we focus on NN models and Sharpe Ratio as the performance measure, and compare to that of linear models.

The results are reported in Table 4. According to the top panel, the global equal-weighted (value-weighted) long-short portfolio based on NNs yields a Sharpe Ratio of 3.90 (1.69), a large improvement from previous tables. This is also much higher than the Sharpe Ratio of the market portfolio 0.96 (0.53) and the best performing linear model 2.59 (1.04). One should take the high Sharpe Ratio with caution for investment purposes, as the estimates here do not take into account transaction costs or other frictions, such as short-sale constraints, in the international equity markets.

We next examine the risk of the machine learning based long-short 10–1 portfolios. Following GKX, we first look at the maximum drawdown (MaxDD), maximum one-month loss (Max 1M Loss), and portfolio turnover rate. The maximum drawdown of a strategy is defined as,

$$\text{MaxDD} = \max_{0 \leq t_1 \leq t_2 \leq T}(Y_{t_1} - Y_{t_2}), \tag{4}$$

where Y_t is the cumulative log return from month zero through t. The maximum one-month loss is the lowest monthly return of the trading strategy. For equal-weighted portfolios, NN4-based strategies have the lowest maximum drawdown and one-month loss. For value-weighted portfolios, NN4 models have the lowest maximum drawdown, but OLS-3 has the smallest one-month loss.

The portfolio average monthly turnover is calculated as,

$$\text{Turnover} = \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i}\left|w_{i,t+1} - \frac{w_{i,t}(1+r_{i,t+1})}{\sum_{j}(1+r_{j,t+1})}\right|\right), \tag{5}$$

where $w_{i,t}$ is the weight of stock i in the portfolio at month t. It appears that the monthly turnover rate of NN-based strategies is approximately 150%, which is about 20 to 30% higher than the number shown in GKX based on the U.S. market. Given the larger pool of stocks and the important role of price trend predictors in machine learning models, it is not surprising that the outperformance is achieved with a relatively higher portfolio turnover rate.

The previous results are all based on raw returns. Last, we turn to risk-adjusted returns to examine whether the machine learning models capture something beyond the commonly known factors. We adopt three international asset pricing models to calculate risk-adjusted returns: the Fama-French five-factor model augmented with a momentum factor, the 6-factor model developed by Hou et al. (2011), and the partial-segmentation Carhart model in Karolyi and Wu (2018).[18] In the Fama-French model, we include a set of the 6 factors for developed markets and a set for emerging markets. That is, in total 12 factors are used for the risk adjustment of the global portfolio returns.

The bottom panels of Table 4 report the results. The monthly equal-weighted (value-weighted) alphas based on the best performing NN model are significantly positive, at 3.84%–4.89% (2.12%–2.31%) with t-statistics well above 5. Those existing factor models exhibit low R^2 for the NN-based portfolios. Information ratio (IR) ranges from 1.15 to 1.18 for

[18]Fama-French factor data are downloaded from Kenneth R. French's website. The Fama-French five factors include the excess return on the value-weighted market portfolio and portfolios formed on size, book-to-market, operating profitability, and investment. See Fama and French (2016, 2017) for more details. We thank Andrew Karolyi and Ying Wu for sharing the factor data from Hou et al. (2011) and Karolyi and Wu (2018). The model proposed by Hou et al. (2011) contains 6 factors: the market portfolios and factor-mimicking portfolios based on momentum and cash flow-to-price for developed markets and emerging markets. Their data are available from 1981 to 2010. Karolyi and Wu (2018) add a new factor to the global Carhart model to account for externalities driven by the incomplete accessibility to stocks and stock markets. The data are available from 1990 to 2010.

equal weighting and 0.48 to 0.52 for value weighting. For most measures, NN models, which take into account nonlinear and complex interaction effects, significantly outperform linear models.

4.2 Cross-Market Integration

In Section 3.2, we show that the U.S. equity market is relevant for many other markets. We study cross market integration in this subsection, specifically, whether the information derived from U.S. stocks can improve our predictions of international stock returns. We first start with the pooled sample (excluding the U.S.) and then examine market-specific models.

Pooling all non-U.S. stocks

While there are multiple ways to extract information from U.S. stocks, we add new variables that are similar to those commonly used in the literature. We construct three types of state variables:

1. U.S. Factors: In each month, for each of the 36 characteristics, we sort U.S. stocks into 10 deciles in descending order and compute the value-weighted returns for each decile. Then we define a factor as the return of the top decile portfolio minus the return of the bottom decile portfolio. This is similar to the way that common risk factors are constructed, such as Fama and French (1993, 2015, 2017).

2. U.S. Characteristic Gaps: In each month, we compute the characteristic gap as the divergence between the 95th percentile and the 5th percentile of a corresponding stock characteristic in the U.S. market. Cohen et al. (2003) and Huang (2022) show that gaps in book-to-market and in past returns, respectively, can predict future value and momentum returns.

3. Local Factors: As a comparison, we compute local factors in the same way as the U.S. factors. Stocks that are in the same market as the stock in question are used.

We also compute the interaction terms for each stock characteristic and its respective factor or characteristic gap. Therefore, on top of the 36 raw stock characteristics plus 30 country dummies, the augmented model in this section adds 36 × 3 factors or characteristic gaps + 36 × 3 interaction terms = 216 independent variables.[19]

Panel A of Table 5 reports the difference in equal-weighted and value-weighted Sharpe Ratios between the augmented models and the original models using only 36 stock characteristics plus country dummies. For most NN models, the augmented model generally improves equal- and value-weighted Sharpe Ratios. The last column shows the difference between the best performing of the original NNs and that of the augmented ones. The improvement is economically significant and equals 0.57 for equal-weighted and 0.54 for value-weighted Sharpe Ratios.

In Panel B, we reduce the number of additional variables by focusing on the top 10 characteristics. With the pooled sample of all non-U.S. stocks, we first train its NN models (NN1–NN5). Then, we select the top 10 characteristics according to their variable importance in the best NN model (based on the value-weighted Sharpe Ratio).[20] Therefore, in each test we add 10 × 3 factors or characteristic gaps + 10 × 3 interaction terms = 60 independent variables (on top of the 36 stock characteristics). With a reduced number of model inputs, the robustness of model estimation can be enhanced. The performance of augmented NN models with top 10 characteristics shows even higher equal- and value-weighted Sharpe Ratios. Comparing with the best original NN model, the best augmented NN model's equal-(value-) Sharpe Ratio is higher by 0.75 (0.68).

The difference between Panels A and B highlights that NN models do not necessarily become more powerful when having more independent variables. Even with a large number of observations, the full augmented NN models with 36 characteristics do not generate the

[19] We could also input local characteristic gaps, but it would be redundant to do so as stock level characteristics are model inputs and machine learning models allow such nonlinear relationships if they are useful.

[20] The best NN model is NN5, and the top 10 characteristics include *stddolvol*, *mom_1*, *retvol*, *mvel1*, *mom_12*, *turn*, *ill*, *stdturn*, *mom_6*, and *maxret*.

best results.

Figure 5 graphs the variable importance of each type of variables in the best augmented NN model using top 10 characteristics. The sum of variable importance is normalized to one. Stock characteristics are the most important (45%), followed by the U.S. characteristic gaps (34%) and U.S. factors (15%). Local factors have the lowest variable importance (8%).

Taken together, incorporating the information or state variables of the U.S. market can significantly improve the return predictability in other markets, supporting the conjecture of international market integration.

Market-specific augmented models

Now we rerun the market-specific models with the additional variables from the U.S. market. Because the market-specific models contain a much lower number of observations, we focus on a subset of additional independent variables and a subset of markets in order to have reliable estimates. Given our findings in the previous subsection, we only add U.S. characteristic gaps and U.S. factors based on the top 10 characteristics in each market. That is, for each market, we select the top 10 characteristics based on their variable importance in the best market-specific NN model in Section 3.3. Therefore, we add 10 $\times$ 2 factors or characteristic gaps + 10 $\times$ 2 interaction terms = 40 independent variables (on top of the 36 stock characteristics) in each market.[21] Also, given that NN performs more robust in markets with more observations (Table 2 and Figure 2), we only examine the top 25 markets ranked on the total number of observations.

In Table 6, we report the summary across all the 25 markets. In Panel A, the augmented models include both U.S. characteristic gaps and U.S. factors. In Panels B and C, the augmented models include only U.S. characteristic gaps and U.S. factors, respectively. Focusing

[21] We use two markets as examples to illustrate this procedure. According to the best market-specific NN model (based on value-weighted Sharpe ratio), in Japan the top 10 characteristics in terms of variable importance are: *stddolvol*, *chmom*_6, *stdturn*, *turn*, *mom*_6, *mom*_1, *mvel*1, *indmom*_*a*_12, *dovol*, and *ill*. The 10 U.S. factors and 10 U.S. characteristic gaps are chosen based on this list. The 20 interaction terms refer to the 10 variables in the list interacting with the corresponding U.S. factor and the corresponding U.S. characteristic gap. Then, in a different market, we use a different list. For example, the top 10 characteristics in China are: *stddolvol*, *mom*_1, *stdturn*, *chmom*_6, *mom*_12, *mvel*1, *dy*, *bm*, *retvol*, and *turn*.

on the best NN models, Panels A and B show similar results while Panel C is weaker. The best augmented NN models in Panels A and B yield higher Sharpe Ratios by 0.10–0.29 on average (64%–88% of markets with positive improvement), when compared with the best NN models using only 36 stock characteristics (shown in Table 2).

While the above results suggest that international markets seem to be partially integrated, are the U.S. variables more important in markets that are more integrated with the world? We explore this possibility using the market-specific NN models of the 25 markets. The degree of market integration in each market is proxied by three metrics: the segmentation measure constructed by Bekaert et al. (2011) and the economic integration and financial integration measures developed by Akbari et al. (2020).[22]

Figure 6 plots the relationship between the variable importance of U.S. characteristic gaps and the country rank based on the degree of market integration. (Not all the 25 markets appear in the Figure because the integration measures do not cover some of the markets.) We observe that the variable importance decreases with Bekaert et al. (2011)'s segmentation metric (which is the opposite to integration) and increases with Akbari et al. (2020)'s economic integration measure. U.S. variables are more important in countries that are more integrated with the world (such as United Kingdom) than in countries that are less integrated (such as Greece). The relationship between the variable importance and Akbari et al. (2020)'s financial integration measure is weaker.

Overall, information from U.S. stocks seems to be useful in producing better rankings of local stocks' predicted returns, and hence higher Sharpe Ratios, especially in markets that are closer to full integration. While NN models cannot explain why U.S. characteristic gaps are more important than U.S. factors, one possible reason is that the U.S. characteristic

[22] Bekaert et al. (2011)'s segmentation metric is constructed based on the earnings yield and the assumption of equal earnings yields across countries under the null of full integration. Derived using a return decomposition approach, Akbari et al. (2020) define economic integration as a common cash-flow dynamic and financial integration as a common risk-pricing dynamic. Akbari et al. (2020) highlight the difference between economic and financial integration using China and Ireland as examples. China is the second-largest economy but is considered as financially segmented from the world market. Ireland is one of the world's largest offshore financial centers but contributes little to global economic growth.

gaps contain information about global cash-flow news. The bottom two graphs of Figure 6 provide suggestive evidence that the U.S. characteristic gaps are more relevant for global cash-flow news than global discount rate news. In Panels A and B of Table 6, where we add the U.S. characteristic gaps to the market-specific NNs, we see the improvements in value-weighted Sharpe Ratios are larger than those in equal-weighted Sharpe Ratios, implying that the return predictability increases more for larger stocks. To the extent that cash-flows of larger stocks are more globally integrated (e.g., because of their multinational nature), this also indicates that the U.S. characteristic gaps may contain global cash-flow information. U.S. factors may carry such information too, but returns can be contaminated by noise and other variables. On the other hand, local factors do not appear to help enhance return predictability.

5 Conclusion

We construct a dataset of 32 international markets and use machine learning models to predict the cross-section of stock returns. In the U.S. market, even with only 36 characteristics, the predictive power and profitability of complex machine learning models are comparable to those documented in previous studies using hundreds of variables. More important, training our models using U.S. data and applying them on international stocks—a stringent test to address potential overfitting issues—concludes that machine learning models outperform linear models, particularly in forming profitable portfolios and predicting return rankings.

