Machine Learning Approaches to Return Prediction and Portfolio Strategies in Global Equity Markets

Author:	Eszter Kósa
SNR:	2139134
ANR:	995500
Supervisor:	dr. Amir Salarkia
Second Reader:	dr. Fatemeh Hosseini Tash
Date:	13th August 2025, Tilburg, Netherlands

Table of Contents

Abstract

This thesis investigates the predictive and economic performance of machine learning (ML) methods for cross-sectional stock return forecasting across six major equity markets: the United States, Germany, the United Kingdom, Japan, China, and France. Using 153 firm-level predictors from the Jensen, Kelly, and Pedersen (2023) global factor library and a rolling-window framework inspired by Gu et al. (2020), the study evaluates Ordinary Least Squares, Lasso, Ridge, Elastic Net, Random Forests, and feedforward neural networks.

Regularized linear models—particularly Ridge and Lasso—consistently outperform both OLS and nonlinear models. Ridge dominates in the U.S., U.K., China, and France, while Lasso leads in Germany; no model performs strongly in Japan. ML-based long-short portfolios generate economically meaningful, factor-robust returns, with Sharpe ratios up to 0.84. Results show that while ML enhances portfolio performance, optimal model choice is market-dependent, underscoring the importance of aligning predictive methods with local market structures.

1.Introduction

The prediction of asset returns has long been a central research subject in empirical finance, with models ranging from the Capital Asset Pricing Model (CAPM) to multi-factor frameworks like those of Fama and French (1993, 2015). While these models provide foundational insights into systematic risk premia, their explanatory power for the cross-section of returns often remains modest, particularly at the individual stock level. This persistent challenge has driven a growing interest in more flexible statistical approaches, especially machine learning (ML) methods, which can accommodate high-dimensional predictor spaces and complex nonlinear relationships. Recent contributions, most notably Gu et al. (2020), demonstrate that ML can improve the out-of-sample (OOS) prediction of returns and enhance portfolio construction strategies in the U.S. market. However, an important open question remains: do these benefits extend to international markets with differing structures, investor bases, and regulatory environments?

This thesis addresses this gap by systematically applying machine learning techniques to forecast returns and construct portfolios across seven major equity markets: the United States, Germany, the United Kingdom, Japan, China, and France. Using a dataset of 153 firm-level characteristics from the Jensen, Kelly, and Pedersen (2023) global factor library, covering the period from 1990 to 2024 (which varies depending on the country since the data coverage is not the same in every case), the study evaluates multiple predictive models—including Ordinary Least Squares (OLS), regularized regressions (Lasso, Ridge, Elastic Net), Random Forests, and feedforward Neural Networks (NN1–NN4)—within a rolling-window framework. By adopting the methodology of Gu et al. (2020) but limiting the predictor space to firm-level variables (excluding macroeconomic interactions and sector dummies), this research highlights the predictive value of cross-sectional characteristics, offering new insights into the economic relevance of these models across diverse markets.

In nearly all markets, regularized linear models (Lasso, Ridge, and Elastic Net) consistently deliver lower Root Mean Squared Errors (RMSEs) and higher OOS R^2 values than both OLS and nonlinear approaches. For instance, in the U.S., Ridge regression achieved the highest long-short spread of 1.55% with a Sharpe ratio of 0.84, while Random Forests and Neural Networks underperformed both in predictive accuracy and portfolio profitability. Similar findings emerge in Germany and China, where Ridge and Lasso strategies dominate, though spreads and Sharpe ratios differ in magnitude, reflecting local market dynamics. Importantly, these performance differences persist even after adjusting returns for well-known

systematic factors using the Fama-French three-factor (FF3), Carhart four-factor (FF4), and Fama-French five-factor (FF5) models, suggesting that the machine learning portfolios capture signals not fully explained by traditional risk premia. However, the explanatory power of these factor models remains low (R^2 typically below 5%), underscoring the unique, idiosyncratic nature of the ML-driven alpha.

The cross-country evidence indicates that model performance is highly context-dependent. For example, while Ridge dominates in the U.S. and China, Lasso leads in Germany, highlighting how structural features—such as liquidity constraints, ownership concentration, and the prevalence of retail versus institutional investors—affect the transferability of predictive models. Furthermore, nonlinear models like Random Forests and Neural Networks, often celebrated for their flexibility, frequently underperform in these low-signal, high-noise environments, likely due to overfitting.

These findings lead to two research questions:

1. **How effectively do machine learning methods predict one-month-ahead stock returns across international equity markets when using firm-level characteristics alone?**
2. **How effectively do machine learning forecasts translate into profitable, risk-adjusted portfolio strategies across international equity markets, and are these gains robust to adjustments for common risk factors?**

By addressing these questions, this thesis contributes to the growing literature on machine learning in asset pricing, extending the conversation beyond U.S.-centric studies. It also offers practical insights for global investors: which models are most effective for cross-sectional stock selection, and how much value do they add after accounting for widely accepted risk factors?

Ultimately, this research bridges the gap between predictive modelling and portfolio construction in a multi-market context, demonstrating that while machine learning holds promise for improving return forecasts, its success depends critically on model choice, predictor selection, and sensitivity to market-specific dynamics.

2.Literature review

2.1. History of Empirical Asset Pricing

Asset pricing is one of the foundational pillars of modern financial theory and its development has gone through several significant theoretical evolutions over the past decades. Initially, the Capital Asset Pricing Model (CAPM) introduced by Sharpe (1964) and Lintner (1965) served as the dominant framework. In line with Markowitz's (1959) portfolio theory, which seeks the optimal trade-off between return and risk, CAPM distinguishes between diversifiable and non-diversifiable sources of risk and provides a theoretical model for expected returns based on systematic risk exposure.

Despite its popularity, CAPM faced substantial empirical criticism, which motivated the development of more comprehensive models. One of the most influential advances was the three-factor model by Fama and French (1993, 1996), which added size (SMB) and value (HML) factors to the market premium. Later, Carhart (1997) extended the model by incorporating momentum, leading to the well-known Carhart four-factor model.

Further refinements were made by Fama and French (2014), who introduced a five-factor model by adding profitability (RMW) and investment (CMA) factors, drawing on insights from the dividend discount model. Their subsequent empirical tests across international markets (Fama & French, 2015) revealed regional differences in factor significance, indicating that some factors may have varying explanatory power depending on market characteristics. In a 2018 (Fama and French, 2018) extension, they also tested the inclusion of momentum in the five-factor framework.

Recent studies have emphasized the context-dependence of factor relevance. For instance, Dirkx and Peter (2020) highlight that factors which are prominent in one country, such as Germany, may not generalize across borders. In emerging markets like Poland's WSE, momentum has shown particularly strong predictive power. Additionally, researchers like Roy and Shijin (2018) proposed the incorporation of human capital as an asset pricing factor, opening new dimensions in explaining return variation.

Together, these developments underline that asset pricing models have evolved into increasingly complex frameworks, reflecting the heterogeneity of market structures and investor behaviour. This growing complexity has also motivated the exploration of machine learning methods, which can flexibly accommodate high-dimensional predictors and nonlinear interactions in the return prediction process.

2.2. Machine Learning in Asset Pricing

As mentioned above, the recent asset pricing literature studies how the usage of Machine Learning methods result in more accurate asset return prediction. Moreover, they try to examine which methods provide the best fits when it comes to asset pricing. In this section I would like to review the most important findings and results in the field of asset pricing with ML

Gu et al. (2020) conduct one of the most comprehensive studies on applying machine learning techniques to the cross-sectional prediction of stock returns. They compare multiple ML algorithms and evaluate how precisely they estimate expected returns, comparing them to traditional linear models. The authors are using a huge dataset (approximately 30000) of U.S. stocks from 1957 to 2016 and incorporating 94 firm-level characteristics, their interactions with 8 macroeconomic variables and 74 industry dummies. Their study included linear methods like lasso, ridge or elastic net; dimension reduction methods such as PCA or PLS; tree-based methods like random forest and gradient boosted trees, neural networks and generalized linear models. Gu et al. (2020) find that machine learning significantly outperforms traditional models, especially tree-based models and neural networks. Additionally, their results suggest that the most informative predictors are the following: price trends such as momentum, short-term reversal or industry momentum; liquidity such as market capitalization, dollar volume or bid-ask spread; volatility such as idiosyncratic volatility or beta. Their work lays a foundation for modern ML-based asset pricing and motivates many subsequent studies, including those that extend the framework to international markets (e.g., Drobetz et al., 2021) and to different asset classes such as bonds (e.g., Feng et al., 2025).

Rapach et al. (2013) study the predictive power of U.S. stock returns for international equity markets, which offers insight into cross-country return forecasting frameworks. With using a monthly dataset covering twelve countries – including Japan, the Netherlands and Germany – they implement out-of-sample forecasting regressions and Granger causality tests to test if lagged U.S. market returns improve return predictions abroad. The authors use both linear forecasting models and compare them to ML methods such as random forests or neural networks. Rapach et al. (2013) find evidence that U.S. returns have signific ant predictive power for foreign market. In contrast, the returns of other countries do not help predict the U.S. returns, which suggests that the U.S. market tends to lead others. Furthermore, they find that machine learning models generally outperform traditional ones.

Moritz and Zimmermann (2016) propose a method called tree-based conditional portfolio sorts, which uses decision tree and random forest techniques to model expected returns in a data-driven way. The authors conduct their analysis using data from the U.S. stock market. With using lagged-return deciles, their empirical results show that recent short-term returns are the strongest predictors of future stock returns, outperforming traditional medium-term momentum indicators.

Giglio and Xiu (2017) address the issue of relevant risk factors being either unobserved or measured with noise. This phenomenon might cause bias in the estimated risk premia in traditional models like the Fama-MacBeth regression. Therefore, they propose a three-pass estimation procedure that combines principal component analysis (PCA) with cross-sectional and time-series regressions. Their results point out that traditional methods tend to overstate the importance of certain macroeconomic predictors, whereas liquidity-related factors retain significance even after correction. This suggests that some widely-used predictors may not capture priced risk as suggested, while others are more robust.

Messmer (2017) examined if deep feedforward neural networks (DFNs) can improve the prediction of cross-sectional stock returns compared to traditional linear models. Focusing on predicting monthly returns based on 68 firm characteristics (such as momentum, book-to-marekt, beta or profitability, the study uses a sample of U.S. large-and mid-cap stocks from 1970 to 2014. Deep feedforward neural networks can flexibly approximate the nonlinear relationships between the firm characteristics and expected returns. Therefore, as the key findings of Messmer's (2017) study show, DFNs outperform traditional linear models when it comes to predictive accuracy. Additionally, the Sharpe ratios of the constructed long-short portfolios are also higher than in the case of traditional models. Considering the characteristics, Messmer (2017) finds that price-based variables, especially short-term reversal and 12-month momentum are the most influential predictors.

Kelly et al (2019) use instrumented principal component analysis (IPCA) when conducting their study. IPCA introduces the innovation of allowing firm characteristics (like size, value or momentum) to function as instruments for estimating time-varying exposures to latent risk factors. Therefore, the model is able to combine insights from cross-sectional anomalies and latent factor models. The authors apply their model to approximately 12000 U.S. stocks from 1962 to 2014 and use 36 firm characteristics, including profitability, leverage, momentum, liquidity and valuation metrics. Kelly et al. (2019) find that IPCA achieves a significantly higher R^2 than the Fama-French 5-factor model. Their key findings about the

characteristics emphasise that only 10 out of 36 matter significantly, which include value-related, momentum and reversal, liquidity and volatility parameters.

Green, Hand, and Zhang (2017) and Harvey, Liu, and Zhu (2016) critique the expansion of return predictors—often called the "factor zoo"—emphasizing the risks of data mining and multiple testing biases in asset pricing research. This criticism led to the development of statistically rigorous testing frameworks, such as the double-selection LASSO method by Feng et al. (2020). Feng et al. (2020) apply a double-selection LASSO method that extends the Fama-MacBeth regression so that it can be used in high-dimensional settings. The authors address the problem of the rapidly increasing number of proposed pricing factors (factor zoo) and the question whether the factors actually add explanatory power for expected returns. After conducting two LASSO selections and one post-selection OLS regression, the authors present that many new factors are redundant. However, factors such as profitability, investment or Quality Minus Junk still provide statistically significant explanatory power, even after controlling for the factor zoo.