We achieve even stronger results if we train the neural network (NN) models separately for each market, allowing the models to pick up market-specific return-characteristic relationships. These results are more prominent when the market-specific model is less similar to the U.S.-trained model (measured based on the centered kernel alignment (CKA) index) and for NN models with more hidden layers. However, there are signs that regression trees overfit the in-sample data and underperform linear models, especially in markets where there are few observations.

While the return-generating process seems to vary across markets, international markets are not totally segmented. Market-specific NN models, especially in countries that are more integrated with the world, are even more powerful when we add U.S. characteristic gaps and the interactions between stock characteristics and their respective U.S. characteristic gap as independent variables.

We conclude that NN models, previously focusing on the U.S. market, can be applied to equity markets around the world. With a reduced set of predictors, one can examine more closely the return-characteristic relationships generated by the algorithms and link them to the market-specific structure. For example, Leippold et al. (2021) show that the most relevant variables when using NN models to predict Chinese stock returns are liquidity and fundamental factors, which they attribute to the short-termism of retail investors in China. Future research can provide more economic insights into other variables and other markets.

Another possible future research direction is to better explain the power of our NN models using an asset pricing model. We follow GKX and use characteristics to forecast returns, while the traditional asset pricing literature focuses on systematic risk factors and betas. Feng et al. (2022) combine deep learning optimization with asset pricing factor models. Their methodology, applied on U.S. equity data, starts from firm characteristics, generates risk factors, and fits the cross-sectional returns. Our results suggest that market-specific nonlinear and complex interactions among the predictors should not be overlooked, and the additional information carried by U.S. characteristics is valuable in international markets. It is interesting to see how the market-specific return-characteristic relationships and market integration can be linked to equilibrium asset pricing.

Appendix A List of Stock Characteristics

The table lists the acronym and definition of the 36 stock characteristics used as model inputs.

Acronym	Definition
absacc	Absolute accruals
acc	Working capital accruals
agr	Asset growth
bm	Book to market
bm_ia	Industry-adjusted book to market
cashdebt	Cash flow to debt
cashpr	Cash productivity
cfp	Cash flow to price ratio
cfp_ia	Industry-adjusted cash flow to price ratio
chmom_6	Change in mom_6
chpmia	Industry-adjusted change in profit margin
depr	Depreciation / PP&E
dolvol	Dollar trading volume
dy	Dividend to price
egr	Growth in common shareholder equity
ep	Earnings to price
herf	Industry sales concentration
ill	Illiquidity
indmom_a_12	Industry 12-month equal-weighted momentum
lev	Leverage
lgr	Growth in long-term debt
maxret	Maximum daily return
mom_1	1-month reversal
mom_12	12-month momentum
mom_6	6-month momentum
mve_ia	Industry-adjusted size
mvel1	Log market capitalization
pctacc	Percent accruals
retvol	Return volatility (standard deviation) of daily return
roe	Return on equity
salecash	Sales to cash
sgr	Sales growth
sp	Sales to price
stddolvol	Volatility of liquidity (dollar trading volume)
stdturn	Volatility of liquidity (share turnover)
turn	Share turnover

Appendix B List of International Markets

The table below lists the name of markets in our sample, along with the sample periods and the number of observations.

Market	Train	Valid	Test	# Rows
USA	[1963, 1979]	(1979, 1989]	(1989, 2017]	2456110
Japan	[2008, 2010]	(2010, 2011]	(2011, 2017]	349030
China	[1999, 2004]	(2004, 2007]	(2007, 2017]	277265
India	[2007, 2010]	(2010, 2012]	(2012, 2017]	230459
Korea	[1997, 2003]	(2003, 2007]	(2007, 2017]	224998
Hong_Kong	[1997, 2003]	(2003, 2007]	(2007, 2017]	174678
Taiwan	[2007, 2010]	(2010, 2012]	(2012, 2017]	93079
France	[1995, 2001]	(2001, 2005]	(2005, 2017]	92427
United_Kingdom	[2005, 2008]	(2008, 2010]	(2010, 2017]	68740
Thailand	[1997, 2003]	(2003, 2007]	(2007, 2017]	68082
Australia	[2008, 2010]	(2010, 2011]	(2011, 2017]	65555
Singapore	[2007, 2010]	(2010, 2012]	(2012, 2017]	50412
Sweden	[2001, 2005]	(2005, 2008]	(2008, 2017]	43510
South_Africa	[1997, 2003]	(2003, 2007]	(2007, 2017]	41985
Poland	[2006, 2009]	(2009, 2011]	(2011, 2017]	40630
Israel	[2005, 2008]	(2008, 2010]	(2010, 2017]	37071
Vietnam	[2010, 2012]	(2012, 2013]	(2013, 2017]	35671
Italy	[2001, 2005]	(2005, 2008]	(2008, 2017]	35491
Turkey	[2006, 2009]	(2009, 2011]	(2011, 2017]	33537
Switzerland	[2002, 2006]	(2006, 2009]	(2009, 2017]	28259
Indonesia	[2005, 2008]	(2008, 2010]	(2010, 2017]	27329
Greece	[2006, 2009]	(2009, 2011]	(2011, 2017]	20216
Philippines	[2006, 2009]	(2009, 2011]	(2011, 2017]	16963
Norway	[2007, 2010]	(2010, 2012]	(2012, 2017]	16451
Sri_Lanka	[2010, 2012]	(2012, 2013]	(2013, 2017]	16430
Denmark	[2007, 2010]	(2010, 2012]	(2012, 2017]	12309
Finland	[2007, 2010]	(2010, 2012]	(2012, 2017]	12305
Saudi_Arabia	[2010, 2012]	(2012, 2013]	(2013, 2017]	11708
Jordan	[2009, 2011]	(2011, 2012]	(2012, 2017]	11431
Egypt	[2010, 2012]	(2012, 2013]	(2013, 2017]	9342
Spain	[2011, 2012]	(2012, 2013]	(2013, 2017]	7493
Kuwait	[2012, 2013]	(2013, 2014]	(2014, 2017]	6123

Appendix C Hyperparameters of the Machine Learning Models

	LASSO	RIDGE	RF	GBRT+H	NN1 - NN5
Huber loss, ξ				99.9% quantile	
Penalty	$\lambda_1 \in (10^{-3}, 10^3)$	$\lambda_2 \in (10^{-3}, 10^3)$			$\lambda_1 \in (10^{-5}, 10^{-3})$
Max Depth			[1, 6]	[1, 2]	
Max Features			{3, 5}	{3, 5}	
Estimators			300	[1, 1000]	10
Weighting Scheme				{0.01, 0.1}	
Learning Rate					0.01
Activation Function					ReLU
Batch Size					10000
Epoches					100
Patience					5
Batch Normalization					✓
Neurons					[32, 16, 8, 4, 2]

References

Akbari, A., L. Ng, and B. Solnik (2020). Emerging markets are catching up: economic or financial integration? *Journal of Financial and Quantitative Analysis 55*(7), 2270–2303.

Avramov, D., S. Cheng, and L. Metzker (2022). Machine learning versus economic restrictions: Evidence from stock return predictability. *Management Science, forthcoming*.

Bali, T. G., A. Goyal, D. Huang, F. Jiang, and Q. Wen (2020). Different strokes: Return predictability across stocks and bonds with machine learning and big data. *Georgetown McDonough School of Business Research Paper* (3686164), 20–110.

Bekaert, G., C. R. Harvey, C. T. Lundblad, and S. Siegel (2011). What segments equity markets? *Review of Financial Studies 24*(12), 3841–3890.

Bekaert, G., R. J. Hodrick, and X. Zhang (2009). International stock return comovements. *Journal of Finance 64*(6), 2591–2626.

Bianchi, D., M. Büchner, T. Hoogteijling, and A. Tamoni (2021). Corrigendum: Bond risk premiums with machine learning. *Review of Financial Studies 34*(2), 1090–1103.

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

Cakici, N., C. Fieberg, D. Metko, and A. Zaremba (2022). Machine learning goes global: Cross-sectional return predictability in international stock markets. *Working Paper*.

Cakici, N. and A. Zaremba (2022). Empirical asset pricing via machine learning: The global edition. *Available at SSRN 4028525*.

Carrieri, F., I. Chaieb, and V. Errunza (2013). Do implicit barriers matter for globalization? *Review of Financial Studies 26*(7), 1694–1739.

Caruana, R., S. Lawrence, and C. Giles (2000). Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. *Advances in Neural Information Processing Systems 13*.

Chen, L., M. Pelger, and J. Zhu (2019). Deep learning in asset pricing. *Working Paper*.

Chinco, A., A. D. Clark-Joseph, and M. Ye (2019). Sparse signals in the cross-section of returns. *Journal of Finance 74*(1), 449–492.

Cohen, R. B., C. Polk, and T. Vuolteenaho (2003). The value spread. *Journal of Finance 58*(2), 609–641.

Cong, L. W., G. Feng, J. He, and X. He (2022). Asset pricing with panel tree under global split criteria. *Working Paper*.

Daniel, K. and S. Titman (1997). Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance 52*(1), 1–33.

Daniel, K., S. Titman, and K. J. Wei (2001). Explaining the cross-section of stock returns in japan: factors or characteristics? *Journal of Finance 56*(2), 743–766.

Davis, J. L., E. F. Fama, and K. R. French (2000). Characteristics, covariances, and average returns: 1929 to 1997. *Journal of Finance 55*(1), 389–406.

De Jong, F. and F. A. De Roon (2005). Time-varying market integration and expected returns in emerging markets. *Journal of Financial Economics 78*(3), 583–613.

Dimopoulos, Y., P. Bourret, and S. Lek (1995). Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processing Letters 2*(6), 1–4.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (1998). Value versus growth: The international evidence. *Journal of Finance 53*(6), 1975–1999.

Fama, E. F. and K. R. French (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics 105*(3), 457–472.

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

Fama, E. F. and K. R. French (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies 29*(1), 69–103.

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

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *Journal of Finance 75*(3), 1327–1370.

Feng, G. and J. He (2021). Factor investing: A bayesian hierarchical approach. *Journal of Econometrics*.

Feng, G., N. Polson, and J. Xu (2022). Deep learning in characteristics-sorted factor models. *Working Paper*.

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

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

Griffin, J. M. (2002). Are the fama and french factors global or country specific? *Review of Financial Studies 15*(3), 783–803.

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

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

Han, Y., A. He, D. Rapach, and G. Zhou (2018). What firm characteristics drive us stock returns. *Working Paper*.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *Review of Financial Studies 29*(1), 5–68.

Hau, H. (2011). Global versus local asset pricing: A new test of market integration. *Review of Financial Studies 24*(12), 3891–3940.

Hou, K., G. A. Karolyi, and B.-C. Kho (2011). What factors drive global stock returns? *Review of Financial Studies 24*(8), 2527–2574.

Hou, K., C. Xue, and L. Zhang (2020). Replicating anomalies. *Review of Financial Studies 33*(5), 2019–2133.

Huang, S. (2022). The momentum gap and return predictability. *Review of Financial Studies 35*(7), 3303–3336.

Ince, O. S. and R. B. Porter (2006). Individual equity return data from thomson datastream: Handle with care! *Journal of Financial Research 29*(4), 463–479.

Kan, R., X. Wang, and X. Zheng (2022). In-sample and out-of-sample sharpe ratios of multi-factor asset pricing models. *Available at SSRN 3454628*.

Karolyi, G. A. and R. M. Stulz (2003). Are financial assets priced locally or globally? *Handbook of the Economics of Finance 1*, 975–1020.

Karolyi, G. A. and S. Van Nieuwerburgh (2020). New methods for the cross-section of returns. *Review of Financial Studies 33*(5), 1879–1890.

Karolyi, G. A. and Y. Wu (2018). A new partial-segmentation approach to modeling international stock returns. *Journal of Financial and Quantitative Analysis 53*(2), 507–546.

Kelly, B. T., S. Malamud, and K. Zhou (2022). The virtue of complexity in return prediction. *Journal of Finance, forthcoming*.