In addition to the literature from the past five years, Pelger (2020) estimates time-varying systematic risk factors by applying principal component analysis (PCA). Their data consists of high-frequency (5-minute) returns from S&P 500 stocks using Trade and Quote millisecond data. Pelger (2020) splits the returns into two types: continuous returns and jump returns, where jump returns capture large, sudden movements throughout the trading day. Thus, he can treat the different returns as two separate sources of risk. Moreover, he builds two different covariance matrices (one for continuous returns and one for jump returns) to see how the stocks move together. When extracting the most important risk factors from the continuous covariance matrix, he finds four dominant factors: market factor, energy-related factor, financial factor and sector-specific factor. The jump covariance matrix shows only one significant component, which can be interpreted as system-wide risk factor that causes movement in all stocks' prices at once (e.g., central bank announcements or unexpected interest rate changes).

In 2021 Dobretz et al. performed an asset pricing analysis on the European equity markets, using data from 19 Eurozone countries between 1990 and 2020. They filtered their choice of firms based on their market capitalizations; the minimum market cap had to be €25M. When building their models, Dobretz et al. (2021) used 22 firm-level characteristics (e.g., size, book-to-market, momentum, ROA or profitability). The applied machine learning methods consisted of elastic net, principal component regression, partial least squares, random forests, gradient boosted trees, neural networks and support vector machines. After conducting their

study, Dobretz et a. (2021) find that neural networks and tree-based methods provide the best out-of-sample predictions and that the ordinary least squares method underperforms due to overfitting. This paper further supports the previous discoveries that ML-based top decile portfolios earn higher Sharpe ratios and risk-adjusted returns than traditional models. The most informative predictors were observed to be momentum, earnings-to-price and book-to-market, however, their importance seemed to vary across models.

Bianchi et al. (2021) conduct a study in which they explore how machine learning methods can improve the prediction of U.S. treasury bond risk premiums, more precisely: the excess returns earned by holding longer-term bonds over short term ones. Their dataset covers U.S. government bonds (1-to 5-yar maturities), 134 macro predictors, yield curve variables and latent factors. With applying several machine learning models (elastic net, random forests, principal component regression, partial least squares and neural networks) they find that ML methods significantly outperform linear models, especially elastic net and random forests. Furthermore, the used methods discover that the predictors have time-varying importance. For instance, inflation-related variables are more important in times of recession. Additionally, the ML forecast based trading strategies managed to yield higher Sharpe ratios and certainty equivalent returns than traditional models.

Feng et al. (2025) further investigate the prediction of individual corporate bond returns with applying machine learning methods to forecast them. Their dataset covers both public and private bonds from 1976 to 2020. In their framework the authors use 26 bond-level characteristics, such as credit spread, duration, downside risk and return skewness. Moreover, they incorporate 25 macroeconomic and market-wide predictors, including GDP growth, VIX or corporate bond excess returns. In their evaluation they apply various ML models: lasso, ridge, elastic net, partial least squares, principal component analysis, random forest, boosted trees and neural networks. The key findings reveal that ML models once again outperform traditional methods (Fama-French three and five factor models), with random forest being the best performer. Moreover, the ML strategies generated considerable investment value. Besides identifying the key predictors (e.g., HML, GDP growth, corporate bond excess returns, downside risk, short-term reversal), they also examine that predictability is stronger during periods of high risk aversion and low economic activity.

2.3. Machine Learning Portfolios

Gu et al. (2020) show that machine learning–based long–short decile portfolios deliver stronger performance than those formed from traditional linear models. In their U.S. equity sample from 1957–2016, the best-performing ML methods—particularly gradient boosted trees and deep neural networks—achieve out-of-sample monthly alphas of roughly 0.5–0.6% and annualized Sharpe ratios around 1.2–1.4, compared to Sharpe ratios below 0.8 for standard OLS benchmarks. These excess returns remain economically and statistically significant after controlling for Fama–French and momentum factors, suggesting that ML models identify predictive structure in the cross-section of returns beyond what is captured by conventional factor models.

3.Methodology

3.1. Overarching model

Based on to Gu et al.'s (2020) paper, the statistical model describing the general functional form for risk premium predictions (in each examined country) is:

$$\mathbf{r_{(i,t+1)} = E_t(r_{(i,t+1)}) + \epsilon_{(i,t+1)},}$$

where the predictor function is the following:

$$\mathbf{E_t(r_{(i,t+1)}) = g^*(z_{(i,t)}).}$$

The asset's excess return can be described as an additive prediction error model. In the equation, $r_{(i,t+1)}$ is the excess return for stock i at time t+1, E_t ($r_{(i,t+1)}$) is the conditional expected return, modeled by a function g* ($z_{(i,t)}$) of predictors and $\epsilon_{(i,t+1)}$ is the prediction error term. The vector $z_{(i,t)}$ is a vector of stock-level predictors (e.g., firm size, book-to-market ratio or momentum). To ensure consistency and to leverage a large amount of data for risk premium estimation, g* applies uniformly across all stocks and time periods. Moreover, to reduce the risk of overfitting, the model does not utilize prior return history of stock i or the characteristics of other stocks and therefore avoids complex dependencies.

Furthermore, Gu et al. (2020) construct machine learning-based portfolios to assess the economic value of their return predictions. This strategy involves ranking stocks based on predicted returns and forming long-short decile portfolios. This way I would be able to evaluate their performance based on their Sharpe ratios and see the predictive power of the ML models.

The objective of this study is to extend the framework of Gu, Kelly, and Xiu (2020) – who conducted their study on the U.S. equity market – to broader international context. By applying similar Machine Learning models to global (both developed and emerging) markets,

my study aims to test if the predictive models can reveal universal pricing signals or if their performance is unforeseen on local market dynamics.

First, I describe the data sources and the firm-level characteristics I am using based on the WRDS Global Stock Returns and Characteristics dataset. Second, I present my sample splitting methods and then I introduce the implemented machine learning algorithms. These include linear regression (OLS), regularized regressions (Lasso, Ridge, Elastic Net), tree-based models (Random Forests and Gradient Boosted Trees), and feedforward neural networks. The models - varying from linear and additive ones to non-linear and highly flexible ones - are chosen to demonstrate the diversity of statistical patterns. Lastly, I describe the model performance evaluation methods, using both statistical and economic metrics. Out-of-sample R^2 and mean squared error (MSE) are measuring predictive accuracy.

3.2. Portfolio Construction and Evaluation

In addition to the test of machine learning models, I assess the economic value of machine learning forecasts by constructing and then evaluating value-weighted decile portfolios, following Gu et al.'s (2020) methodology. This step helps us to examine whether the predicted excess returns from each model could be employed to form profitable investment strategies. Therefore, in each country, for each machine learning model and rolling window I build out-of-sample decile and long-short portfolios based on predicted excess returns and evaluate their average realized and predicted returns and Sharpe-ratios.

First, I collect the out-of-sample predictions and the corresponding realized excess returns for every rolling window for all firms in each test period. Then, using the models' predictions I sort the stocks into ten deciles based on their predicted 1-month ahead excess returns. I assess the stocks with the lowest predicted returns to the first decile, while in the tenth decile there are stocks with the highest predicted returns. This decile assignment is performed cross-sectionally for each month.

Second, for each decile and date I compute value-weighted average realized excess returns, using each firm's lagged market capitalization as portfolio weights. This yields ten time series of monthly returns for each decile-sorted portfolio. To evaluate the performance of the signals from each model, I form a long-short strategy that goes long in Decile 10 and short in Decile 1. Thus, I compute the excess return of the long-short portfolio at each point in time as the difference between the two portfolios' value-weighted returns.

Lastly, I evaluate the performance of the portfolio by calculating the Sharpe ratio at annual frequency:

$$\textbf{Sharpe ratio}_{\textbf{annual}} = \frac{\boldsymbol{Mean}\,(r_t)}{\boldsymbol{SD}(r_t)}\boldsymbol{x}\sqrt{\mathbf{12}}$$

Where r_t denotes the monthly excess return of the long-short portfolio in month *t*. Sharpe ratio captures the return per unit of risk and serves as a standard measure of risk-adjusted performance.

To further evaluate the economic significance of the machine learning forecasts, I conduct time-series regression of the long-short (High-Low) decile portfolios on the academically well-known asset pricing models: the Fama-French three-factor model (FF3), the Carhart four-factor model (FF3 + MOM), and the Fama-French five factor model (FF5). In each rolling window I estimate these regressions and then average the results across windows. For each model, I estimate:

$$\boldsymbol{r_t^{LS} = \alpha + \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 MOM_t + \beta_5 RMW_t + \beta_6 CMA_t + \epsilon_t}$$

Where r_t^{LS}represents the monthly return of the long-short portfolio, and the factors vary depending on the chosen model.

The intercept (alpha=α) in these regressions measures the risk-adjusted abnormal return of the long-short portfolio that is not explained by the included factors. A statistically significant and positive alpha indicates that the machine learning portfolios generate returns beyond the well-known risk premia, providing evidence of their economic value. Contrarily, a low or insignificant alpha suggests that the observed returns are largely explained by standard risk exposures rather than unique insights from the ML models.

3.2. Data Sources and Variable Construction

The study is conducted on stock-level panel data spanning January 1990 to December 2024. The data was extracted from the WRDS Global Stock Returns and Characteristics database. The analysis covers seven major equity markets – Japan, Germany, the United Kingdom, China and France – and applies firm-specific predictors derived from JKP factor library.

I construct the predictive models of 1-month ahead excess returns using 153 firm-level characteristics from the JKP factor library (Jaenson, Kelly and Pedersen, 2023). All data was extracted from the WRDS Global Stock Returns and Characteristics dataset. The dependent variable is the 1-month ahead excess return, *ret_exc_lead1m*. According to the documentation, this variable is computed as the forward-looking return (from price at time t to t+1) minus the risk-free rate. The characteristics are organized into economic categories such as Valuation (e.g., book-to-market, dividend yield), Profitability (e.g., return on assets, gross profitability),

Investment (e.g., asset growth, accruals), Financing and issuance (e.g., equity issuance, debt issuance), Price trends and momentum (e.g., short-and long-term past returns), Risk and frictions (e.g., market beta, idiosyncratic volatility) or Intangibles.

To mitigate the impact of outliers while keeping most of the observations the predictors were winsorized at the 1st and the 99th percentiles in each test window. This ensures that the excessively large or small predictor and return values – which may be due to data errors or rare market events – do not distort model training or evaluation. Moreover, by winsorizing the data, the model robustness is improved.

Furthermore, each month, all firm characteristics are cross-sectionally normalized into the [-1,1] interval, following Gu et al.'s (2020) methodology. For each predictor and time period, I compute the cross-sectional rank of each observation, and map these ranks to the [-1,1] interval using the transformation:

$$x_{i,t}^{norm} = 2 \times \frac{rank(x_{i,t})}{N_t} - 1$$

where $rank(x_{i,t})$ denotes the rank of firm I's predictor value at time t, and N_t is the number of firms with non-missing observations in period t. This transformation ensures that each variable scale-free and robust to outliers which facilitates efficient learning across different model architectures.

3.3. Sample Splitting

To evaluate model performance in a way that it reflects the real-time forecasting challenges faced by investors, I use a rolling window sampling scheme to generate multiple train, validation and test splits. By this approach it is ensured that model training, hyperparameter tuning and evaluation are all performed strictly out-of-sample and in a temporally consistent manner and therefore look-ahead bias is avoided.

Each rolling window is constructed based on the following structure:

- Training set: 5 years (60 months),
- Validation set: 1 year (12 months),
- Test set: 1 year (12 months).

The rolling windows advance by 1 year (12 months) at each step. In the first window, I train the models using data from months t=1 to t=60, validate data on months t=61 to t=72 and evaluate it on months t=73 to t=84 and then shift it forward to the second window by 12 months. For each window the training set is used to fit the model, the validation set is used to select

hyperparameters and the test set is used for out-of-sample performance evaluation including predictive accuracy and economic metrics.

3.4. Machine Learning Models

3.4.1. Linear

Ordinary Least Squares (OLS) regression is the classical linear modelling approach, which I widely used in empirical asset pricing to relate excess returns to firm characteristics:

$$r_{i,t+1} = \beta_0 + \sum_{j=1}^{p} x_{i,t}^{(j)} \beta_j + \varepsilon_{i,t+1}$$

Where $r_{i,t+1}$ denotes the realized excess return for firm i at time $t+1$, $x_{i,t}^{(j)}$ is the j-th predictor (in this case a firm characteristic) observed at time t, and $\varepsilon_{i,t+1}$ is the error term. The OLS estimator minimizes the sum of squared errors and serves as a benchmark for evaluating the predictive power of other, more complex models. OLS is prone to overfitting in high-dimensional settings and highly sensitive to multicollinearity, which are both common in financial data. However, in the paper of Green et al. (2017) it is shown, that many firm-level predictors are correlated, which can undermine the stability of OLS estimates.

3.4.2. Ridge Regression

Traditional linear models such as Ordinary Least Squares (OLS) have a higher tendency to perform poorly out-of-sample due to overfitting, multicollinearity, and noise accumulation when it comes to high-dimensional asset pricing. Regularization methods use penalties on the model coefficients to ensure balance in the trade-off between bias and variance, which later guarantees more stable and generalizable predictions. To answer my research question, I applied three of the most widely used regularized linear models: Ridge Regression, Lasso Regression and Elastic Net.

Ridge Regression (Hoerl and Kennard, 1970) modifies the OLS objective by introducing an L_2 penalty on the size of the coefficients. Formally the ridge estimator solves:

$$\hat{\beta}^{ridge} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} (r_{i,t+1} - x'_{i,t}\beta)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}$$

where $\lambda \geq 0$ is the regularization parameter controlling the degree of shrinkage. Increasing λ means that the model penalizes large coefficients more heavily, while shrinking them toward zero without eliminating any variable entirely.

From an economic perspective, when many characteristics contribute marginally to return prediction (so the signal is believed to be dense), ridge regression is particularly useful. Ridge

regression reduces the multicollinearity problem appearing in high-dimensional settings by stabilizing the inversion of the X'X matrix and reducing model variance. However, ridge regression keeps all coefficients in the model, since it does not perform variable selection. This might mean a drawback if the true signal is concentrated in a small subset of variables.

3.4.3. Lasso Regression

The Least Absolute Shrinkage and Selection Operator (Lasso), introduced by Tibshirani (1996), replaces the L_2 penalty with an L_1 penalty:

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} (r_{i,t+1} - x'_{i,t}\beta)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

The key feature of Lasso is that it shrinks some coefficients exactly to zero. Therefore, it can effectively perform variable selection. In high-dimensional asset pricing, where there are many predictors, this property makes Lasso an appealing model for identifying a subset of relevant characteristics. For example, the idea that only a small number of firm fundamentals are truly priced by the market justifies Lasso's selective nature. Still, Lasso faces plenty of limitations, since it may behave irregularly when predictors are highly correlated, selecting one variable from a group, even when all are informative.

3.4.4. Elastic Net

Zou and Hastie (2005) proposed the Elastic Net to combine the strength of Ridge and Lasso while also mitigating the individual weaknesses of them. The Elastic Net uses both L_2 and L_1 penalties to the loss function:

$$\hat{\beta}^{EN} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} (r_{i,t+1} - x'_{i,t}\beta)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right\}$$

While encouraging grouped variable selection, Elastic Net allows correlated predictors to enter the model jointly. In empirical asset pricing many economically similar variables are highly correlated, therefore this feature is especially valuable.

Empirical Implementation in R

When applying the regularized linear models – Ridge, Lasso, and Elastic Net – I implemented a rolling-window regression framework in R using the *glmnet* package. This methodology aligns with Gu et al.'s (2020) forecasting-based design, ensuring a separation between training, validation and test sets. Each regularized model was trained using the *cv.glmnet()* function, which performs k-fold (in this case: default, 10 -fold) cross-validation to select the optimal penalty parameter λ:

- Ridge: set alpha = 0 for an L_2 penalty,
- Lasso: set alpha = 1 for an L_1 penalty,
- Elastic Net: iteratively searched across values of $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, selecting a combination of α and λ that minimized the validation RMSE.

3.4.5. Random Forests

Random Forests (Breiman, 2001) is a nonparametric machine learning technique which manages to aggregate predictions based on multiple randomized regression trees. In the case of empirical asset pricing, where predictor-response relationships are often nonlinear and high-dimensional, random forests can be well-suited and a good choice.

Traditional linear models assume additive and linear relationships between the predictors and the (excess) returns. This assumption can cause failure in capturing complex interactions and threshold effects that are present in the real markets. Random Forests allow nonlinear and non-additive effects amongst characteristics. Moreover, they capture interactions without requiring explicit specification, while providing robustness to overfitting through averaging across trees. Both Gu et al. (2020) and Dobretz et al. (2021) demonstrated that they are particularly effective in the context of cross-sectional return prediction.

The algorithmic design of Random Forests is the following:

- *B* individual regression trees, each trained on a bootstrap sample of the training data,
- At each node split, a random subset of predictors (of size *mtry*) is considered,
- Predictions are averaged across trees, reducing variance and enhancing stability.

Mathematically, the forest prediction for firm *i* in month t + 1 is:

$$\widehat{r_{i,t+1}^{RF}} = \frac{1}{B}\sum_{b=1}^{B} T_b(z_{i,t})$$

Where $T_b(.)$ denotes the b[th] decision tree and $z_{i,t}$ is the vector of lagged characteristics.

In addition to their many strengths, Random Forests also have limitations: it is computationally intensive for large *B* and *p* values, their results are harder to interpret globally, and the importance of the variables can be biased toward high-variance variables.

Empirical implementation in R

Random Forests were implemented using the *randomForest* package in R. The training data was subsampled at 5% (*sampsize = 0.05 x N*) to enhance speed without degrading performance. The *maxnodes* parameter was set to 256, limiting tree depth to avoid overfitting.

I evaluated RMSE across multiple values of *ntree* (number of trees: 50, 100, 200, 300) during model tuning.

3.4.6. Feedforward Neural Networks (1-4 layers)

Feedforward Neural Networks capture nonlinearities and complex interactions between firm-level characteristics. In this paper I used one- to four-layer fully connected NNs. A standard feedforward neural network consists of an input layer, one or more hidden layers of nonlinear transformations, and an output layer. For return prediction, the network learns to approximate the conditional expectation function:

$$\hat{r}_{i,t+1} = f_\theta(z_{i,t})$$

where $z_{i,t}$ is the vector of firm characteristics at time t and $f_\theta$is the neural network with parameters θ learned through stochastic gradient descent. Each hidden layer performs a transformation of the form:

$$a^{(l)} = \phi(W^{(l)}a^{(l-1)} + b^{(l)})$$

where $W^{(l)}, b^{(l)}$ are weights and biases for layer l, ϕ is the activation function (ReLU in my paper) and $a^{(0)} = z_{i,t}$, the input features.

3.4.7. Machine Learning Model Evaluation Metrics

The predictive performance of each model is assessed by employing evaluation metrics such as out-of-sample R-squared (R^2), and Root Mean Squared Error (RMSE).

The R-squared statistic measures the proportion of variance in future excess returns explained by the model's predictions. Higher out-of-sample explanatory power is indicated by a higher R-squared, however in asset pricing R^2 values tend to be negative due to high noise-to-signal ratios in returns.

The RMSE values demonstrate the average magnitude of prediction errors, with lower RMSEs indicating that the model's predicted returns are closer on average to the actual realized returns. When comparing models in terms of prediction accuracy, RMSE is especially useful.

4.Empirical results

4.1. USA

To evaluate the predictive performance of machine learning models in the case of the U.S. equity market, I analysed a large panel of data with 1,668,162 monthly firm-level observations. To ensure the robustness, I excluded all characteristics from the JKP factor set

which had more than 50% of missing data and filtered out the micro-cap stocks with a market equity below USD 300,000. Furthermore, I imputed the remaining missing values with cross-sectional medians. After the data-cleaning I had 142 predictors to work with.

4.1.1. Model Performance Evaluation

The out-of-sample (OOS) performance of the employed models is presented in Table 1. As the table reports, the OOS R^2 values across different algorithms remain negative. The highest R^2 was reached by Elastic Net ($\alpha = 0.5$) and the lowest value was -7.22% yielded by the two-layer Neural Network model (NN2). The root mean squared errors (RMSE) correspond with the OOS R^2 values ranging between 0.1255 and 0.1282.

Table 1-Out-of-Sample R-Squared and RMSE values for the U.S. equity market, data collected from WRDS database

Machine Learning Method	R^2	RMSE
OLS	-2.31%	0.1256
Lasso	-2.31%	0.1256
Ridge	-2.06 %	0.1255
ENet (α=0.5)	-2.05%	0.1255
RF	-3.22%	0.1262
NN1	-4.89%	0.1270
NN2	-7.22%	0.1282
NN3	-5.08%	0.1271
NN4	-5.61%	0.1273

However, we can claim that the predictive power is low on individual stock level, there are patterns to observe in the model performance. First, the regularized linear models, such as Ridge and Elastic net outperform OLS, which suggests that the regularization mitigates overfitting in the case of a high-dimensional predictor set. This aligns with Gu et al.'s (2020) findings which suggest that penalization is a critical component of return prediction models. Second, the implemented nonlinear models such as Random Forests and Neural Networks underperform relative to regularized linear models. This underperformance might be due the low signal-to-noise ratio in monthly returns, when more flexible models tend to capture noise rather than systematic structure (Harvey et al., 2016). As shown in Table 1, one-layer Neural Network outperforms the deeper algorithms, which supports Gu et al.'s (2020) result that shallow networks provide better out-of-sample performance.

4.1.2. Variable Importance

Figure 1 depicts the top 20 predictors identified by the variable importance analysis of Random Forest. The dominant predictors include *div12m_me*, which is the Dividend-to-Market

Equity ratio and measures the total dividends paid over the past 12 months scaled by the firm's market equity and therefore captures shareholder payout intensity. In addition to Dividend-to-Market Equity ratio, the second most important payout and value characteristic is *eqpo_me* (equity net payout scaled by market equity). Both Lintner (1956) and Fama and French (1993) found that higher payout yields to higher returns and that dividends and repurchases could serve as signals of firm quality. The variable *ni_inc8q* (net income including extraordinary items, aggregated over eight quarters) came out as the third most influential characteristic, which emphasizes that sustained earnings performance could determine the future returns of the company. Academic literature such as Novy-Marx (2013) and Fama and French (2015) both suggest that quality and profitability factors impact future returns. Momentum and seasonality are captured by the predictors *ret_1_0*, *ret_3_1*, *ret_6_1*, *ret_9_1* and *ret_12_1*. The presence of these factors show that the momentum effect is dominant when it comes to predicting returns.

In conclusion, the findings are in contrast with Gu et al.'s (2020) results, where momentum, liquidity and volatility are reported as the dominant categories in the U.S. equity market. The difference lies in the composition of the predictor sets; Gu et al (2020) construct their variable set from approximately 900 signals, combining 94 firm-level characteristics with their interactions with 8 macroeconomic time-series variables and 74 industry sector dummies. This paper focuses on the 153 firm characteristics derived from the Jensen, Kelly and Pedersen (2023) global factor library. The limitation of the predictor set to purely firm-level characteristics and excluding macroeconomic interactions and sector dummies lets us examine the predictive power of cross-sectional stock level information.

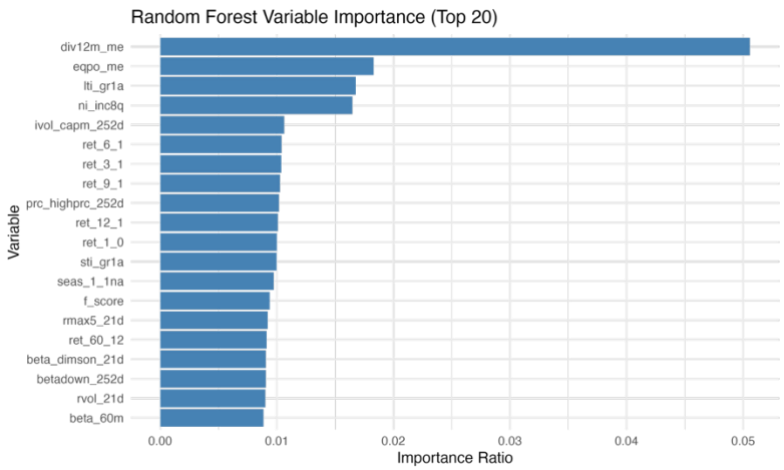

Figure 1-Random Forest normalized feature importance ratios - the 20 highest values in the case of the U.S.