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

Kornblith, S., M. Norouzi, H. Lee, and G. Hinton (2019). Similarity of neural network representations revisited. In *International Conference on Machine Learning*, pp. 3519–3529. PMLR.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

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

Lewis, K. K. (2011). Global asset pricing. *Annual Review of Financial Economics 3*(1), 435–466.

Liu, J., R. F. Stambaugh, and Y. Yuan (2019). Size and value in china. *Journal of Financial Economics 134*(1), 48–69.

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

Rapach, D. E., J. K. Strauss, and G. Zhou (2013). International stock return predictability: What is the role of the united states? *Journal of Finance 68*(4), 1633–1662.

Rapach, D. E. and G. Zhou (2020). Time-series and cross-sectional stock return forecasting: new machine learning methods. *Machine Learning for Asset Management: New Developments and Financial Applications*, 1–33.

Rouwenhorst, K. G. (1998). International momentum strategies. *Journal of Finance 53*(1), 267–284.

Tang, C., D. Garreau, and U. von Luxburg (2018). When do random forests fail? *Advances in Neural Information Processing Systems 31*.

Table 1. Performance of International Portfolios based on Predictions of U.S.-Estimated Machine Learning Models

This table reports the performance of machine learning models on the entire sample periods of each international market. All stocks are sorted into deciles based on predicted returns in the next month. Predictions are based on machine learning models estimated with U.S. stock market data of the 36 stock characteristics (listed in Appendix A). We report the annualized Sharpe Ratio of value- and equal-weighted long-short portfolio returns in Panel A, out-of-sample R-square of individual stocks in Panel B.1, Ranking Correlation in Panel B.2, and Decile Score Distance in Panel B.3. Models include OLS using only size, book-to-market, and momentum (OLS-3), OLS with all variables (OLS), LASSO, RIDGE, random forest (RF), gradient boosted regression trees with Huber loss (GBRT+H), and neural networks with one to five layers (NN1–NN5). Markets are sorted in a descending order on the number of observations. In Panel A, we also report the Sharpe Ratio of the market portfolio. The portfolio that generates the highest value of the performance measure is highlighted in color. Panel C compares the mode performance by categories. For example, in the column labeled as "Tree–Linear," *difference* refers to the highest Sharpe Ratio in tree models (i.e., RF and GBRT+H) minus the highest in linear models (i.e., OLS-3, OLS, LASSO, and RIDGE). "# of +" refers to the number of markets with a positive value of *difference*, and the fraction of markets with positive *difference* is also reported.

Panel A: Equal- and value-weighted Sharpe ratio

	Equal-Weighted												Value-Weighted											
	Market	OLS-3	OLS	LASSO	RIDGE	RF	GBRT+H	NN1	NN2	NN3	NN4	NN5	Market	OLS-3	OLS	LASSO	RIDGE	RF	GBRT+H	NN1	NN2	NN3	NN4	NN5
Japan	0.83	0.46	1.19	0.81	1.20	1.54	1.50	1.86	1.73	1.75	1.74	1.72	0.43	0.42	0.79	0.63	0.79	0.48	0.57	0.77	0.91	0.76	0.91	0.80
China	0.50	0.83	1.54	1.37	1.54	1.86	1.94	1.88	1.85	1.79	1.70	1.80	0.40	0.81	1.06	0.86	1.06	1.00	1.18	1.39	1.44	1.32	1.18	1.23
India	0.65	0.74	1.36	0.87	1.36	1.79	1.94	2.15	2.03	2.25	2.25	2.21	0.52	0.41	0.60	0.02	0.60	-0.06	0.13	0.73	0.47	0.70	0.86	0.62
Korea	0.58	1.32	1.71	1.54	1.71	2.17	2.20	2.13	2.14	2.20	1.89	1.94	0.49	0.70	0.76	0.72	0.76	1.07	0.92	0.98	1.02	1.10	0.96	0.91
Hong_Kong	0.47	0.88	1.14	0.82	1.13	1.88	2.06	2.31	2.14	2.34	2.54	2.24	0.28	0.44	0.27	0.30	0.26	0.60	0.62	0.50	0.97	0.86	0.96	0.94
Taiwan	0.43	0.87	0.57	0.29	0.57	1.20	1.30	0.95	1.26	1.12	1.07	1.08	0.40	0.38	0.06	-0.09	0.06	0.49	0.10	0.35	0.40	0.39	0.35	0.43
France	0.68	0.84	1.31	1.33	1.32	2.08	2.11	2.12	2.32	2.33	2.27	2.35	0.45	0.60	0.77	0.59	0.78	0.43	0.45	0.70	1.05	0.83	0.77	0.67
United_Kingdom	0.42	1.03	1.11	0.96	1.11	1.84	2.33	2.14	2.09	1.83	2.04	1.90	0.43	0.38	0.30	0.08	0.30	0.07	0.34	0.67	0.92	0.87	0.93	0.34
Thailand	0.78	0.82	0.35	0.20	0.34	1.04	1.07	1.43	1.33	1.42	1.47	1.36	0.48	0.31	0.30	0.03	0.30	0.37	0.40	0.81	0.86	0.91	0.77	0.85
Australia	0.76	1.19	2.30	2.29	2.31	3.31	3.94	3.81	3.83	3.64	3.67	3.50	0.54	0.58	0.72	0.42	0.72	1.01	1.19	1.60	1.72	1.98	1.88	1.54
Singapore	0.24	0.90	2.03	2.12	2.03	2.89	3.36	3.49	3.29	3.25	3.43	3.30	0.30	0.45	0.63	0.38	0.66	1.11	0.78	1.75	1.46	1.41	1.47	1.54
Sweden	0.62	0.96	1.44	1.22	1.42	1.15	1.61	1.98	1.73	1.80	1.83	1.71	0.44	0.55	0.60	0.65	0.61	0.65	0.54	0.82	0.77	0.84	0.74	0.58
South_Africa	1.22	1.65	1.82	1.92	1.81	2.15	2.16	2.65	2.45	2.49	2.51	2.62	0.75	0.92	0.72	0.78	0.72	0.75	0.63	1.11	0.96	0.98	0.92	0.94
Poland	0.25	1.07	1.16	1.11	1.16	1.47	1.60	1.40	1.65	1.60	1.88	1.55	0.24	0.55	0.24	-0.01	0.24	-0.08	0.00	0.89	0.91	1.12	1.31	0.95
Israel	0.63	1.11	1.41	1.09	1.41	1.12	1.47	1.68	1.78	1.80	1.62	1.63	0.30	1.16	0.50	0.03	0.51	0.97	0.67	1.15	1.23	1.19	1.23	1.15
Vietnam	0.65	1.17	1.85	1.85	1.80	3.33	2.64	3.32	3.54	3.41	3.27	3.36	0.56	0.79	0.36	-0.03	0.36	0.96	0.72	0.88	0.98	1.10	1.27	1.27
Italy	0.15	0.93	0.58	0.42	0.58	1.05	0.90	1.13	1.29	1.30	1.36	1.25	0.23	0.62	0.57	0.53	0.56	0.72	0.34	0.61	0.67	0.85	0.85	0.67
Turkey	0.79	0.32	0.62	0.23	0.62	0.58	0.65	0.77	0.82	0.81	0.93	0.76	0.61	0.28	0.59	0.45	0.59	-0.05	0.04	0.14	0.08	0.18	0.39	0.06
Switzerland	0.86	0.53	0.84	1.06	0.85	0.56	0.84	0.86	0.84	0.80	0.95	0.72	0.63	0.65	0.78	0.48	0.79	0.59	0.46	0.68	0.34	0.45	0.91	0.44
Indonesia	0.94	0.76	-0.05	-0.20	-0.06	0.45	0.49	0.74	0.66	0.89	0.63	0.57	0.88	0.48	0.19	0.09	0.12	0.19	0.09	0.81	0.56	0.93	0.79	0.76
Greece	0.25	0.57	2.76	2.62	2.72	2.84	2.95	3.64	3.35	3.37	3.22	3.25	-0.17	0.48	1.25	1.07	1.26	1.42	0.92	1.64	1.54	1.65	1.61	1.73
Philippines	1.07	0.36	1.09	1.09	1.09	1.62	1.97	2.06	1.81	1.91	1.73	1.73	0.78	0.45	0.60	0.61	0.59	0.55	0.77	0.98	0.99	0.96	0.73	0.81
Norway	0.14	0.57	0.98	0.35	0.95	0.77	0.66	0.98	1.15	1.15	1.19	0.88	0.29	0.42	0.71	0.16	0.71	0.67	0.32	0.52	0.69	0.80	1.15	1.04
Sri_Lanka	0.65	0.82	2.05	1.89	2.07	2.07	2.29	2.34	2.45	2.46	2.69	2.83	0.72	0.50	0.90	0.81	0.91	1.34	1.59	1.73	1.24	1.81	1.64	1.70
Denmark	0.19	0.35	1.01	0.71	1.03	1.09	1.45	1.66	1.72	1.59	1.60	1.63	0.63	-0.02	0.39	0.26	0.38	0.46	0.47	0.32	0.41	0.41	0.38	0.35
Finland	0.38	0.86	0.87	0.85	0.86	0.54	1.11	1.35	1.34	1.41	1.35	1.27	0.21	0.32	0.66	0.69	0.65	0.16	0.23	0.14	0.52	0.58	0.68	0.48
Saudi_Arabia	0.35	0.62	0.93	0.30	0.93	0.86	0.83	1.19	1.13	1.16	1.01	1.27	0.40	0.42	1.03	0.15	1.04	0.37	0.33	0.69	0.74	0.53	0.39	1.02
Jordan	0.40	0.91	1.23	0.60	1.24	1.49	1.02	1.86	1.62	1.52	1.80	1.65	-0.06	0.31	0.70	-0.10	0.73	1.06	0.20	1.32	1.08	0.74	1.10	1.44
Egypt	0.48	0.33	0.43	0.21	0.42	0.46	0.55	0.75	0.71	0.66	0.86	0.83	0.45	0.66	0.57	0.05	0.57	-0.38	-0.11	0.63	0.35	0.07	0.45	0.56
Spain	0.43	0.35	0.19	-0.28	0.22	0.56	0.35	0.43	0.11	0.46	0.44	0.31	0.48	0.18	-0.34	-0.14	-0.32	0.31	-0.19	0.34	-0.22	0.91	0.47	0.25
Kuwait	0.46	0.65	1.37	1.04	1.36	1.36	1.10	1.44	1.14	1.22	1.26	1.27	0.30	0.35	1.17	0.70	1.16	0.61	0.91	1.29	0.82	0.96	1.03	0.96