4.1.3. Portfolio Analysis

With constructing value-weighted decile portfolios I was able to evaluate the economic value of machine learning portfolios, shown in Table 2. The linear regressions such as OLS, Lasso, Ridge and Elastic Net produced economically meaningful long-short spreads. Out of the four models, Ridge regression achieved the highest average (monthly) realised long-short return of 1.55% with an annualized Sharpe ratio of 0.84. The second-best performing machine learning portfolio was captured by OLS with a monthly average realized return of 1.37% and an annualized out-of-sample Sharpe ratio of 0.93. Elastic Net delivered the third highest return of 1.05% (SR=0.72) followed by Lasso with a spread of 1.02% and an annualized out-of-sample Sharpe ratio of 0.70. According to Gu et al.'s (2020) study, the best performing linear method is Lasso, therefore these results are in contrast with the mentioned literature.

Moving on to nonlinear models, we can claim that they underperform all linear models. Random Forests generate a long-short return of 0.95% and an annualized Sharpe ratio of 0.53. The Neural Network algorithms (NN1-NN4) yield spreads between 0.24% (NN2) and 0.89% (NN4) and their Sharpe ratios range from 0.19 to 0.71. Compared to Gu et al's (2020) findings, we can exhibit different portfolio performances, since their results claim that neural networks and tree-based models outperform linear approaches. This could be due to the predictor space being limited to firm-level predictors without macroeconomic and interaction terms.

The performance of the portfolios is consistent with the out-of-sample evaluation of the machine learning models. Presented in Table1, it is visible that Ridge managed to get the least negative R^2 value (-2.06%) and managed to yield the highest portfolio return (1.55%). The consistency is further supported by the underperformance of portfolio-returns of nonlinear models: Random Forests and Neural Networks both exhibited more negative R^2 values (-3.22% and -7.22% for NN2, respectively) and yielded the lower portfolio returns than linear models (0.95% for RF and 0.24% for NN2).

Table 2 - Performance of the machine learning models in the case of the U.S.

	OLS				Lasso				Ridge			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	-0.42%	-1.35%	6.89%	-0.21	0.05%	-0.29%	6.86%	0.02	-0.44%	-0.28%	7.97%	-0.19
2	0.39%	-0.51%	5.48%	0.24	0.34%	0.08%	6.34%	0.19	0.18%	0.10%	7.06%	0.09
3	0.50%	-0.06%	5.09%	0.34	0.49%	0.29%	5.31%	0.32	0.21%	0.31%	6.02%	0.12
4	0.71%	0.27%	4.70%	0.53	0.41%	0.45%	4.96%	0.28	0.51%	0.47%	5.39%	0.33
5	0.66%	0.56%	4.73%	0.49	0.77%	0.59%	4.89%	0.55	0.55%	0.60%	5.03%	0.38
6	0.75%	0.83%	4.52%	0.58	0.55%	0.73%	4.67%	0.41	0.76%	0.73%	4.84%	0.54
7	0.82%	1.10%	4.73%	0.60	0.70%	0.86%	4.67%	0.52	0.88%	0.85%	4.63%	0.66
8	0.78%	1.39%	4.60%	0.59	0.68%	0.99%	4.79%	0.49	0.83%	0.98%	4.47%	0.64
9	0.88%	1.75%	5.13%	0.60	0.90%	1.16%	4.74%	0.66	0.91%	1.14%	4.41%	0.72
High	0.95%	2.41%	5.71%	0.58	1.07%	1.45%	5.31%	0.70	1.11%	1.42%	4.71%	0.82
High-Low	**1.37%**	**3.77%**	**5.10%**	**0.93**	**1.02%**	**1.73%**	**5.04%**	**0.70**	**1.55%**	**1.70%**	**6.39%**	**0.84**
	ENet (α=0.5)				**RF**				**NN1**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	0.08%	-0.26%	6.91%	0.04	-0.13%	-0.89%	7.63%	-0.06	0.01%	-3.25%	6.18%	0.01
2	0.40%	0.09%	6.42%	0.22	0.38%	-0.02%	6.13%	0.22	0.53%	-1.63%	5.24%	0.35
3	0.51%	0.30%	5.37%	0.33	0.51%	0.35%	5.30%	0.33	0.70%	-0.80%	4.72%	0.51
4	0.47%	0.46%	5.09%	0.32	0.58%	0.61%	4.84%	0.42	0.57%	-0.17%	4.55%	0.44
5	0.71%	0.60%	4.78%	0.51	0.67%	0.83%	4.87%	0.47	0.76%	0.37%	4.51%	0.58
6	0.58%	0.72%	4.78%	0.42	0.82%	1.11%	4.85%	0.59	0.66%	0.88%	4.60%	0.50
7	0.71%	0.85%	4.60%	0.53	0.82%	1.50%	5.13%	0.55	0.72%	1.40%	4.86%	0.51
8	0.75%	0.98%	4.72%	0.55	0.86%	1.82%	5.33%	0.56	0.81%	1.98%	4.82%	0.58
9	0.81%	1.14%	4.76%	0.59	0.81%	2.20%	5.59%	0.50	0.79%	2.71%	5.29%	0.52
High	1.13%	1.42%	5.16%	0.76	0.82%	3.03%	7.34%	0.39	0.89%	4.08%	5.97%	0.52
High-Low	**1.05%**	**1.69%**	**5.06%**	**0.72**	**0.95%**	**3.93%**	**6.21%**	**0.53**	**0.88%**	**7.33%**	**4.29%**	**0.71**
	NN2				**NN3**				**NN4**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	0.46%	-3.97%	6.60%	0.24	0.12%	-3.31%	7.21%	0.06	-0.02%	-3.77%	7.92%	-0.01
2	0.52%	-1.65%	5.11%	0.36	0.37%	-1.12%	5.51%	0.23	0.33%	-1.31%	5.48%	0.21
3	0.72%	-0.71%	4.84%	0.52	0.50%	-0.32%	4.83%	0.36	0.46%	-0.44%	4.56%	0.35
4	0.71%	-0.05%	4.60%	0.54	0.60%	0.19%	4.54%	0.46	0.67%	0.11%	4.27%	0.54
5	0.78%	0.50%	4.52%	0.59	0.75%	0.58%	4.46%	0.58	0.76%	0.53%	4.26%	0.62
6	0.75%	1.01%	4.55%	0.57	0.64%	0.95%	4.49%	0.50	0.64%	0.92%	4.39%	0.50
7	0.63%	1.54%	4.54%	0.48	0.72%	1.32%	4.58%	0.55	0.77%	1.32%	4.66%	0.57
8	0.64%	2.16%	5.14%	0.43	0.86%	1.75%	4.61%	0.65	0.90%	1.81%	5.03%	0.62
9	0.81%	3.01%	5.55%	0.50	0.88%	2.37%	5.39%	0.57	0.81%	2.54%	5.66%	0.49
High	0.69%	4.99%	6.71%	0.36	0.96%	3.97%	6.62%	0.50	0.88%	4.45%	7.17%	0.42
High-Low	**0.24%**	**8.96%**	**4.42%**	**0.19**	**0.84%**	**7.28%**	**4.29%**	**0.68**	**0.89%**	**8.22%**	**4.90%**	**0.63**

In this table, I report the performance of the prediction-sorted, value-weighted decile portfolios over the out-of-sample testing period. All stocks are sorted into deciles based on their predicted returns for the next month. Columns "Pred" provides the predicted monthly returns for each decile, "Avg" provides the average realized monthly returns, "SD" provides the standard deviations and "SR" reports the Sharpe ratios. This follows Gu et al.'s (2020) methodology.

4.1.4. Factor Model Regressions

Table 3 reports the statistics of the Fama-French Three-Factor (FF3), the Carhart (FF3+MOM) and the Fama-French 5 Factor (FF5) model specifications in the case of the United States. Ridge regression achieved an alpha of 1.6% (t=4.58), OLS's alpha is 1.4% (t=4.84) and Lasso produced an alpha of 1.1% (t=4.14). These results suggest that the employed machine learning strategies can generate economically meaningful abnormal returns beyond what is captured by the market, size and value factors. The R^2 values of the regressions are fairly low, with NN1 having the lowest (0.36%) and OLS having the highest (1.24%), meaning

that only about 0.36–1.24% of the month-to-month movements in the long–short portfolio returns are explained by Fama–French factors. Looking at the results of the Carhart (1997) model, which includes the momentum factor in addition to the FF3 model. The intercepts (alphas) are slightly reduced but not completely eliminated: Ridge's alpha declines from 1.6% to 1.5% (t=4.24) and OLS yields only 1.2% (t=4.75) in contrast to the prior 1.4%. These changes imply that the inclusion of the momentum factor strengthen the explanatory power of return premiums. The results are consistent with the reported importance of the variables, where both short- and medium-term predictors (*ret_1_0*, *ret_6_1* etc.) turned out to be in the top 20 performing variables. The alphas remain statistically significant, which means that the models extract signals beyond the momentum an FF3 exposures. The regression analysis of the Fama-French Five-Factor model (FF5), where profitability and investment factors are added to the original FF3 model shows that Ridge's alpha further declines to 1.4% (t=4.18) and OLS yields an intercept of 1.3% (t=4.08). Considering these results and that the alphas remain statistically significant, we can claim that a part of the returns remain unexplained by the FF5 factors.

Overall, the low explanatory power (R^2 values) of the models means that machine learning strategies can generate alpha that is independent of the traditional risk factors. Moreover, the machine learning long-short decile portfolios produce statistically significant abnormal returns that are robust to standard factor adjustments in the U.S. equity market.

Table 3 - Risk-adjusted performance of machine learning portfolios in the case of the U.S.

		OLS	Lasso	Ridge	ENet (α=0.5)	RF	NN1	NN2	NN3	NN4
FF3	mean_ret	1.37%	1.02%	1.55%	1.05%	0.95%	0.88%	0.15%	0.84%	0.89%
	α	0.014	0.011	0.016	0.012	0.010	0.009	0.003	0.009	0.009
	t (α)	4.84	4.14	4.58	4.18	3.08	3.69	1.19	3.85	3.44
	R^2	1.24%	2.04%	0.36%	1.00%	0.51%	1.18%	1.40%	0.10%	0.21%
Carhart	mean_ret	1.37%	1.02%	1.55%	1.05%	0.95%	0.88%	0.15%	0.84%	0.89%
	α	0.014	0.011	0.015	0.012	0.009	0.010	0.003	0.009	0.009
	t (α)	4.75	3.80	4.24	3.83	2.77	4.00	1.17	3.81	3.33
	R^2	1.70%	2.04%	0.62%	1.10%	2.13%	3.26%	1.42%	0.24%	0.91%
FF5	mean_ret	1.37%	1.02%	1.55%	1.05%	0.95%	0.88%	0.15%	0.84%	0.89%
	α	0.013	0.011	0.014	0.011	0.010	0.009	0.003	0.008	0.009
	t (α)	4.08	3.86	4.18	3.86	3.17	3.33	1.09	3.47	3.09
	R^2	2.05%	2.24%	2.61%	2.02%	0.55%	2.36%	1.73%	0.39%	0.24%

In this table, I report the average monthly returns, the alphas and the R^2 values with respect to the Fama-French Three-Factor, the Carhart and the Fama-French Five-Factor models.

4.2. Germany

After excluding variables with more than 50% of missing observations and imputing missing values with cross-sectional medians I obtained a final sample with 225,458 monthly firm-level observations and 124 characteristics from the JKP variable set.

4.2.1. Model Performance Evaluation

Table 4 - Out-of-Sample R-squared and RMSE values for the German equity market, data collected from WRDS database

Machine Learning Method	R^2	RMSE
OLS	-1.36%	0.1666
Lasso	-1.04%	0.1724
Ridge	-1.08%	0.1724
ENet (α=0.5)	-1.06%	0.1724
RF	-3.41%	0.1727
NN1	-7.98%	0.1778
NN2	-10.89%	0.1803
NN3	-7.51%	0.1775
NN4	-6.63%	0.1768

Table 4 shows the out-of-sample predictive performance of the machine learning models for the German equity market. The linear models (OLS, Lasso, Ridge, Elastic Net) exhibit stronger performance than the nonlinear ones. While OLS has the lowest RMSE value (0.1666), Lasso achieved the highest, but still negative R^2 value of -1.04%. Since all R^2 values across models remain negative, we an suggest that the models do not outperform a simple historical mean forecast when it comes to explaining cross-sectional variation in one-month-ahead retrns. The nonlinear approaches such as Random Forest (RF) algorithms or Neural Networks (NN1-NN4) underperform compared to the linear ones. RF deliver an OOS RMSE of 0.1727 and an OOS R^2 of -3.41%, while Neural Networks perform even worse with RMSE values ranging from 0.1768 (NN4) to 0.1803 (NN2) and R^2 values being between -10.89% (NN2) and -6.63% (NN4). Consequently, we can conclude that greater model complexity does not necessarily mean better predictive accuracy in a low-signal environment such as the German equity market. These results align with the ones observed while analysing the U.S. equity market, since linear regularized models provide more stable OOS predictions than nonlinear ones in the German market as well.