Panel B.1: R^2_{oos}

	OLS-3	OLS	LASSO	RIDGE	RF	GBRT+H	NN1	NN2	NN3	NN4	NN5
Japan	-0.19	-0.51	-0.14	-0.51	-0.42	-1.87	-0.54	-0.45	-0.46	-0.37	-0.32
China	-0.02	0.01	0.04	0.01	-1.26	-8.75	-0.41	-0.35	-0.40	-0.27	-0.34
India	0.08	0.04	0.00	0.04	0.11	-0.63	0.33	0.32	0.34	0.37	0.35
Korea	0.25	0.30	0.23	0.30	-0.22	-0.38	0.44	0.43	0.39	0.41	0.39
Hong_Kong	0.15	-0.01	0.00	-0.01	0.07	-0.99	0.32	0.30	0.28	0.36	0.35
Taiwan	-0.03	-1.02	-0.51	-1.02	-0.82	-6.48	-0.89	-0.70	-0.78	-0.65	-0.58
France	0.17	0.07	0.21	0.07	-0.10	-7.47	0.25	0.23	0.17	0.30	0.33
United_Kingdom	0.05	-0.31	-0.12	-0.30	-0.44	-2.11	-0.30	-0.31	-0.57	-0.24	-0.10
Thailand	0.13	-0.49	-0.29	-0.49	0.15	-2.44	0.13	0.16	0.13	0.24	0.19
Australia	0.12	0.37	0.27	0.37	0.87	0.63	1.16	1.14	1.07	1.19	1.10
Singapore	0.17	0.46	0.28	0.46	0.65	0.06	1.54	1.49	1.53	1.58	1.39
Sweden	0.12	0.10	0.15	0.11	-3.12	-6.95	0.16	0.10	-0.10	0.10	0.15
South_Africa	0.33	0.45	0.53	0.45	1.46	-2.14	1.64	1.54	1.55	1.57	1.42
Poland	0.11	-0.09	0.00	-0.09	-0.33	-3.03	0.43	0.49	0.44	0.55	0.49
Israel	0.24	-0.01	0.12	-0.01	-3.11	-5.32	-0.43	-0.59	-0.75	-0.35	-0.13
Vietnam	0.16	0.28	0.21	0.29	0.72	0.62	0.79	0.75	0.78	0.78	0.75
Italy	0.12	-0.90	-0.42	-0.90	-1.44	-5.67	-1.14	-0.91	-1.12	-0.75	-0.67
Turkey	-0.04	-0.54	-0.31	-0.54	-0.74	-3.23	-0.58	-0.69	-0.67	-0.64	-0.46
Switzerland	-0.18	-0.99	-0.50	-0.98	-12.33	-32.63	-3.08	-3.55	-4.20	-3.42	-2.86
Indonesia	0.21	-0.27	-0.24	-0.27	-0.32	-2.78	-0.01	-0.04	-0.02	0.02	0.07
Greece	0.03	0.85	0.55	0.85	0.65	-0.37	1.74	1.71	1.80	1.77	1.54
Philippines	0.08	-0.12	0.01	-0.11	0.10	-2.73	0.76	0.70	0.59	0.60	0.61
Norway	0.06	-0.20	0.00	-0.20	-4.37	-3.86	0.09	0.22	0.30	0.31	0.23
Sri_Lanka	0.03	0.57	0.35	0.57	0.13	0.62	0.90	0.81	0.72	0.80	0.91
Denmark	-0.17	-0.22	0.20	-0.21	-0.33	-2.80	0.16	0.34	0.04	0.21	0.25
Finland	0.15	-0.54	-0.02	-0.53	-4.26	-14.51	-0.68	-1.08	-1.04	-0.61	-0.34
Saudi_Arabia	-0.06	-0.81	-0.09	-0.81	-1.22	-2.48	-0.62	-0.18	-0.45	-0.16	-0.10
Jordan	0.10	0.01	0.08	0.01	-1.10	-6.90	0.43	0.32	0.22	0.44	0.55
Egypt	-0.13	-1.00	-0.32	-1.00	-0.73	-2.83	-0.52	-0.27	-0.52	-0.39	-0.40
Spain	-0.15	-1.32	-0.62	-1.31	-0.65	-1.83	-1.94	-1.99	-2.14	-1.46	-1.31
Kuwait	-0.03	-0.06	0.30	-0.05	-0.57	0.01	0.10	0.08	-0.02	0.26	0.11

Panel B.2: Rank Correlation

	OLS-3	OLS	LASSO	RIDGE	RF	GBRT+H	NN1	NN2	NN3	NN4	NN5
Japan	0.82	4.20	4.21	4.20	4.45	4.75	5.33	5.31	5.48	5.36	5.17
China	2.36	7.65	6.92	7.65	6.53	6.84	6.37	6.23	5.79	6.14	5.70
India	1.97	4.15	3.23	4.15	5.16	5.46	5.91	5.45	5.98	5.85	6.10
Korea	5.49	7.89	7.29	7.88	8.80	8.04	8.39	8.36	8.17	8.24	8.12
Hong_Kong	3.65	4.97	4.37	4.98	7.93	7.60	7.42	7.34	7.29	7.61	7.60
Taiwan	1.91	2.99	2.05	3.00	4.22	3.95	4.08	3.66	3.72	3.68	3.78
France	4.44	6.03	6.57	6.05	8.03	7.45	7.46	7.95	7.88	7.88	7.87
United_Kingdom	1.97	1.17	0.61	1.19	4.49	6.36	4.13	4.13	4.38	4.66	4.37
Thailand	3.59	4.02	3.41	4.02	5.66	5.12	6.60	6.47	6.49	6.79	6.65
Australia	2.66	4.43	4.60	4.45	7.67	9.23	7.41	7.39	7.16	7.38	7.40
Singapore	3.42	7.04	6.64	7.05	8.98	10.09	10.71	10.05	10.57	10.71	10.73
Sweden	4.73	5.24	5.13	5.26	6.82	7.51	6.29	6.32	6.31	6.61	6.40
South_Africa	5.01	6.98	7.57	6.99	7.15	7.54	8.62	8.11	8.44	8.56	8.52
Poland	2.51	2.10	2.29	2.11	5.02	6.67	6.08	5.88	5.97	6.11	6.25
Israel	5.35	4.34	3.39	4.34	4.34	6.43	5.44	6.00	5.93	6.15	6.05
Vietnam	3.25	7.30	5.75	7.31	7.48	8.15	8.74	8.34	8.73	8.70	8.80
Italy	4.14	1.69	2.62	1.71	5.84	5.28	4.99	5.23	5.14	5.42	5.67
Turkey	1.27	2.67	1.72	2.68	4.24	4.85	4.71	4.79	4.94	4.86	4.52
Switzerland	1.73	3.39	3.36	3.41	3.94	5.06	5.19	5.63	5.36	5.29	5.27
Indonesia	1.06	0.55	-0.40	0.55	3.62	3.69	3.92	3.49	3.66	3.94	3.99
Greece	2.11	8.34	10.68	8.36	8.10	10.06	10.67	10.64	11.25	10.85	10.94
Philippines	-0.20	3.66	4.33	3.67	6.74	7.89	7.85	7.23	7.18	7.18	7.77
Norway	2.81	3.00	1.66	3.02	5.64	6.43	5.73	5.89	6.03	6.05	5.96
Sri_Lanka	1.02	9.84	9.70	9.86	8.55	11.16	10.62	10.18	10.50	10.63	11.30
Denmark	1.38	2.28	3.88	2.30	4.92	7.16	4.85	5.22	5.41	5.24	4.86
Finland	4.27	4.88	4.59	4.89	4.74	5.61	6.48	7.48	7.60	7.38	7.07
Saudi_Arabia	3.33	5.13	4.51	5.12	5.22	5.80	6.37	6.85	6.84	6.78	6.95
Jordan	2.27	6.77	4.53	6.77	5.72	5.83	7.97	7.07	7.20	8.08	8.06
Egypt	3.54	4.74	2.45	4.74	5.45	5.81	6.08	6.26	5.78	5.33	5.82
Spain	-0.72	-2.00	-1.02	-2.01	1.08	2.94	0.03	0.64	0.15	0.97	0.92
Kuwait	-0.53	2.80	5.98	2.79	3.08	4.47	3.64	3.49	3.65	3.66	4.09

Panel B.3: Decile Score Distance

	OLS-3	OLS	LASSO	RIDGE	RF	GBRT+H	NN1	NN2	NN3	NN4	NN5
Japan	0.06	0.38	0.42	0.38	0.45	0.50	0.52	0.53	0.54	0.52	0.51
China	0.27	0.72	0.65	0.72	0.71	0.74	0.70	0.67	0.66	0.66	0.65
India	0.13	0.39	0.32	0.39	0.45	0.50	0.60	0.60	0.66	0.66	0.65
Korea	0.49	0.82	0.79	0.82	0.84	0.88	0.86	0.90	0.89	0.87	0.86
Hong_Kong	0.34	0.51	0.47	0.51	0.75	0.76	0.84	0.82	0.87	0.88	0.87
Taiwan	0.21	0.30	0.24	0.30	0.42	0.45	0.38	0.45	0.42	0.40	0.37
France	0.38	0.60	0.67	0.60	0.80	0.72	0.73	0.77	0.78	0.73	0.75
United_Kingdom	0.21	0.08	0.15	0.08	0.33	0.60	0.41	0.42	0.40	0.44	0.39
Thailand	0.37	0.35	0.29	0.34	0.59	0.54	0.76	0.73	0.75	0.80	0.76
Australia	0.30	0.42	0.54	0.43	0.85	1.12	0.88	0.96	0.93	0.89	0.84
Singapore	0.32	0.77	0.83	0.77	1.03	1.17	1.30	1.30	1.33	1.33	1.28
Sweden	0.44	0.59	0.61	0.59	0.54	0.74	0.67	0.65	0.67	0.66	0.59
South_Africa	0.54	0.64	0.74	0.64	0.67	0.83	0.94	0.89	0.89	0.86	0.93
Poland	0.31	0.28	0.32	0.28	0.44	0.67	0.54	0.65	0.62	0.70	0.60
Israel	0.55	0.49	0.40	0.48	0.36	0.60	0.56	0.66	0.62	0.56	0.59
Vietnam	0.38	0.56	0.58	0.55	0.84	0.77	0.86	0.91	0.87	0.86	0.91
Italy	0.39	0.19	0.29	0.20	0.52	0.53	0.48	0.57	0.55	0.57	0.50
Turkey	0.07	0.37	0.28	0.37	0.33	0.46	0.52	0.52	0.53	0.54	0.49
Switzerland	0.16	0.38	0.46	0.39	0.28	0.43	0.40	0.39	0.38	0.42	0.34
Indonesia	0.13	-0.10	-0.12	-0.10	0.22	0.25	0.35	0.37	0.46	0.37	0.38
Greece	0.16	0.89	1.12	0.89	1.10	1.15	1.37	1.38	1.36	1.30	1.34
Philippines	-0.14	0.43	0.59	0.43	0.86	0.91	1.06	0.98	0.96	0.95	0.99
Norway	0.31	0.33	0.13	0.32	0.42	0.50	0.45	0.55	0.54	0.57	0.48
Sri_Lanka	0.09	0.95	1.05	0.97	0.98	1.21	1.21	1.16	1.21	1.22	1.27
Denmark	0.21	0.41	0.45	0.41	0.52	0.83	0.79	0.72	0.82	0.74	0.71
Finland	0.35	0.42	0.49	0.42	0.32	0.54	0.63	0.67	0.72	0.67	0.60
Saudi_Arabia	0.29	0.60	0.29	0.60	0.58	0.52	0.70	0.69	0.69	0.61	0.71
Jordan	0.33	0.70	0.37	0.71	0.60	0.52	0.91	0.74	0.75	0.87	0.86
Egypt	0.22	0.54	0.34	0.54	0.41	0.44	0.55	0.53	0.51	0.58	0.60
Spain	-0.01	-0.07	-0.02	-0.06	0.04	0.31	0.18	0.14	0.25	0.24	0.15
Kuwait	-0.06	0.45	0.60	0.45	0.40	0.49	0.56	0.48	0.52	0.46	0.54

Panel C: Comparison of model performance

	Sharpe Ratio (EW)			Sharpe Ratio (VW)			R^2_{oos}		
	Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
difference	0.41	0.65	0.25	-0.05	0.37	0.42	-1.17	0.01	1.18
# of +	26	30	25	15	27	29	8	17	30
fraction of +	0.84	0.97	0.81	0.48	0.87	0.94	0.26	0.55	0.97

	Rank Correlation			Decile Score Distance		
	Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
difference	1.69	1.75	0.06	0.15	0.22	0.07
# of +	26	29	15	26	28	22
fraction of +	0.84	0.94	0.48	0.84	0.90	0.71

Table 2. Performance of International Portfolios based on Predictions of Market-Specific Machine Learning Models

This table reports the performance of machine learning models on the testing periods of each international market. All stocks are sorted into deciles based on their predicted returns for the next month. Predictions are based on machine learning models estimated with each market's data of the 36 stock characteristics (listed in Appendix A). We report the annualized Sharpe Ratio of value- and equal-weighted long-short portfolio returns in Panel A. Panel B presents Rank Correlation and Decile Score Distance. Models include OLS using only size, book-to-market, and momentum (OLS-3), OLS with all variables (OLS), LASSO, RIDGE, random forest (RF), gradient boosted regression trees with Huber loss (GBRT+H), and neural networks with one to five layers (NN1–NN5). Markets are sorted in a descending order on the number of observations. In Panel A, we also report the Sharpe Ratio of the market portfolio. The portfolio that generates the highest value of the performance measure is highlighted in color. Panel C compares the mode performance by categories. For example, in the column labeled as "Tree–Linear," *difference* refers to the highest Sharpe Ratio in tree models (i.e., RF and GBRT+H) minus the highest in linear models (i.e., OLS-3, OLS, LASSO, and RIDGE). "# of +" refers to the number of markets with a positive value of *difference*, and the fraction of markets with positive *difference* is also reported. Panel C reports the statistics for the sample of all markets and for subsamples equally split by the number of observations of the market (labeled as "Top half" and "Bottom half").

Panel A: Equal- and value-weighted Sharpe ratio

	Equal-Weighted												Value-Weighted											
	Market	OLS-3	OLS	LASSO	RIDGE	RF	GBRT+H	NN1	NN2	NN3	NN4	NN5	Market	OLS-3	OLS	LASSO	RIDGE	RF	GBRT+H	NN1	NN2	NN3	NN4	NN5
USA	0.72	0.79	1.85	1.76	1.88	2.30	2.65	2.77	2.91	2.81	2.78	2.91	0.75	0.52	0.76	0.69	0.76	0.73	0.78	1.02	1.19	1.13	1.16	1.31
Japan	1.49	0.55	1.73	0.66	1.63	1.61	2.04	2.12	1.76	1.72	1.87	1.53	1.10	0.48	0.29	0.33	0.45	0.30	0.52	0.73	0.83	0.66	0.59	0.33
China	0.45	0.66	2.30	2.24	2.34	3.12	2.80	3.01	2.69	2.70	2.79	2.89	0.25	0.49	1.31	1.36	1.39	2.25	1.81	2.13	1.61	1.66	1.93	1.99
India	1.26	0.95	2.50	2.01	2.46	3.02	3.17	4.14	4.52	4.59	4.44	4.53	0.96	0.70	0.35	0.80	0.20	1.23	0.31	1.88	2.78	2.06	2.83	2.05
Korea	0.65	1.82	2.69	2.55	2.55	3.05	3.25	3.50	3.46	3.51	3.31	3.32	0.40	0.98	0.79	1.15	1.08	1.85	1.94	1.79	1.85	2.00	1.72	2.00
Hong_Kong	0.56	1.12	1.62	1.50	1.58	2.44	2.35	2.70	2.95	2.81	2.83	2.45	0.43	1.03	1.02	1.07	0.97	0.57	0.22	1.20	1.27	1.16	1.29	1.46
Taiwan	0.95	0.13	0.98	0.63	0.80	1.08	0.95	0.68	0.64	0.77	0.99	-0.19	1.09	0.10	-0.65	-0.43	-0.40	0.70	0.29	0.09	0.05	-0.24	0.24	-0.29
France	0.54	1.09	2.33	2.16	2.19	2.34	2.95	2.37	2.93	2.70	2.28	2.35	0.37	0.68	0.32	0.47	0.25	0.43	0.47	0.93	0.68	0.61	0.84	0.86
United_Kingdom	0.67	1.12	1.81	1.82	2.06	2.15	2.09	2.09	1.52	2.38	1.66	2.29	0.80	0.75	1.65	1.24	1.63	1.44	1.22	0.80	0.56	1.13	1.27	1.63
Thailand	1.00	1.26	1.94	1.30	1.85	1.53	1.94	2.26	2.59	2.35	2.13	2.32	0.71	0.87	0.85	0.47	0.85	0.69	0.77	0.88	1.17	1.23	0.91	0.90
Australia	1.22	1.44	2.02	1.21	2.05	1.78	1.99	2.69	3.20	3.10	3.12	2.59	0.98	1.20	0.01	0.04	0.20	-0.58	-0.56	1.36	1.05	1.35	1.34	0.60
Singapore	0.32	1.02	2.90	2.41	3.17	4.11	3.75	4.39	4.74	4.12	4.41	4.69	0.35	0.69	-0.47	-0.31	0.33	0.78	0.55	1.91	3.46	2.69	2.56	2.53
Sweden	1.25	0.94	1.65	1.58	1.99	1.58	1.65	1.92	2.29	2.03	2.33	2.28	1.15	0.39	0.83	0.72	0.66	0.15	0.25	0.21	0.25	0.72	0.91	0.74
South_Africa	1.20	0.87	1.91	2.15	2.13	2.08	2.19	2.44	2.63	2.80	2.46	2.47	0.82	0.68	0.45	0.61	0.49	0.85	-0.06	0.58	0.28	0.42	0.43	0.38
Poland	0.62	0.46	0.65	0.42	0.68	0.89	1.22	1.24	1.51	1.23	1.28	0.99	0.66	0.40	0.33	0.72	0.38	0.22	-0.36	1.04	1.37	1.47	0.74	0.73
Israel	0.67	1.55	1.53	0.86	1.29	1.47	1.56	1.49	1.42	1.25	1.12	1.23	0.01	1.64	0.31	-0.03	0.31	0.32	-0.72	0.73	1.11	0.35	0.13	0.15
Vietnam	1.89	0.99	0.08	-0.02	0.11	2.20	1.30	3.07	2.81	2.92	2.11	1.89	0.98	0.52	-0.32	0.22	-0.62	0.22	0.49	1.96	1.12	1.44	1.16	-0.32
Italy	0.47	0.94	0.89	0.71	0.83	1.18	1.40	1.34	1.48	1.43	1.26	1.12	0.50	0.29	0.34	-0.19	0.04	0.39	0.14	0.38	0.43	0.26	0.02	0.09
Turkey	1.03	-0.31	1.17	0.55	0.60	0.94	0.81	1.12	1.05	0.36	0.63	0.23	0.81	-0.24	0.42	0.10	0.15	0.33	0.40	0.69	0.32	0.56	0.08	-0.06
Switzerland	1.00	1.18	2.23	2.40	2.12	1.42	1.20	1.99	1.67	1.73	1.16	1.60	0.88	0.43	0.88	0.96	0.26	0.09	0.74	0.90	0.54	0.92	0.26	0.81
Indonesia	1.16	1.55	0.87	0.68	1.01	0.90	1.24	0.52	0.37	0.94	1.14	0.92	0.98	1.07	-0.04	0.10	0.01	-0.36	-0.31	-0.04	0.02	0.30	0.72	0.36
Greece	1.14	1.23	3.16	2.56	2.93	4.09	3.74	4.20	4.65	4.24	3.96	4.63	0.36	1.18	1.37	1.63	1.60	0.75	1.73	2.22	1.68	1.89	1.78	2.47
Philippines	1.25	0.69	1.20	0.98	1.36	0.84	0.46	1.90	1.62	1.57	1.91	1.53	0.90	0.58	0.71	0.81	0.74	0.59	0.51	0.78	1.03	0.02	0.78	0.86
Norway	0.87	2.10	1.50	1.51	1.53	0.89	1.40	1.04	1.04	0.67	1.28	1.55	1.14	1.67	0.32	0.33	0.39	0.70	0.73	0.09	0.24	0.07	0.82	1.07
Sri_Lanka	0.60	0.95	2.87	2.71	2.85	1.84	2.66	4.47	4.10	3.88	3.56	4.80	0.48	1.17	1.04	1.12	1.39	0.38	1.24	2.05	2.32	1.89	1.67	2.26
Denmark	2.01	0.94	0.73	0.87	0.83	0.80	1.16	0.74	0.66	1.35	0.97	1.50	1.39	0.74	0.06	0.53	0.19	0.54	-0.34	0.59	0.39	0.80	-0.05	0.61
Finland	1.43	0.37	-0.17	0.53	-0.02	-0.53	0.07	0.21	0.09	0.29	0.57	0.17	1.24	-0.06	-0.39	0.58	-0.15	0.01	0.51	0.77	0.45	0.50	0.70	0.73
Saudi_Arabia	-0.03	0.33	0.66	0.89	0.37	1.52	1.13	1.46	0.87	1.09	0.65	0.73	0.10	0.34	0.48	-0.19	-0.40	0.22	0.20	0.38	-0.05	0.28	0.18	0.01
Jordan	0.94	0.79	1.30	1.78	1.54	1.23	0.79	1.49	1.61	1.67	1.35	1.17	-0.03	0.02	0.76	1.06	0.81	1.20	0.61	0.33	0.29	1.35	0.97	-0.18
Egypt	0.65	0.63	0.12	-0.34	-0.03	0.26	0.34	0.90	1.04	0.01	0.89	0.69	0.67	0.19	0.48	-0.10	0.23	-0.78	-0.38	0.65	1.23	0.46	0.31	0.56
Spain	0.44	-0.27	0.95	0.45	1.01	0.16	1.30	0.45	0.39	0.76	1.03	0.77	0.47	-0.27	0.50	0.07	0.90	0.44	0.10	0.41	0.41	0.83	0.89	1.04
Kuwait	0.25	0.38	0.69	0.88	0.48	-0.21	0.84	1.14	1.08	0.78	1.14	0.80	0.27	0.30	0.95	0.93	0.62	-0.34	1.07	1.04	0.97	0.43	0.67	0.49