4.2.2. Variable Importance

Figure 2 visualises the importance ratios of the top 20 most influential predictors identified on the German equity market. The most dominant predictor is *ret_1_0* which is the 1-month return variable. This captures the short-term momentum effect and is consistent with the results documented by Jegadeesh and Titman (1993). *Bidaskhl_21d* (21-day bid-ask spread) serves as a proxy for liquidity and trading costs and emphasises the pricing of illiquidity risk in German equities. The presence of predictor *age* (firm age) in the most influential variables indicates that firm maturity contributes to return predictability, due to older firms yielding more

stable earnings and being less exposed to idiosyncratic risk. Other significant predictors are *ret_3_1* and *ret_6_1* (3- and 6-month momentum), which confirms that momentum works across different time horizons. Lastly, another important variable is *seas_1_1na* capturing seasonality. The factor picks up on calendar-based return patterns suggesting that seasonality is present in the German equity market.

In conclusion, we can claim that the predictive signals in the German market are mostly controlled by momentum across multiple horizons and liquidity-related variables. The presence of the bid-ask measure shows that both trading costs and liquidity play a more significant role in pricing German stocks than U.S. equities.

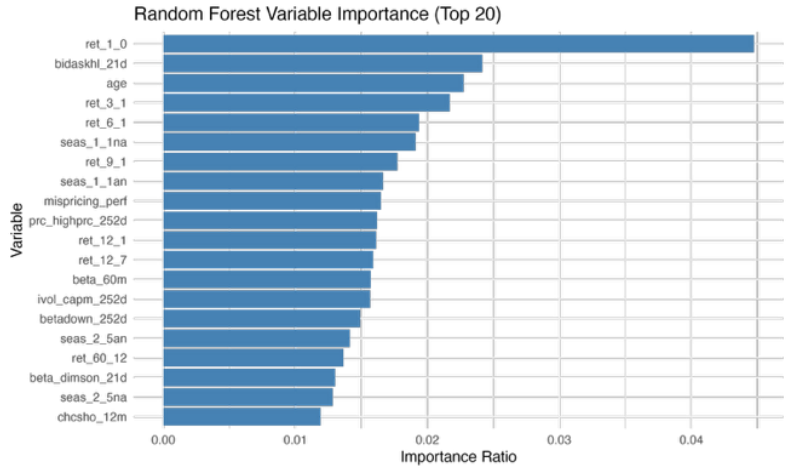

Figure 2 - Random Forest normalized feature importance ratios - the 20 highest values in the case of Germany

4.2.3. Portfolio Analysis

The performance of the value-weighted decile portfolios is documented in Table 5. As the table reports, we can see that Lasso delivers the highest spread of 1.98% with an annualized OOS Sharpe ratio of 0.82. This is followed by Ridge with a realized average monthly return of 1.89% and a corresponding Sharpe ratio of 0.80. These models outperform the other two linear ones, since Elastic Net yields a spread of 1.78% (SR=0.72) and OLS reports an average monthly return of 1.73% for the High-Low long-short portfolio with an annualized Sharpe ratio of 0.71.

Moving on to the nonlinear methods, it is visible that Random Forest algorithms performed poorly in comparison with the linear methods. RF produced a spread of 0.28% with an annualized Sharpe ratio of only 0.13. The deep learning methods (NN1-NN4) managed to yield long-short spreads between 0.20% (NN3) and 0.95% (NN2) with annualized Sharpe ratios ranging from 0.11 (NN3) to 0.61 (NN2).

Table 5 - Performance of the machine learning portfolios in the case of Germany

	OLS				Lasso				Ridge			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	-1.23%	-3.01%	7.93%	-0.54	-1.51%	-2.26%	8.75%	-0.60	-1.36%	-2.19%	8.05%	-0.59
2	0.18%	-1.57%	7.30%	0.09	0.22%	-1.12%	7.39%	0.10	0.04%	-1.07%	6.66%	0.02
3	0.07%	-0.78%	7.03%	0.03	0.20%	-0.48%	6.56%	0.10	0.36%	-0.47%	6.84%	0.18
4	0.59%	-0.18%	6.67%	0.31	0.26%	0.01%	6.56%	0.14	0.25%	0.00%	6.70%	0.13
5	0.48%	0.34%	6.45%	0.26	0.49%	0.41%	6.96%	0.24	0.27%	0.39%	6.29%	0.15
6	0.49%	0.82%	6.53%	0.26	0.60%	0.79%	6.57%	0.32	0.56%	0.76%	6.29%	0.31
7	0.51%	1.32%	6.83%	0.26	0.62%	1.16%	6.68%	0.32	0.46%	1.12%	6.68%	0.24
8	0.50%	1.86%	7.02%	0.25	0.37%	1.57%	6.43%	0.20	0.64%	1.53%	6.70%	0.33
9	0.53%	2.52%	6.66%	0.28	0.78%	2.06%	6.75%	0.40	0.67%	2.02%	6.88%	0.34
High	0.49%	3.58%	7.39%	0.23	0.47%	2.81%	7.22%	0.23	0.52%	2.82%	7.43%	0.24
High-Low	**1.73%**	**1.73%**	**8.44%**	**0.71**	**1.98%**	**5.07%**	**8.39%**	**0.82**	**1.89%**	**5.01%**	**8.22%**	**0.80**
	ENet (α=0.5)				**RF**				**NN1**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	-1.48%	-2.31%	8.43%	-0.61	0.23%	-2.06%	9.48%	0.08	0.00%	-7.93%	6.24%	0.00
2	0.25%	-1.15%	7.43%	0.12	0.58%	-0.91%	7.51%	0.27	0.45%	-4.57%	6.38%	0.25
3	0.15%	-0.50%	6.55%	0.08	0.21%	-0.34%	6.56%	0.11	0.34%	-2.67%	6.71%	0.17
4	0.12%	-0.01%	6.41%	0.06	0.71%	0.05%	6.25%	0.40	0.31%	-1.17%	6.64%	0.16
5	0.47%	0.40%	6.99%	0.23	0.65%	0.38%	6.08%	0.37	0.02%	0.14%	6.79%	0.01
6	0.52%	0.79%	6.17%	0.29	0.81%	0.70%	5.87%	0.48	0.56%	1.38%	6.40%	0.30
7	0.61%	1.17%	6.80%	0.31	0.95%	1.03%	5.91%	0.56	0.47%	2.67%	6.60%	0.25
8	0.27%	1.58%	6.64%	0.14	0.86%	1.44%	6.66%	0.45	0.71%	4.14%	7.07%	0.35
9	0.75%	2.09%	6.74%	0.38	0.53%	2.09%	6.60%	0.28	0.34%	6.00%	6.92%	0.17
High	0.31%	2.88%	7.43%	0.14	0.51%	3.51%	7.21%	0.25	0.56%	9.60%	7.38%	0.26
High-Low	**1.78%**	**5.19%**	**8.62%**	**0.72**	**0.28%**	**5.58%**	**7.58%**	**0.13**	**0.56%**	**17.53%**	**4.88%**	**0.40**
	NN2				**NN3**				**NN4**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	0.08%	-9.80%	7.23%	0.04	0.20%	-7.56%	8.52%	0.08	-0.21%	-7.11%	7.98%	-0.09
2	0.30%	-5.82%	6.49%	0.16	0.59%	-3.78%	6.91%	0.30	0.29%	-3.59%	7.12%	0.14
3	0.84%	-3.65%	6.29%	0.46	0.24%	-1.94%	6.45%	0.13	0.64%	-1.96%	6.30%	0.35
4	0.32%	-2.05%	6.97%	0.16	0.49%	-0.72%	6.46%	0.26	0.11%	-0.86%	6.59%	0.06
5	0.54%	-0.70%	6.94%	0.27	0.68%	0.30%	6.18%	0.38	0.49%	0.07%	6.64%	0.25
6	0.31%	0.61%	6.43%	0.16	0.07%	1.20%	6.45%	0.04	0.62%	0.92%	6.64%	0.32
7	0.05%	1.96%	6.78%	0.03	0.55%	2.20%	6.57%	0.29	0.46%	1.83%	6.56%	0.24
8	0.60%	3.52%	6.86%	0.30	0.57%	3.39%	6.99%	0.28	0.07%	2.93%	6.78%	0.04
9	0.25%	5.62%	7.23%	0.12	0.48%	5.07%	6.97%	0.24	0.66%	4.61%	7.43%	0.31
High	1.04%	9.66%	7.87%	0.46	0.41%	8.63%	7.97%	0.18	0.39%	8.08%	9.23%	0.15
High-Low	**0.95%**	**19.46%**	**5.46%**	**0.61**	**0.20%**	**16.19%**	**6.39%**	**0.11**	**0.60%**	**15.19%**	**8.09%**	**0.26**

In this table, I report the performance of the prediction-sorted, value-weighted decile portfolios over the out-of-sample testing period. All stocks are sorted into deciles based on their predicted returns for the next month. Columns "Pred" provides the predicted monthly returns for each decile, "Avg" provides the average realized monthly returns, "SD" provides the standard deviations and "SR" reports the Sharpe ratios. This follows Gu et al.'s (2020) methodology.

High-decile portfolios managed to earn spreads between 0.31% and 1.04% per month on average, while low-decile portfolios mostly produced negative or near-zero returns. However, as Table 7 shows, the low overall explanatory power of the models resulted in the stock not always being correctly sorted into the corresponding decile. Overall, the portfolio performance aligns with the OOS model-level results provided in Table 4, where linear models had the lowest RMSE and least negative R^2 values.

4.2.4. Factor Model Regressions

Reported in Table 6, it is visible that under the FF3 model the long-short portfolios retain positive and statistically significant alphas across the linear methods. Lasso provides the highest alpha value of 2% (t=3.47) followed by Ridge with an alpha of 1.9% (t=3.27) and OLS at 1.8% (t=2.91). This suggests that the implemented machine learning strategies generate economically meaningful abnormal returns that are not captured by the market, size and value factors. The R^2 values of the linear models vary from 1.43% (Lasso) to 2.19% (Elastic Net) showing that the three Fama-French factors explain only a very small part of the return variance. For the nonlinear models, the performance of the portfolios is much weaker. For instance, Random Forests generate a slightly negative alpha of -0.7% (t=-0.88), while Neural Networks produce near-zero or insignificant alphas. This implies that the more complex models can not extract stable, factor-adjusted signals and in some cases, they are value destroying after accounting for the standard risk factors.

In the case of the Carhart model, also presented in Table 6, when the momentum factor is added to the FF3 regression, the alphas slightly decrease but remain statistically significant for the linear models. To begin with, Lasso's alpha drops slightly to 1.9% (t=3.18) while the alpha of Ridge remains 1,9% (t=3.21). This implies that their performance, especially Lasso's, was probably linked to momentum exposure. The alpha of OLS regression (1.9%, t=2.95) grows and the priorly also negative alpha of RF slightly increases to -0.6%, which means that those portfolios' returns are less tied to the simple momentum effects. Overall, the results suggest that momentum only explains a small portion of the performance for the regularized linear models but adds a little explanatory value for OLS and RF portfolios.

Continuing the analysis with the Fama French Five Factor (FF5) model, Table 6 shows the results of the regressions when we added profitability and investment factors. It is visible that the alphas of the linear models (Lasso, Ridge and OLS) maintain alphas, which lets us believe that these new factors do not fully explain the returns. Random Forest continues to show a weak alpha value and Neural Networks remain inconsistent. This suggests that the linear models are capturing return components beyond the expanded factor set, while the nonlinear models fail to deliver robust, factor-adjusted excess returns.

Table 6 - Risk-adjusted performance of machine learning portfolios in the case of Germany

		OLS	Lasso	Ridge	ENet (α=0.5)	RF	NN1	NN2	NN3	NN4
FF3	mean_ret	1.73%	1.98%	1.89%	1.78%	0.28%	0.56%	0.95%	0.20%	0.60%
	α	0.018	0.020	0.019	0.018	-0.007	0.006	0.010	0.001	0.006
	t (α)	2.91	3.47	3.27	2.99	-0.88	1.88	2.36	0.18	1.08
	R^2	1.64%	1.43%	1.59%	2.19%	9.56%	2.78%	2.93%	6.72%	0.30%
Carhart	mean_ret	1.73%	1.98%	1.89%	1.78%	0.28%	0.56%	0.95%	0.20%	0.60%
	α	0.019	0.019	0.020	0.018	-0.006	0.006	0.011	0.001	0.008
	t (α)	2.95	3.18	3.21	2.81	-0.70	1.80	2.53	0.36	1.18
	R^2	1.83%	1.58%	1.68%	2.20%	9.68%	2.84%	3.99%	6.91%	1.43%
FF5	mean_ret	1.73%	1.98%	1.89%	1.78%	0.28%	0.56%	0.95%	0.20%	0.60%
	α	0.018	0.019	0.018	0.016	-0.009	0.007	0.009	-0.002	0.005
	t (α)	3.28	3.38	3.40	2.85	-1.21	1.85	1.88	-0.31	0.95
	R^2	4.99%	2.35%	3.82%	3.90%	13.03%	2.85%	6.47%	7.46%	6.42%

In this table, I report the average monthly returns, the alphas and the R^2 values with respect to the Fama-French Three-Factor, the Carhart and the Fama-French Five-Factor models.

4.3. China

After excluding variables with more than 50% of missing observations and imputing missing values with cross-sectional medians I obtained a final sample with 641,979 monthly firm-level observations and 145 characteristics from the JKP variable set.

4.3.1. Model Performance Evaluation

Generally, after looking at the results in Table 7 we can say that the models performed rather poorly. The best performing models' out-of-sample R^2 values are still very low: Lasso, Ridge and Elastic Net have R^2 values of approximately -4.77%, -4.78% and 4.78% respectively. The third best performing model was OLS, with producing an R^2 of -5.23%. When looking at RMSEs, all three regularized linear models generate values of approximately 0.1286, followed closely by OLS's RMSE of 0.1288.

Nonlinear models once again perform considerably worse. Random Forest exhibits a much larger decrease in explanatory power with an R^2 of -14.71% (RMSE 0.1337). Neural Networks perform the weakest overall, with NN2 and NN4 producing R^2 values of -16.56% and -17.05%, respectively and RMSEs above 0.1349. These results imply that the employed nonlinear models fail to generalize effectively in the Chinese market, likely because of the presence of higher market noise and unique structural features of Chinese equities.

Table 7 - Out-of-Sample R-Squared and RMSE values for the Chinese equity market, data collected from the WRDS database

Machine Learning Method	R^2	RMSE
OLS	-5.23%	0.1288
Lasso	-4.77%	0.1286
Ridge	-4.78%	0.1286
ENet (α=0.5)	-4.78%	0.1286
RF	-14.71%	0.1337
NN1	-12.87%	0.1330
NN2	-16.56%	0.1349
NN3	-13.81%	0.1334
NN4	-17.05%	0.1351

4.3.2. Variable Importance

Figure 3 visualises the twenty most significant predictors in the Chinese equity market. The most dominant factor turned out to be the variable *age* (firm age), which indicates that younger firms in China behave differently from mature firms in terms of return patterns. The predictors *eqnetis_at*, *eqnpo_me*, *chcsho_12m* capture net equity issuance, equity payouts and share count changes and therefore imply the strong presence of equity financing and payout activity. Thus, in the Chinese equity market the financing decisions are highly informative for returns. The presence of the variables *rd_me* and *rd_sale* underscores that investment in innovation pays a role in predicting expected returns. This statement may reflect both investor optimism for high-R&D firms and associated risk premia. Profitability measures also made it to the top twenty most influential predictors: *netis_at* and *ni_inc8q* are net income scaled by assets and earnings persistence over eight quarters. These highlight the significance of stable and improving profitability in predicting cross-sectional returns.

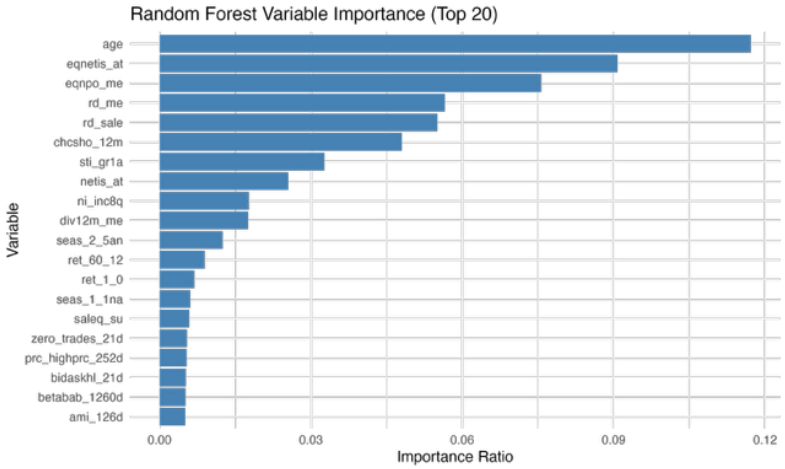

Figure 3 - Random Forest normalized feature importance ratios - the 20 highest values in the case of China

Overall, the results suggest that Chinese stock return predictability is driven by several firm fundamentals such as profitability, R&D or earnings persistence, by financing behaviour (issuance and payouts) or price-based signals such as momentum (*ret_1_0*, *ret_60_12*) or seasonality (*seas_1_1na*, *seas_2_5an*). In China's market, secondary offerings and share issuance are very frequent, which means that the liquidity and financing variables are likely capturing the effects of share sales on ownership structure and how investors respond to firms raising new capital (Chen and Chen, 2007).

4.3.3. Portfolio Analysis

The performance of the machine learning portfolios in the Chinese equity market is presented by Table 8. As exhibited, among the linear models Ridge delivers the highest long-short spread at 1.98% with an annualized out-of-sample Sharpe ratio of 1.09. This is followed by Lasso and Elastic Net with both having long-short spreads of 1.89% and annualized Sharpe ratios of 0.96. OLS is slightly behind the regularized linear models with an average monthly realized return of 1.79% and a Sharpe ratio of 1.05. The performance of the regularized linear models underscores their effectiveness in ranking Chinese equities, with Ridge delivering the best approach in this market.

Once again, the nonlinear models underperform. To begin with, Random Forest performs poorly by generating a negative long-short spread of -0.12% with a Sharpe ratio of -0.06, implying limited economic value in its forecasts. The deep-learning methods (NN1-NN4) exhibit spreads between 0.64% (NN3) and 0.87% (NN1) with Sharpe ratios ranging from 0.60 to 0.73 and therefore still being far below Ridge and Lasso.

The High-decile portfolios are earning around 0.87% and 0.91% per month in the case of the linear models and thus show a clear pattern. Additionally, the low-decile portfolios produce negative or near-zero returns which further supports the mentioned structure. Meanwhile, Random Forest and Neural Networks fail to produce competitive long-short spreads, further supporting the conclusion that more complex nonlinear methods are less robust in the Chinese context. Both higher market noise and structural market features could challenge generalization for these models.

In conclusion, the presented results successfully mirror the out-of-sample model performance metrics, where Ridge and Lasso had the lowest RMSEs and the less negative R^2 values.

Table 8 - Performance of the machine learning portfolios in the case of China

	OLS				Lasso				Ridge			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	-0.87%	-1.79%	7.65%	-0.40	-0.96%	-0.55%	7.89%	-0.42	-1.07%	-0.63%	8.06%	-0.46
2	-0.28%	-0.68%	7.06%	-0.14	-0.26%	0.04%	7.35%	-0.12	-0.30%	0.00%	7.53%	-0.14
3	-0.06%	-0.09%	7.06%	-0.03	0.02%	0.37%	7.33%	0.01	-0.10%	0.34%	7.30%	-0.05
4	0.31%	0.37%	7.31%	0.15	0.17%	0.63%	7.25%	0.08	0.21%	0.61%	7.41%	0.10
5	0.08%	0.78%	7.14%	0.04	0.29%	0.87%	7.30%	0.14	0.31%	0.85%	7.05%	0.15
6	0.24%	1.17%	7.04%	0.12	0.24%	1.09%	7.42%	0.11	0.38%	1.09%	7.21%	0.18
7	0.52%	1.57%	7.55%	0.24	0.37%	1.31%	7.61%	0.17	0.37%	1.32%	7.10%	0.18
8	0.66%	2.00%	7.63%	0.30	0.53%	1.55%	7.79%	0.24	0.39%	1.58%	7.33%	0.18
9	0.85%	2.52%	7.75%	0.38	0.75%	1.84%	7.78%	0.33	0.62%	1.88%	7.68%	0.28
High	0.91%	3.51%	8.13%	0.39	0.92%	2.29%	8.22%	0.39	0.91%	2.40%	7.72%	0.41
High-Low	**1.79%**	**5.30%**	**5.91%**	**1.05**	**1.89%**	**2.84%**	**6.80%**	**0.96**	**1.98%**	**3.02%**	**6.32%**	**1.09**
	ENet (α=0.5)				**RF**				**NN1**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	-0.99%	-0.59%	7.84%	-0.44	-0.09%	-2.30%	7.76%	-0.04	-0.39%	-5.63%	7.65%	-0.18
2	-0.28%	0.02%	7.28%	-0.13	0.09%	-1.03%	7.59%	0.04	-0.01%	-2.89%	6.90%	0.00
3	0.00%	0.35%	7.40%	0.00	0.00%	-0.31%	7.45%	0.00	0.10%	-1.52%	6.92%	0.05
4	0.13%	0.62%	7.27%	0.06	0.28%	0.31%	7.15%	0.14	0.13%	-0.47%	7.03%	0.06
5	0.31%	0.87%	7.32%	0.15	0.33%	0.79%	7.06%	0.16	0.15%	0.47%	7.14%	0.07
6	0.23%	1.09%	7.44%	0.11	0.21%	1.23%	6.90%	0.10	0.22%	1.38%	7.03%	0.11
7	0.36%	1.32%	7.59%	0.17	0.05%	1.67%	7.03%	0.02	0.29%	2.31%	7.05%	0.14
8	0.58%	1.57%	7.71%	0.26	0.09%	2.13%	7.10%	0.04	0.38%	3.35%	7.21%	0.18
9	0.78%	1.86%	7.91%	0.34	-0.01%	2.69%	7.33%	0.00	0.41%	4.68%	7.45%	0.19
High	0.90%	2.33%	8.27%	0.38	-0.21%	3.83%	7.47%	-0.10	0.48%	7.28%	7.74%	0.21
High-Low	**1.89%**	**2.91%**	**6.81%**	**0.96**	**-0.12%**	**6.13%**	**4.95%**	**-0.09**	**0.87%**	**12.91%**	**4.08%**	**0.73**
	NN2				**NN3**				**NN4**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	-0.47%	-7.11%	7.80%	-0.21	-0.43%	-5.65%	7.44%	-0.20	-0.43%	-7.08%	7.21%	-0.21
2	-0.12%	-3.64%	7.10%	-0.06	-0.03%	-2.55%	7.09%	-0.01	-0.13%	-3.46%	6.76%	-0.06
3	-0.09%	-2.05%	7.13%	-0.05	0.05%	-1.21%	7.22%	0.02	-0.02%	-1.97%	6.93%	-0.01
4	0.19%	-0.85%	6.98%	0.09	0.00%	-0.24%	6.95%	0.00	0.22%	-0.88%	6.98%	0.11
5	0.26%	0.18%	7.12%	0.13	0.13%	0.60%	7.04%	0.07	0.23%	0.05%	7.25%	0.11
6	0.31%	1.16%	7.17%	0.15	0.21%	1.38%	7.12%	0.10	0.27%	0.94%	7.12%	0.13
7	0.36%	2.19%	7.26%	0.17	0.43%	2.21%	6.90%	0.22	0.24%	1.87%	7.20%	0.12
8	0.36%	3.33%	6.86%	0.18	0.47%	3.16%	7.16%	0.23	0.40%	2.91%	7.25%	0.19
9	0.42%	4.84%	7.14%	0.20	0.51%	4.45%	7.49%	0.24	0.46%	4.33%	7.52%	0.21
High	0.22%	8.13%	7.37%	0.11	0.21%	7.57%	7.66%	0.09	0.27%	7.88%	7.78%	0.12
High-Low	**0.69%**	**15.24%**	**3.96%**	**0.60**	**0.64%**	**13.22%**	**4.17%**	**0.53**	**0.70%**	**14.96%**	**3.49%**	**0.70**

In this table, I report the performance of the prediction-sorted, value-weighted decile portfolios over the out-of-sample testing period. All stocks are sorted into deciles based on their predicted returns for the next month. Columns "Pred" provides the predicted monthly returns for each decile, "Avg" provides the average realized monthly returns, "SD" provides the standard deviations and "SR" reports the Sharpe ratios. This follows Gu et al.'s (2020) methodology.