Panel B: Rank Correlation and Decile Score Distance

	Rank Correlation											Decile Score Distance										
	OLS-3	OLS	LASSO	RIDGE	RF	GBRT+H	NN1	NN2	NN3	NN4	NN5	OLS-3	OLS	LASSO	RIDGE	RF	GBRT+H	NN1	NN2	NN3	NN4	NN5
USA	1.57	4.97	4.74	4.96	5.29	6.13	5.89	6.08	6.05	6.16	6.02	0.15	0.51	0.51	0.51	0.56	0.66	0.61	0.64	0.63	0.63	0.63
Japan	-2.19	2.89	1.28	2.59	1.27	2.13	3.04	2.66	2.31	2.54	1.73	-0.21	0.29	0.03	0.24	0.20	0.34	0.31	0.22	0.21	0.25	0.17
China	2.76	11.47	12.10	11.60	12.40	12.03	12.56	12.41	12.15	12.36	12.54	0.25	1.04	1.13	1.06	1.34	1.22	1.26	1.23	1.19	1.20	1.23
India	-0.56	4.79	4.30	4.76	6.69	7.03	8.04	8.12	8.32	8.39	8.10	-0.18	0.49	0.44	0.49	0.73	0.71	0.88	0.95	0.89	0.90	0.87
Korea	5.20	9.81	10.41	10.06	11.01	11.10	10.54	10.58	10.92	10.48	10.95	0.50	0.97	1.05	0.99	1.20	1.21	1.15	1.16	1.17	1.15	1.14
Hong_Kong	-0.05	3.91	2.56	3.88	5.32	5.23	7.17	7.73	7.47	7.33	6.62	0.04	0.43	0.33	0.44	0.65	0.61	0.84	0.86	0.82	0.84	0.77
Taiwan	-0.91	3.25	2.86	2.80	4.41	4.22	2.91	3.80	4.24	3.23	2.18	-0.09	0.44	0.22	0.36	0.58	0.45	0.31	0.40	0.43	0.45	0.07
France	5.36	4.73	6.04	4.85	4.86	6.60	5.16	5.57	5.55	5.28	5.37	0.49	0.53	0.63	0.51	0.42	0.58	0.51	0.60	0.58	0.54	0.52
United_Kingdom	6.26	3.38	4.51	4.24	7.42	6.78	4.38	4.49	5.49	6.30	6.48	0.54	0.37	0.41	0.48	0.54	0.46	0.36	0.28	0.48	0.49	0.51
Thailand	0.67	3.44	2.48	3.36	2.77	4.08	5.44	5.63	5.50	5.09	4.49	0.03	0.37	0.13	0.32	0.21	0.29	0.60	0.64	0.61	0.50	0.52
Australia	-1.50	2.05	1.33	2.03	4.06	3.49	5.00	4.91	5.67	5.90	4.96	-0.23	0.31	0.09	0.30	0.32	0.38	0.68	0.71	0.76	0.84	0.69
Singapore	-4.64	6.28	5.53	7.40	12.19	11.70	13.47	14.09	12.21	12.89	12.88	-0.37	0.75	0.68	0.94	1.26	1.21	1.42	1.50	1.36	1.41	1.45
Sweden	5.93	7.17	8.47	9.05	7.87	9.06	7.96	9.26	8.33	8.64	9.29	0.52	0.70	0.85	0.96	0.59	0.69	0.73	0.86	0.75	0.75	0.86
South_Africa	1.82	4.88	5.11	4.69	6.77	6.86	6.53	6.54	7.05	7.18	6.79	0.16	0.65	0.69	0.65	0.76	0.82	0.74	0.84	0.94	0.87	0.79
Poland	-2.09	0.34	3.07	0.73	0.37	1.71	1.36	2.33	2.03	2.83	1.51	-0.24	-0.03	0.15	-0.01	0.00	0.20	0.10	0.25	0.16	0.22	0.09
Israel	4.74	5.20	4.61	5.51	5.96	5.29	6.31	7.11	7.26	7.05	6.66	0.55	0.59	0.48	0.52	0.55	0.55	0.55	0.61	0.55	0.53	0.52
Vietnam	2.81	3.91	2.46	3.94	6.62	6.30	8.73	9.38	9.23	8.46	9.20	0.35	0.18	0.18	0.18	0.70	0.64	0.96	0.98	0.98	0.84	0.85
Italy	8.83	8.34	10.02	9.24	8.47	8.99	9.41	10.05	10.79	9.18	9.49	0.90	0.76	0.87	0.88	0.92	0.95	0.97	1.04	1.03	0.90	0.95
Turkey	-5.20	1.61	0.25	0.58	2.46	3.12	2.36	2.50	1.73	1.12	0.71	-0.60	0.20	0.06	0.07	0.27	0.29	0.22	0.20	0.04	0.14	0.05
Switzerland	4.78	6.45	8.78	6.85	3.97	4.50	5.69	5.65	5.50	3.94	5.40	0.45	0.74	0.78	0.67	0.45	0.47	0.68	0.60	0.61	0.40	0.50
Indonesia	4.15	2.66	2.85	2.72	2.07	2.34	1.69	1.91	2.29	3.25	3.06	0.43	0.25	0.16	0.27	0.08	0.26	0.20	0.17	0.45	0.40	0.25
Greece	-2.15	9.79	11.26	10.57	12.16	11.74	14.93	15.39	14.67	16.12	15.65	-0.12	0.96	1.02	1.05	1.30	1.23	1.53	1.74	1.55	1.69	1.79
Philippines	-5.06	3.43	1.11	1.37	-1.12	2.06	3.80	4.79	4.23	4.21	3.33	-0.43	0.33	0.20	0.27	0.19	0.23	0.85	0.94	0.90	1.01	0.68
Norway	12.35	5.96	7.63	8.04	8.00	8.20	6.33	6.98	4.98	7.42	6.13	1.06	0.68	0.66	0.81	0.48	0.88	0.65	0.68	0.39	0.80	0.57
Sri_Lanka	-1.41	8.20	8.84	8.22	3.62	7.37	11.32	11.23	12.21	10.93	12.91	0.06	0.93	1.03	0.90	0.57	0.91	1.50	1.52	1.67	1.36	1.66
Denmark	5.53	4.26	5.08	4.79	7.17	7.01	3.75	5.10	6.49	6.37	6.74	0.52	0.27	0.51	0.29	0.67	0.63	0.39	0.49	0.88	0.56	0.79
Finland	2.48	1.68	2.78	3.49	2.13	3.36	1.68	2.59	3.41	3.59	3.10	0.38	0.08	0.24	0.31	0.14	0.24	0.30	0.22	0.51	0.66	0.28
Saudi_Arabia	1.13	3.44	7.37	4.17	4.54	4.45	4.61	5.42	3.67	4.22	2.78	0.12	0.34	0.42	0.32	0.79	0.52	0.54	0.54	0.45	0.39	0.35
Jordan	2.87	8.65	8.71	9.27	7.39	7.12	9.92	8.29	8.90	9.40	10.78	0.46	0.72	0.95	0.88	0.53	0.49	0.86	0.71	0.88	0.87	0.93
Egypt	0.85	-0.13	4.47	0.00	2.40	3.52	2.33	3.18	-0.43	2.69	2.00	0.07	0.16	0.04	0.05	0.11	0.07	0.41	0.46	-0.11	0.31	0.33
Spain	4.98	6.25	2.79	7.04	6.32	6.15	3.82	3.10	5.41	5.42	6.01	0.24	0.46	0.18	0.86	0.27	0.60	0.49	0.49	0.59	0.71	0.52
Kuwait	-7.40	-0.01	1.03	-1.80	-4.34	-2.67	1.52	-0.29	-0.15	-0.10	-1.12	-0.55	0.20	0.04	-0.18	-0.65	-0.09	0.39	0.17	0.10	0.43	0.02

Panel C: Comparison of model performance

		Sharpe Ratio (EW)			Sharpe Ratio (VW)		
		Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
All markets	*difference*	0.17	0.60	0.44	-0.21	0.41	0.61
	# of +	19	26	25	12	25	28
	fraction of +	0.59	0.81	0.78	0.38	0.78	0.88
Top half	*difference*	0.37	0.77	0.41	-0.14	0.55	0.68
	# of +	13	15	12	8	13	13
	fraction of +	0.81	0.94	0.75	0.50	0.81	0.81
Bottom half	*difference*	0.37	0.77	0.41	-0.14	0.55	0.68
	# of +	6	11	13	4	12	15
	fraction of +	0.38	0.69	0.81	0.25	0.75	0.94
		Rank Correlation			Decile Score Distance		
		Tree−Linear	NN−Linear	NN−Tree	Tree−Linear	NN−Linear	NN−Tree
All markets	*difference*	-0.04	1.08	1.12	0.02	0.20	0.19
	# of +	18	24	23	17	25	23
	fraction of +	0.56	0.75	0.72	0.53	0.78	0.72
Top half	*difference*	1.01	1.69	0.68	0.08	0.18	0.10
	# of +	14	14	12	11	13	10
	fraction of +	0.88	0.88	0.75	0.69	0.81	0.62
Bottom half	*difference*	-1.10	0.47	1.57	-0.05	0.23	0.28
	# of +	4	10	11	6	12	13
	fraction of +	0.25	0.62	0.69	0.38	0.75	0.81

Table 3. Comparison of Performance Between Market-Specific and U.S.-Estimated Models

Panel A reports the comparison of return predictions between market-specific and U.S.-estimated machine learning models. For each market, *difference* is calculated as the equal-weighted (EW) or value-weighted (VW) Sharpe Ratio based on the market-specific model minus that based on the U.S.-estimated model. The corresponding U.S.-estimated model is trained and validated using the U.S. data in the same years that the market-specific model uses. Machine learning models are estimated with the data of the 36 stock characteristics (listed in Appendix A). The models include neural networks with one to five layers (NN1–NN5). "# of +" refers to the number of markets with a positive value of *difference*, and the fraction of markets with positive *difference* is also reported. Panel B reports the improvements of Sharpe ratio (*difference*) for NN1-NN5 models by subsamples equally split by models' CKA similarity.

Panel A: Sharpe Ratio improvement between market-specific and U.S.-estimated model

		NN1	NN2	NN3	NN4	NN5
Sharpe Ratio (EW)	difference	0.74	0.77	0.75	0.69	0.74
	# of +	26	26	26	24	27
	fraction of +	0.84	0.84	0.84	0.77	0.87
Sharpe Ratio (VW)	difference	0.52	0.46	0.40	0.42	0.52
	# of +	24	23	24	23	23
	fraction of +	0.77	0.74	0.77	0.74	0.74

Panel B: CKA similarities and Sharpe Ratio improvement

		NN1	NN2	NN3	NN4	NN5
Sharpe Ratio (EW)	Low CKA	0.92	1.10	0.92	0.78	1.20
	High CKA	0.56	0.44	0.58	0.60	0.30
Sharpe Ratio (VW)	Low CKA	0.49	0.65	0.52	0.54	0.83
	High CKA	0.55	0.27	0.28	0.30	0.21

Table 4. Performance of International Portfolios: Pooling All Stocks

All stocks are pooled together and sorted into deciles based on predicted returns in the next month. Predictions are based on the machine learning models estimated with the data of the 36 stock characteristics (listed in Appendix A) of all stocks and 31 market dummies. We report the annualized Sharpe Ratio of equal- and value-weighted long-short portfolio returns of individual stocks. Max DD is the maximum drawdown of the long-short portfolio. Max 1M Loss is the lowest monthly return of the long-short portfolio. Turnover is the portfolio turnover rate defined as Eq.(5). α is the intercept from regressing monthly long-short portfolio return onto a factor model. FF5+Mom refers to a 12-factor model that includes the international Fama-French five factors plus a momentum factor for developed markets and for emerging markets (available from 1992 to 2017). HKK refers to the 6-factor model in Hou et al. (2011) (available from 1990 to 2010), and KW refers to the partial-segmentation Carhart model in Karolyi and Wu (2018) (available from 1990 to 2017). The t-statistics of α and the R^2 of the regression are reported. Information ratio equals α divided by the standard deviation of the residuals from the regression. The testing sample is from 1992 to 2017. The model that generates the highest Sharpe Ratio, mean return, α, and information ratio, or the lowest Max DD, Max 1M Loss, Turnover, and R^2 is highlighted in color.

	Equal-Weighted										Value-Weighted									
	Market	OLS-3	OLS	LASSO	RIDGE	NN1	NN2	NN3	NN4	NN5	Market	OLS-3	OLS	LASSO	RIDGE	NN1	NN2	NN3	NN4	NN5
Sharpe Ratio	0.96	1.32	2.53	2.39	2.59	3.75	3.81	3.73	3.90	3.78	0.53	0.78	0.85	0.90	1.04	1.45	1.67	1.65	1.69	1.65
Drawdowns and Turnover																				
Max DD (%)	50.78	35.7	35.69	33.99	35.73	21.04	17.90	22.60	16.42	24.24	67.53	30.61	30.89	32.29	30.35	28.53	25.82	35.15	24.83	27.41
Max 1M Loss (%)	19.41	27.35	25.58	24.57	25.58	17.17	15.81	18.58	14.47	19.86	27.89	14.61	22.71	15.54	22.71	24.14	24.95	25.62	21.57	24.25
Turnover (%)		54.16	144.9	153.19	145.03	140.05	142.25	142.28	142.58	142.89		60.27	155.71	161.08	155.64	148.31	150.66	151.29	152.52	151.87
Risk-adjusted Performance using FF5 + Mom model, 1992-2017																				
Mean Return (%)		1.66	2.81	2.65	2.82	4.04	4.11	4.12	4.16	4.10		0.99	1.54	1.31	1.54	2.10	2.13	2.03	2.19	1.95
α		0.94	2.36	2.19	2.37	3.66	3.82	3.80	3.84	3.82		1.21	1.46	1.32	1.47	2.04	2.10	2.09	2.12	2.01
$t(\alpha)$		5.39	10.43	9.44	10.47	15.84	16.11	15.20	16.86	15.27		4.17	4.49	3.97	4.52	6.13	6.75	6.53	7.06	6.96
R^2 (%)		64.59	29.64	28.66	29.60	21.14	16.65	17.06	17.93	18.44		6.74	3.77	3.03	3.79	4.45	3.29	5.64	3.65	3.84
Information Ratio		0.38	0.73	0.66	0.74	1.11	1.13	1.07	1.18	1.07		0.29	0.32	0.28	0.32	0.43	0.48	0.46	0.51	0.49
Risk-adjusted Performance using HKK model, 1990-2010																				
Mean Return (%)		1.74	3.31	3.36	3.30	4.82	4.92	4.95	4.98	4.92		0.89	1.29	1.26	1.29	2.08	2.18	2.07	2.29	2.09
α		1.43	3.22	3.23	3.21	4.74	4.78	4.87	4.76	4.86		1.00	0.98	0.92	0.98	1.87	1.89	1.87	2.16	1.96
$t(\alpha)$		5.08	10.96	10.73	11.26	16.77	16.77	16.39	16.95	15.71		3.29	3.08	2.80	3.09	5.59	6.87	5.75	7.15	6.48
R^2 (%)		30.22	7.91	6.26	8.10	5.34	5.15	5.12	6.40	5.25		2.20	5.60	5.56	5.71	5.26	3.97	5.28	4.09	3.09
Information Ratio		0.34	0.74	0.73	0.76	1.13	1.13	1.11	1.15	1.06		0.22	0.21	0.19	0.21	0.38	0.46	0.39	0.48	0.44
Risk-adjusted Performance using KW model, 1990-2010																				
Mean Return (%)		1.70	3.28	3.31	3.28	4.76	4.88	4.90	4.87	4.88		0.90	1.41	1.37	1.42	2.10	2.28	2.14	2.29	2.16
α		1.13	3.09	3.18	3.10	4.70	4.86	4.84	4.89	4.85		0.80	1.30	1.22	1.30	2.01	2.31	2.09	2.24	2.10
$t(\alpha)$		4.60	10.63	10.51	10.98	16.48	16.38	16.09	16.68	15.40		2.51	3.93	3.54	3.94	5.76	6.75	6.23	7.53	6.69
R^2 (%)		50.82	17.66	13.82	18.24	11.54	10.53	9.82	9.49	10.18		3.57	1.72	1.27	1.77	3.27	2.88	3.79	2.17	1.76
Information Ratio		0.32	0.74	0.73	0.77	1.15	1.14	1.12	1.16	1.08		0.18	0.27	0.25	0.28	0.40	0.48	0.44	0.52	0.47

Table 5. Performance Comparison Between the Augmented and Original Non-U.S. Models

The table reports the comparison of return predictions between augmented and original machine learning models based on a pooled sample of non-U.S. stocks. The testing sample is from 2006 to 2017. Augmented models are estimated with the data of the 36 stock characteristics (listed in Appendix A), market dummies, US factors, US characteristic gaps, and local factors, while original models are estimated with only stock characteristics and market dummies. The models include neural networks with one to five layers (NN1–NN5). For each model, we calculate the difference of the performance measures (i.e., equal-weighted (EW) or value-weighted (VW) Sharpe Ratio) between the augmented and original model (augmented minus original).

Panel A: Augmented model using all stock characteristics, all US factors, all local factors, and all US characteristic gaps vs. original model

	NN1	NN2	NN3	NN4	NN5	Best NN
Sharpe Ratio (EW)	-0.31	0.15	0.72	0.36	0.83	0.57
Sharpe Ratio (VW)	0.29	0.31	0.66	0.43	0.54	0.54

Panel B: Augmented models using all stock characteristics, top 10 US factors, top 10 local factors, and top 10 US characteristic gaps vs. original model

	NN1	NN2	NN3	NN4	NN5	Best NN
Sharpe Ratio (EW)	0.31	0.73	0.76	0.52	0.93	0.75
Sharpe Ratio (VW)	0.23	0.66	0.94	0.45	0.61	0.68

Table 6. Performance Comparison Between the Augmented and Original Market-Specific Models

The table reports the comparison of return predictions between augmented market-specific and original market-specific machine learning models across the top 25 markets in Appendix B. For each market, *difference* is calculated as the equal-weighted (EW) or value-weighted (VW) Sharpe Ratio based on the augmented market-specific model minus that based on the original market-specific model. Augmented models are estimated with the data of the 36 stock characteristics (listed in Appendix A), US factors, and US characteristic gaps, while original models are estimated with only stock characteristics. The models include neural networks with one to five layers (NN1–NN5). "# of +" refers to the number of markets with a positive value of *difference*, and the fraction of markets with positive *difference* is also reported.

Panel A: Augmented models using all stock characteristics, top 10 US factors, and top 10 US characteristic gaps

		NN1	NN2	NN3	NN4	NN5	Best NN
Sharpe Ratio (EW)	difference	0.08	0.14	0.05	0.13	0.17	0.10
	# of +	16	15	16	14	14	16
	fraction of +	0.64	0.6	0.64	0.56	0.56	0.64
Sharpe Ratio (VW)	difference	0.01	0.15	0.13	0.14	0.37	0.15
	# of +	11	17	17	18	22	18
	fraction of +	0.44	0.68	0.68	0.72	0.88	0.72

Panel B: Augmented models using all stock characteristics, and top 10 US characteristic gaps

		NN1	NN2	NN3	NN4	NN5	Best NN
Sharpe Ratio (EW)	difference	0.10	0.07	0.12	0.16	0.04	0.12
	# of +	13	12	16	16	13	18
	fraction of +	0.52	0.48	0.64	0.64	0.52	0.72
Sharpe Ratio (VW)	difference	0.07	0.17	0.33	0.31	0.42	0.29
	# of +	15	16	17	21	19	22
	fraction of +	0.60	0.64	0.68	0.84	0.76	0.88

Panel C: Augmented models using all stock characteristics, and top 10 US factors

		NN1	NN2	NN3	NN4	NN5	Best NN
Sharpe Ratio (EW)	difference	-0.04	-0.07	0.01	0.02	-0.02	-0.08
	# of +	12	8	10	9	10	7
	fraction of +	0.48	0.32	0.4	0.36	0.4	0.28
Sharpe Ratio (VW)	difference	-0.02	-0.01	0.01	-0.02	0.07	-0.04
	# of +	14	13	15	12	12	12
	fraction of +	0.56	0.52	0.6	0.48	0.48	0.48

Figure 1. Relative Importance of the U.S.-estimated Model

Variable importance for the 36 stock characteristics (listed in Appendix A) in each model in the U.S. market. Rows correspond to individual models, and color gradients within each row indicate the most influential (dark blue) to least influential (red) variables. Variable importances within each model are normalized to sum of one.

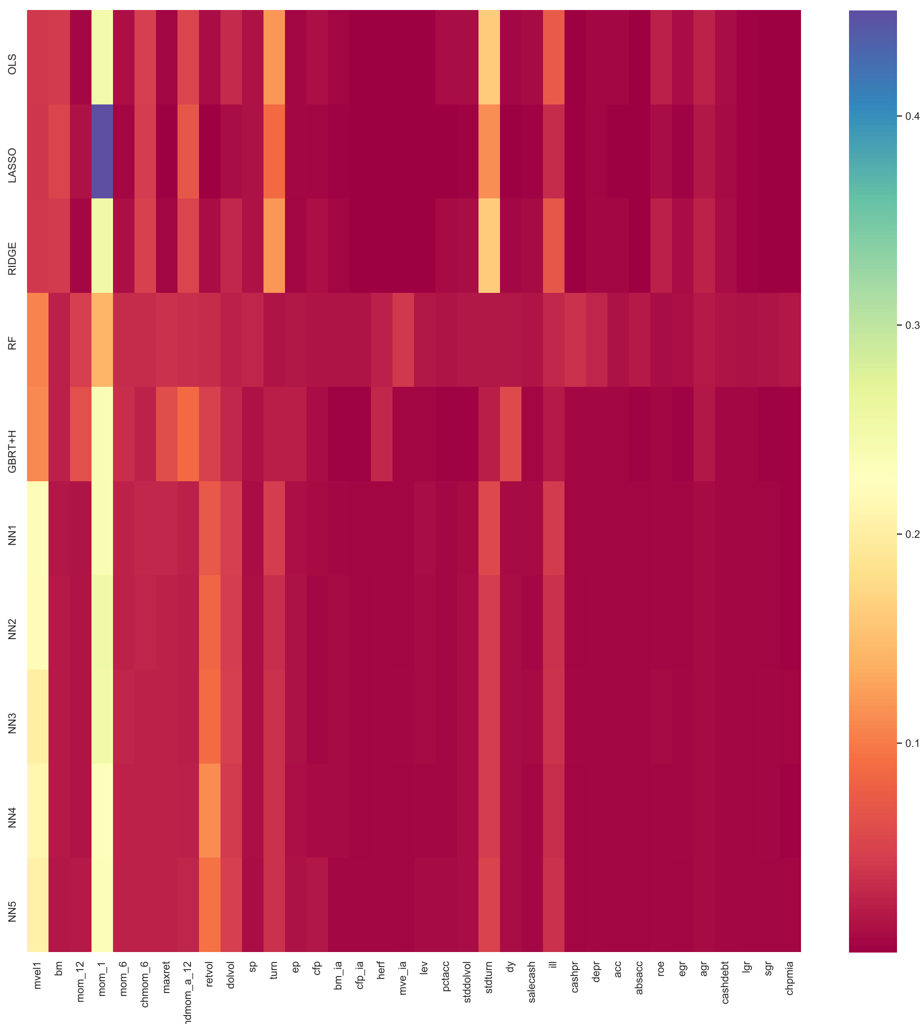
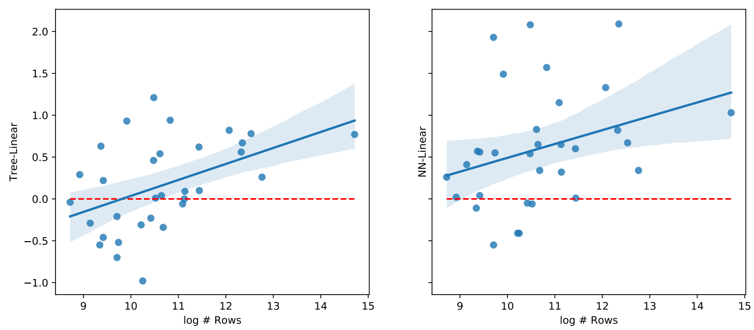
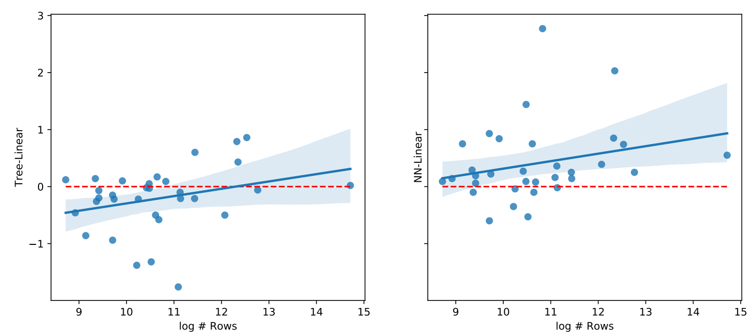


Figure 2. Model Performance and Sample Size

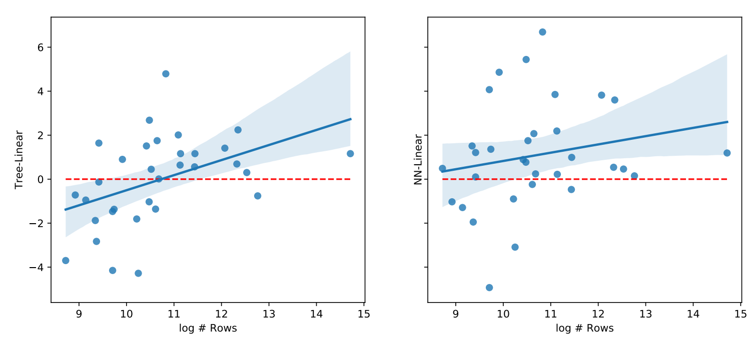
This figure plots the improvements of equal-weighted Sharpe Ratio (Panel (a)), value-weighted Sharpe Ratio (Panel (b)), Rank Correlation (Panel (c)), and Decile Score Distance (Panel (d)) of the best RT and NN models relative to the best linear model against the sample size, with a fitted line with 95% confidence intervals. The horizontal dashed line represents no improvement.



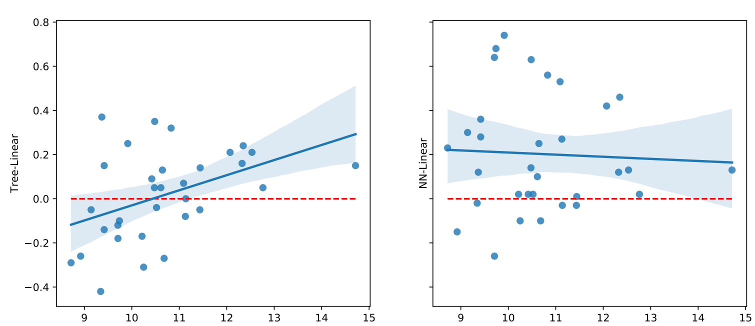
(a) Sharpe Ratio (EW)



(b) Sharpe Ratio (VW)



(c) Rank Correlation



(d) Decile Score Distance

Figure 3. Relative Importance: International Markets

Variable importance for the 36 stock characteristics (listed in Appendix A) in the best performing NN model (based on value-weighted Sharpe Ratio) in each market. Rows correspond to each market, and color gradients within each column indicate the most influential (dark blue) to least influential (red) variables. Variable importances within each market are normalized to sum to one. The figure lists the top 25 markets based on the number of observations.