4.3.4. Factor Model Regressions

Table 9 presents the results of the Fama-French Three-Factor model (FF3) regressions in the case of the machine learning portfolios formed in the Chinese equity market. Ridge achieves the highest alpha at 2.0% (t=3.68), followed by Lasso at 1.9% (t=3.17) and Elastic Net at 1.9% (t=3.20) as well. OLS exhibits a slightly lower alpha of 1.8% (t=3.47). Consequently, we can claim that the linear models' portfolios generate meaningful abnormal returns not captured by market, size, and value factors. Moving on to nonlinear models, we can

observe that Random Forest produces a negative alpha of -0.2% which is consistent with its weak decile performance. All Neural Networks (NN1-NN4) maintain low but positive alphas ranging from 0.6% (NN3) to 0.9% (NN1) with limited statistical significance. The R^2 values of the linear models range from approximately 1.05% (OLS) to 1.54% (Elastic Net), indicating only a weak connection between the portfolio returns and the FF3 factors.

As Table 9 also shows the results of adding momentum factor to the FF3 model, we can see that the alphas of regularized linear models remain the same with Lasso and Ridge staying at 1.9% (t=2.49) and 2.0% (t=3.10), respectively. Moreover, OLS's alpha increases to 2.1% (t=3.32), which implies that momentum only explains a small portion of portfolio performance. The alpha of Random Forest remains negative (-0.2%, t= -0.40) which confirms the model's poor factor-adjusted performance. Overall, the addition of momentum does not meaningfully reduce abnormal returns of the best-performing linear models, indicating that they capture signals beyond simple momentum exposure.

Interestingly, the inclusion of profitability and investment factors somewhat increase the alphas of the best performing models with Ridge yielding an abnormal return of 2.2% (t=3.32) and Lasso having an alpha of 2.0% (t=2.78). Since OLS's alpha remains the same (2.1%, t=3.36) we can suggest that the additional factors do not explain away the returns, rather, they highlight that the long-short portfolios capture pricing information beyond what is already captured by the FF5 model.

The R^2 values increase slightly (up to 2.84% for Ridge), but remain low in general, which emphasizes the statement that long-short portfolio returns are independent of traditional risk factors in the Chinese equity market.

Table 9 - Risk-adjusted performance of machine learning portfolios in the case of China

		OLS	Lasso	Ridge	ENet (α=0.5)	RF	NN1	NN2	NN3	NN4
FF3	mean_ret	1.79%	1.89%	1.98%	1.89%	-0.12%	0.87%	0.69%	0.64%	0.70%
	α	0.018	0.019	0.020	0.019	-0.002	0.009	0.007	0.006	0.007
	t (α)	3.47	3.17	3.68	3.20	-0.51	2.51	2.42	1.95	2.57
	R^2	1.05%	1.35%	1.13%	1.54%	2.38%	3.29%	0.71%	1..90%	2.14%
Carhart	mean_ret	1.79%	1.89%	1.98%	1.89%	-0.12%	0.87%	0.69%	0.64%	0.70%
	α	0.021	0.019	0.020	0.019	-0.002	0.009	0.006	0.006	0.008
	t (α)	3.32	2.49	3.10	2.54	-0.40	2.20	1.86	1.72	2.30
	R^2	1.18%	1.36%	1.14%	1.54%	4.53%	3.30%	0.78%	1.90%	2.30%
FF5	mean_ret	1.79%	1.89%	1.98%	1.89%	-0.12%	0.87%	0.69%	0.64%	0.70%
	α	0.021	0.020	0.022	0.021	-0.001	0.009	0.008	0.007	0.008
	t (α)	3.36	2.78	3.32	2.86	-0.23	2.78	2.43	2.12	2.71
	R^2	2.42%	2.74%	2.84%	3.05%	6.54%	3.43%	1.16%	2.56%	2.54%

In this table, I report the average monthly returns, the alphas and the R^2 values with respect to the Fama-French Three-Factor, the Carhart and the Fama-French Five-Factor models.

4.4. Japan

After excluding variables with more than 50% of missing observations and imputing missing values with cross-sectional medians I obtained a final sample with 1,270,566 monthly firm-level observations and 132 characteristics from the JKP variable set.

4.4.1. Model Performance Evaluation

As presented in Table 10, among the linear models, Ridge delivers the strongest out-of-sample performance with an RMSE of 0.1069 and an R^2 value of -3.63%, narrowly outperforming Lasso (-3.67%, RMSE 0.1069) and Elastic Net (-3.67%, RMSE 0.1069). OLS moderately underperforms with an R^2 value of -3.74% and an RMSE of 0.1070.

Nonlinear methods perform significantly worse. For instance, Random Forest exhibits an R^2 of -7.06% (RMSE 0.1084). Considering the deep learning methods, Neural Networks record the weakest outcomes with R^2 values ranging from -7.72% (NN3) to -10.28% (NN4) and RMSEs all above 0.1090. The clear drop in predictive performance points out the tendency of more complex models to overfit in lower-signal environments.

Table 10 - Out-of-Sample R-Squared and RMSE values for the Japanese equity market, data collected from the WRDS database

Machine Learning Method	R^2	RMSE
OLS	-3.74%	0.1070
Lasso	-3.67%	0.1069
Ridge	-3.63%	0.1069
ENet (α=0.5)	-3.67%	0.1069
RF	-7.06%	0.1084
NN1	-7.75%	0.1090
NN2	-10.20%	0.1103
NN3	-7.72%	0.1090
NN4	-10.28%	0.1103

4.4.2 Variable Importance

Figure 4 presents the twenty most influential predictors in Japan. The results point out the importance of structural firm features, financing decisions, profitability measures and investment related signals. The variable *age* ranks as the top predictor, highlighting that maturity could strongly impact return patterns in Japan. Older firms, often deeply embedded in keiretsu networks, which are interlinked corporate groups characterized by cross-shareholdings and long-term business relationships (Lincoln and Gerlach, 2004), may exhibit different risk

and growth profiles compared to younger companies. Change in shares outstanding (*chcsho_12m*) is another prominent predictor, which reflects the significant role of equity issuance and buyback activity in modelling returns. The presence of *sti_gr1a* indicates that shifts in firms' short-term asset holdings are predictive of returns and potentially capture corporate liquidity management and risk-taking behaviour. The growth in capital expenditures is represented by *capx_gr1a* which could signal management confidence and affect valuation expectations, since it reinforces the market's sensitivity to changes in reinvestment and long-term growth decisions. The predictor *cmp_gpa* (gross profitability scaled by assets) aligns with

Novy-Marx's (2013) evidence that profitability is a key driver od cross-sectional returns and implies that more profitable firms yield a performance premium in Japan's market.

To sum up, the variable importance rankings let us observe that investors in Japan place remarkable weight on firm maturity, equity financing decisions, profitability and capital allocation policies.

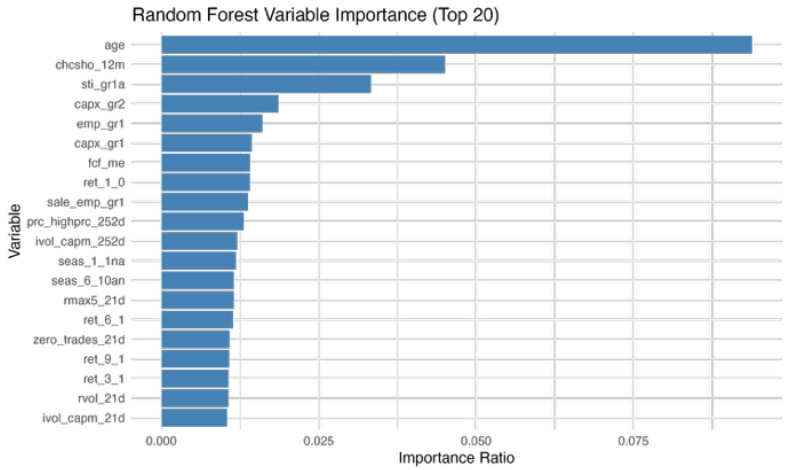

Figure 4 - Random Forest normalized feature importance ratios - the 20 highest values in the case of Japan

4.4.3. Portfolio Analysis

The results of the models in the Japanese equity market are presented in Table 11. We can observe that all models produce relatively modest long-short spreads which reflects the limited cross-sectional predictability in the Japanese market.

OLS yields the highest spread among all linear models at 0.61% with an annualized out-of-sample Sharpe ratio, followed by Elastic Net, which delivers an even weaker spread of 0.38% and a Sharpe ratio of 0.31. The two other regularized linear models, Ridge and Lasso underperform, with long-short spreads of 0.32% (SR=0.25) and 0.22% (SR=0.18), respectively.

The nonlinear models once again underperform, with Random Forest adding only a little economic value and a long-short spread of 0.13% (SR=0.12). Neural Networks produce mixed results, their spreads range from 0.10% (NN1) to 0.42% (NN4), but their higher volatility allows a maximum Sharpe ratio of 0.42 (NN4).

To summarize, the decile progression for most models is relatively flat, with High-decile portfolios earning around 0.53-0.83% per month. The dispersion between the High and Low deciles is modest, with Low-decile portfolios still delivering small positive returns across most models. This limited separation explains why the High–Low spreads remain low, underscoring the challenge of extracting strong predictive signals in Japan's market.