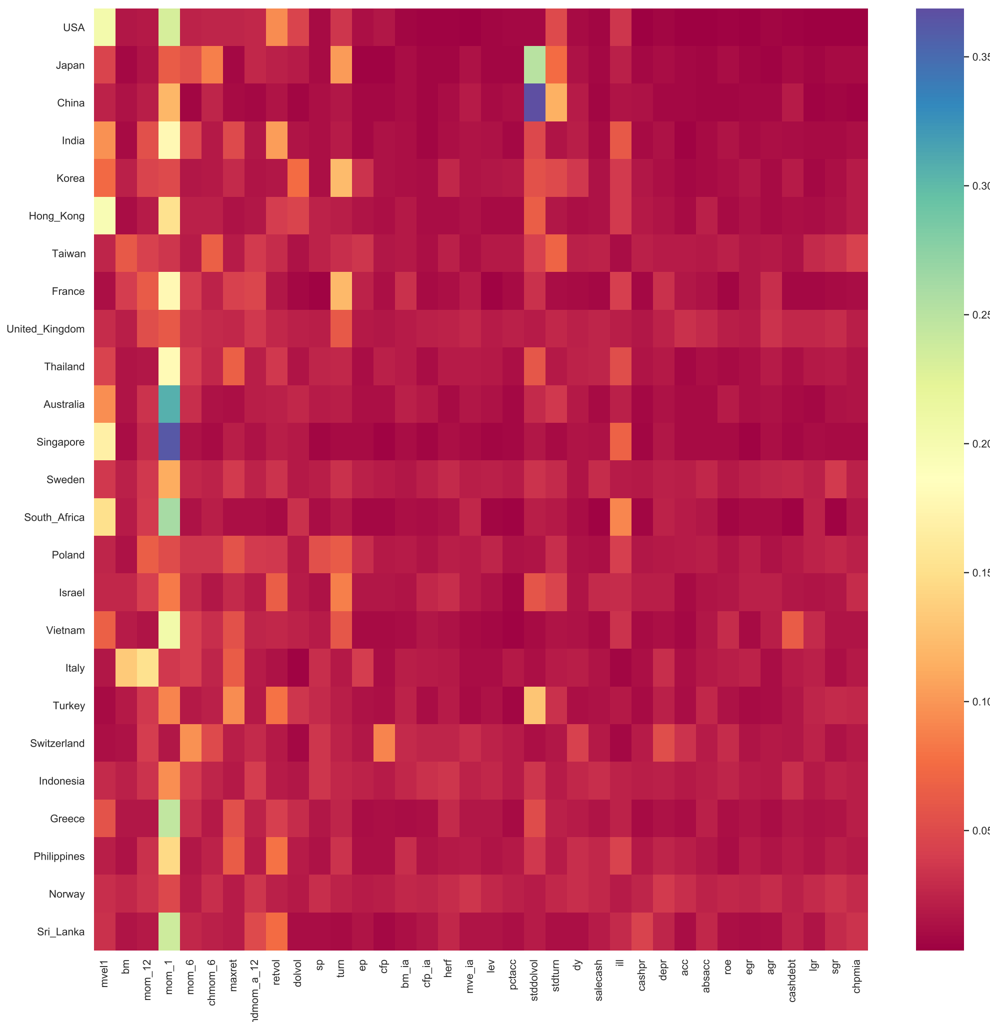
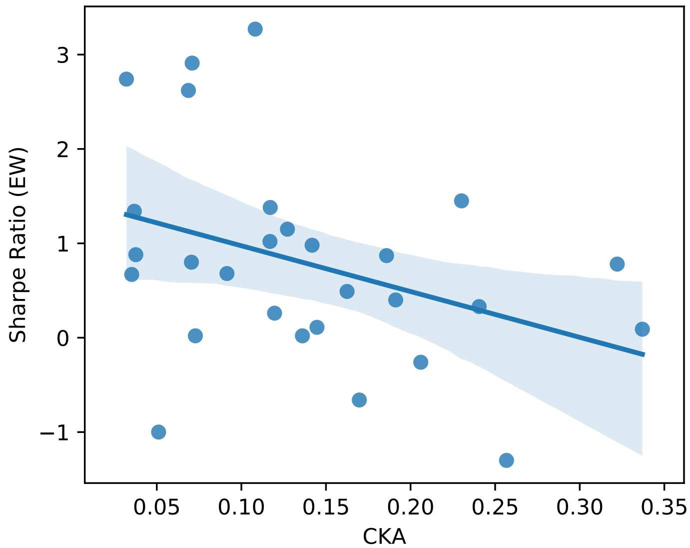
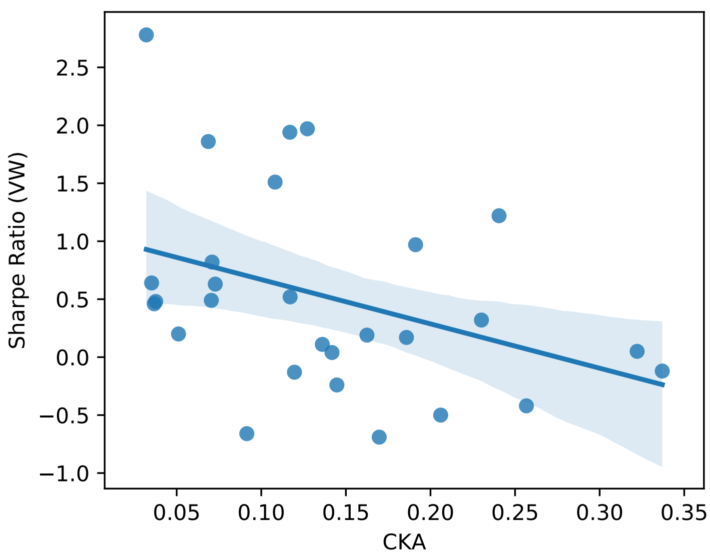


Figure 4. Relations Between Sharpe Ratio Improvements and CKA Similarity

This figure plots the improvements of equal-weighted Sharpe Ratio (Panel (a)) and value-weighted Sharpe Ratio (Panel (b)) from the U.S.-estimated to market-specific NN5 models against the models' CKA similarity, with a fitted line with 95% confidence intervals.



(a) Sharpe Ratio (EW)



(b) Sharpe Ratio (VW)

Figure 5. Group Variable Importance in Augmented NN Models

Group variable importance in the best performing augmented NN model (based on value-weighted Sharpe ratio) using all stock characteristics, top 10 US factors, top 10 local factors, and top 10 US characteristic gaps in a pooled non-U.S. market. Variable importance is normalized to sum to one. Variables are categorized into 4 groups: stock characteristics, US factors, local factors, and US characteristic gaps. We report the sum of variable importance in each variable group.

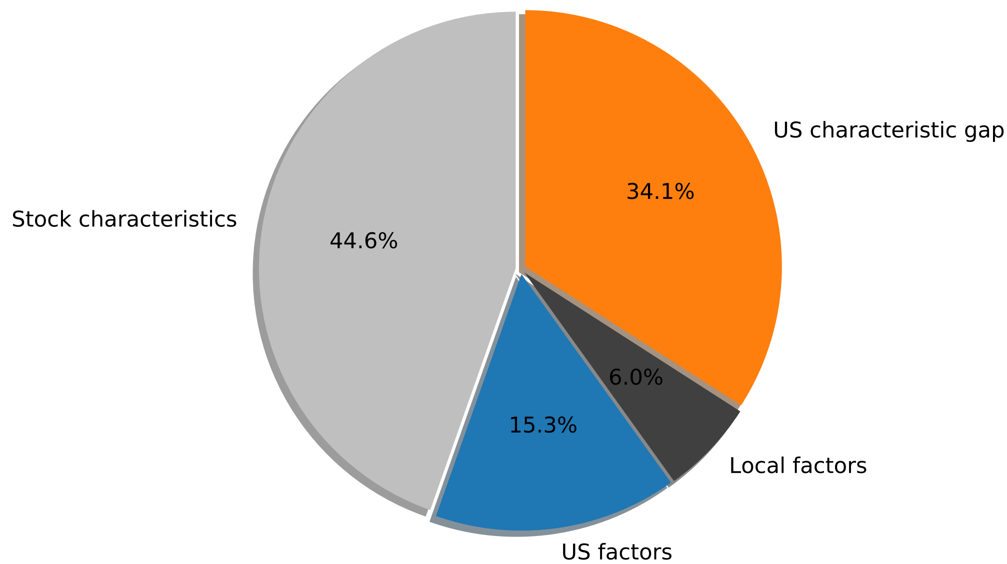
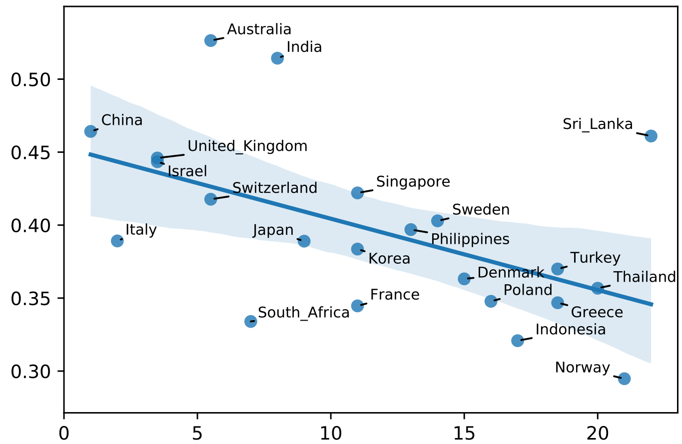
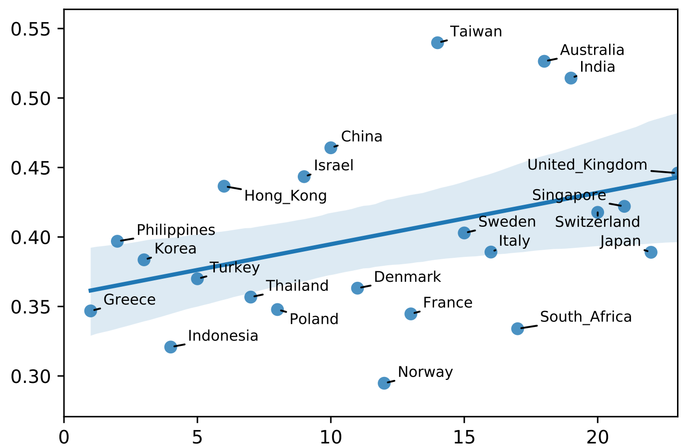


Figure 6. Importance of U.S. Variables and Market Segmentation/Integration Measures

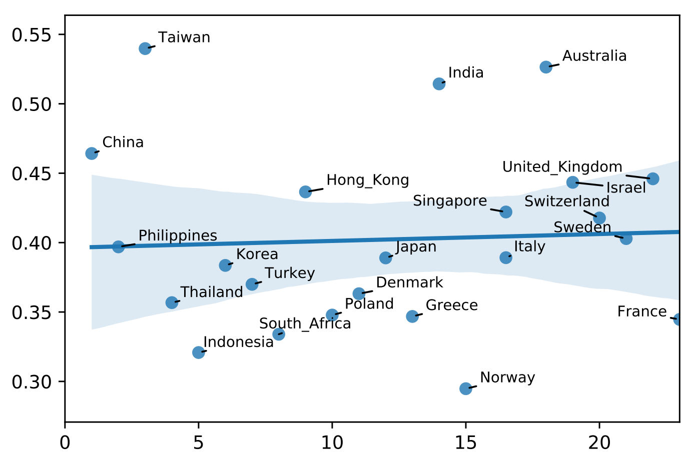
This figure plots the sum of variable importance of U.S. characteristic gaps in the best performing augmented NN model (based on value-weighted Sharpe Ratio) for the top 25 markets in the list of Appendix B against each market's rank of segmentation index in Panel (a), of economic integration index in Panel (b), and of financial integration index in Panel (c).



(a) Segmentation



(b) Economic Integration



(c) Financial Integration