Table 11 - Performance of the machine learning portfolios in the case of Japan

	OLS				**Lasso**				**Ridge**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	0.19%	-1.44%	4.71%	0.14	0.37%	-0.59%	4.78%	0.27	0.34%	-0.46%	5.20%	0.23
2	0.39%	-0.67%	4.41%	0.31	0.45%	-0.17%	4.32%	0.36	0.38%	-0.09%	4.76%	0.28
3	0.32%	-0.24%	4.44%	0.25	0.44%	0.07%	4.38%	0.35	0.44%	0.12%	4.53%	0.34
4	0.46%	0.09%	4.45%	0.36	0.35%	0.25%	4.76%	0.26	0.44%	0.28%	4.62%	0.33
5	0.49%	0.37%	4.64%	0.37	0.30%	0.41%	4.63%	0.23	0.44%	0.42%	4.59%	0.33
6	0.51%	0.64%	4.62%	0.38	0.39%	0.56%	4.77%	0.28	0.56%	0.55%	4.66%	0.41
7	0.49%	0.91%	4.65%	0.36	0.28%	0.72%	4.89%	0.20	0.43%	0.68%	4.75%	0.31
8	0.74%	1.22%	4.71%	0.54	0.39%	0.88%	5.06%	0.27	0.50%	0.83%	4.78%	0.36
9	0.48%	1.59%	5.19%	0.32	0.44%	1.09%	5.07%	0.30	0.46%	1.01%	4.72%	0.34
High	0.80%	2.28%	5.37%	0.52	0.59%	1.47%	5.25%	0.39	0.66%	1.35%	4.67%	0.49
High-Low	**0.61%**	**3.72%**	**4.43%**	**0.48**	**0.22%**	**2.06%**	**4.20%**	**0.18**	**0.32%**	**1.81%**	**4.50%**	**0.25**
	ENet (α=0.5)				**RF**				**NN1**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	0.34%	-0.60%	4.80%	0.24	0.25%	-0.77%	5.40%	0.16	0.43%	-3.56%	4.99%	0.30
2	0.35%	-0.17%	4.32%	0.28	0.41%	0.02%	4.69%	0.30	0.22%	-1.94%	4.84%	0.16
3	0.49%	0.07%	4.38%	0.39	0.38%	0.35%	4.53%	0.29	0.45%	-1.09%	4.56%	0.34
4	0.46%	0.25%	4.67%	0.34	0.56%	0.59%	4.22%	0.46	0.43%	-0.45%	4.56%	0.33
5	0.29%	0.41%	4.70%	0.22	0.52%	0.80%	4.20%	0.43	0.41%	0.11%	4.54%	0.31
6	0.26%	0.56%	4.81%	0.19	0.48%	1.01%	4.36%	0.38	0.37%	0.65%	4.45%	0.28
7	0.45%	0.72%	4.83%	0.32	0.59%	1.23%	4.35%	0.47	0.58%	1.21%	4.56%	0.44
8	0.45%	0.88%	5.09%	0.30	0.34%	1.50%	4.59%	0.25	0.49%	1.83%	4.46%	0.38
9	0.54%	1.09%	5.01%	0.37	0.48%	1.91%	4.73%	0.35	0.61%	2.64%	4.54%	0.46
High	0.72%	1.47%	5.20%	0.48	0.38%	2.91%	5.21%	0.25	0.53%	4.22%	4.79%	0.38
High-Low	**0.38%**	**2.07%**	**4.27%**	**0.31**	**0.13%**	**3.67%**	**3.62%**	**0.12**	**0.10%**	**7.78%**	**3.47%**	**0.10**
	NN2				**NN3**				**NN4**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	0.39%	-4.27%	5.07%	0.27	0.24%	-3.51%	5.19%	0.16	0.25%	-4.11%	5.41%	0.16
2	0.23%	-2.11%	4.54%	0.17	0.35%	-1.54%	4.71%	0.25	0.34%	-1.95%	4.73%	0.25
3	0.53%	-1.18%	4.57%	0.40	0.31%	-0.74%	4.33%	0.25	0.29%	-1.01%	4.36%	0.23
4	0.40%	-0.50%	4.31%	0.32	0.47%	-0.18%	4.32%	0.38	0.29%	-0.38%	4.42%	0.23
5	0.35%	0.07%	4.34%	0.28	0.58%	0.28%	4.59%	0.44	0.50%	0.14%	4.38%	0.39
6	0.48%	0.62%	4.41%	0.37	0.46%	0.71%	4.35%	0.37	0.44%	0.62%	4.18%	0.37
7	0.56%	1.19%	4.28%	0.45	0.44%	1.17%	4.37%	0.35	0.49%	1.11%	4.31%	0.39
8	0.48%	1.85%	4.54%	0.36	0.45%	1.71%	4.68%	0.34	0.59%	1.70%	4.36%	0.47
9	0.40%	2.75%	4.56%	0.31	0.57%	2.49%	4.55%	0.43	0.51%	2.61%	4.71%	0.38
High	0.58%	4.89%	5.11%	0.40	0.50%	4.33%	5.32%	0.33	0.67%	5.23%	5.48%	0.42
High-Low	**0.19%**	**9.17%**	**3.24%**	**0.21**	**0.26%**	**7.84%**	**3.96%**	**0.23**	**0.42%**	**9.35%**	**3.45%**	**0.42**

In this table, I report the performance of the prediction-sorted, value-weighted decile portfolios over the out-of-sample testing period. All stocks are sorted into deciles based on their predicted returns for the next month. Columns "Pred" provides the predicted monthly returns for each decile, "Avg" provides the average realized monthly returns, "SD" provides the standard deviations and "SR" reports the Sharpe ratios. This follows Gu et al.'s (2020) methodology.

4.4.4. Factor Model Regressions

Under the FF3 specification (presented in Table 12), alphas are small across all models, ranging from 0.1% to 0.7% monthly, with OLS delivering the highest alpha at 0.7% (t = 2.41). Ridge and Lasso show alphas of only 0.3% (t ≈ 1.1), below conventional significance levels. The R^2 values are low (0.25%–2.87%), indicating that standard market, size, and value factors explain little of the variance in these long–short strategies.

Adding the momentum factor produces only minor changes, as presented in Table 12. OLS maintains an alpha of 0.007 (t = 2.43), while regularized models remain statistically insignificant (Lasso and Ridge both around 0.3%, t < 1.2). The slightly higher R^2 values (up to 2.87% for Ridge) suggest some incremental explanatory power, but momentum does not fully account for the returns.

Incorporating profitability and investment factor slightly increases explanatory power (R^2 up to 3.2%), but alphas remain low. OLS keeps a small but statistically meaningful alpha (0.007, t = 2.53), while all other models show non-significant alphas between 0.1% and 0.4%.

In conclusion, alphas are modest and mostly lack statistical significance, especially for the regularized models. Only OLS retains a consistently small but meaningful alpha across all specifications. Low R^2 values (all below 3.3%) indicate that traditional factor models explain little of the cross-sectional variation in these ML-based long–short portfolios, even when expanded to include momentum, profitability, and investment factors.

Overall, the machine learning strategies in Japan generate weak excess returns beyond standard risk premia, consistent with the earlier finding that Japan represents a low-signal, harder-to-predict market.

Table 12 - Risk-adjusted performance of machine learning portfolios in the case of Japan

		OLS	Lasso	Ridge	ENet (α=0.5)	RF	NN1	NN2	NN3	NN4
FF3	mean_ret	0.61%	0.22%	0.32%	0.38%	0.13%	0.10%	0.19%	0.26%	0.42%
	α	0.007	0.003	0.003	0.004	0.001	0.001	0.001	0.001	0.005
	t (α)	2.41	0.94	1.12	1.55	0.49	0.31	0.47	0.61	2.15
	R^2	0.79%	0.58%	0.48%	0.25%	2.38%	2.03%	0.31%	0.48%	0.62%
Carhart	mean_ret	0.61%	0.22%	0.32%	0.38%	0.13%	0.10%	0.19%	0.26%	0.42%
	α	0.007	0.003	0.003	0.004	0.001	0.001	0.002	0.002	0.005
	t (α)	2.43	0.97	1.12	1.56	0.48	0.58	0.80	0.77	1.97
	R^2	0.97%	1.58%	-0.52%	0.36%	2.38%	2.87%	1.61%	0.72%	0.63%
FF5	mean_ret	0.61%	0.22%	0.32%	0.38%	0.13%	0.10%	0.19%	0.26%	0.42%
	α	0.007	0.003	0.003	0.004	0.001	0.001	0.003	0.003	0.006
	t (α)	2.53	0.99	1.22	1.61	0.28	0.63	1.42	1.05	2.48
	R^2	1.63%	1.77%	0.74%	1.23%	3.20%	2.92%	4.27%	1.89%	2.04%

In this table, I report the average monthly returns, the alphas and the R^2 values with respect to the Fama-French Three-Factor, the Carhart and the Fama-French Five-Factor models.

4.5. United Kingdom

After excluding variables with more than 50% of missing observations and imputing missing values with cross-sectional medians I obtained a final sample with 640,636 monthly firm-level observations and 113 characteristics from the JKP variable set.

4.5.1. Model Performance Evaluation

The out-of-sample model performances show (Table 13) that predictive accuracy in the UK is modest across all specifications. Linear models (OLS, Lasso, Ridge, and Elastic Net) perform almost identically, with R^2 values of approximately –2.21% and RMSEs of approximately 0.1565, suggesting that regularization does little to improve predictive power over a simple OLS benchmark. Random Forest performs worse with an R^2 of –3.49%, while Neural Networks underperform further, particularly NN2 (R^2 = –9.27%), with other architectures also showing elevated errors. These results suggest that complex, non-linear models fail to extract additional predictive structure from the data, likely due to a relatively low-signal environment and stronger pricing efficiency in the UK market compared to more fragmented or less liquid markets like China.

Table 13 - Out-of-Sample R-Squared and RMSE values for the UK equity market, data collected from WRDS database

Machine Learning Method	R^2	RMSE
OLS	-2.22%	0.1565
Lasso	-2.21%	0.1565
Ridge	-2.21%	0.1565
ENet (α=0.5)	-2.21%	0.1565
RF	-3.49%	0.1574
NN1	-5.52%	0.1589
NN2	-9.27%	0.1616
NN3	-6.91%	0.1598
NN4	-6.91%	0.1599

4.5.2. Variable Importance

Figure 5 visualises the twenty most influential predictors in the United Kingdom. As we can observe, the predictor *age* (firm age) ranks as the most important firm characteristic, which highlights the role of company maturity in shaping expected returns. Moving on, the second most impactful predictor is *sti_gr1a*, which represents sales growth and reflects the market's sensitivity to changes in growth fundamentals. Other than the previously mentioned predictors, momentum indicators also play a role in predicting future returns, since 1- ,3- ,6- , 9-, and 12-month past returns are all present on Figure 5. Seasonality (*seas_1_1na*) appears

among the top predictors as well, which captures the calendar-based return effects that remain relevant even in an efficient market like the UK. In addition to the priorly mentioned predictors, valuation measures such as earnings-to-book (*niq_be*) and dividends-to-market equity (*div12m_me*) underscore the importance of value-oriented signals. As exhibited, liquidity and

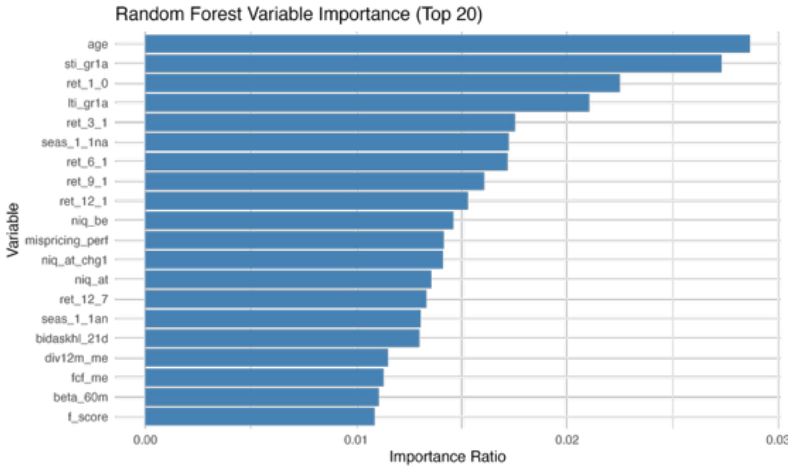

Figure 5 - Random Forest normalized feature importance ratios - the 20 highest values in the case of the UK

trading fractions are also represented, namely by the predictor *bidaskhl_21d*, which indicates that market microstructure effects are relevant for pricing, especially in less liquid stocks. Profitability and cash flow proxies such as free cash flow to market equity (*fcf_me*) and net income scaled by assets (*niq_at*) appear to be mediocre important, which signals that fundamental efficiency metrics play a role in return prediction.

In conclusion, the broad mix of predictor indicates that momentum, growth, value and liquidity remain central drivers of UK cross-sectional returns, as well as profitability and mispricing-based measures. Thus, we can claim that the market prices a wide range of firm characteristics beyond simple risk factors.

4.5.3. Portfolio Analysis

The value-weighted portfolio results for the UK (Table 14) reveal notable spreads in realized returns across models, with the linear models performing especially well. Ridge regression emerges as the strongest performer, since its long-short spread yields 2.06%, with an out-of-sample annualized Sharpe ratio of 1.08. OLS exhibits the second-best performance with a spread of 1.65% (SR=0.95). Lasso reaches a spread of 1.60% and a Sharpe ratio of 0.84. Since Elastic Net provides the fourth biggest spread (1.52%, SR=0.81) we can underline that linear models with regularization continue to extract meaningful cross-sectional signals from the UK equity market.

By contrast, nonlinear models fail to translate they often large predicted spreads into realized gains. Neural Networks let us observe aggressively big predictions, such as a forecast of an 11.31% spread yet their realized long-short returns are minimal, ranging from 0.13% to 0.64% across architectures. Thus, while their forecasts are often extreme, they do not consistently align with actual returns. Random Forest provides s spread of 1.01%, though its Sharpe ratio only reaches 0.59 which is an underperformance compared to linear methods.

To sum up, these findings further highlight the pattern observed in other markets: linear models tend to deliver more reliable and interpretable portfolio outcomes, whereas the more complex methods appear prone to overfitting and struggle to generalize effectively in the UK context.

Table 14 - Performance of the machine learning portfolios in the case of the UK

	OLS				**Lasso**				**Ridge**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	-1.24%	-2.54%	8.06%	-0.53	-1.19%	-1.80%	8.06%	-0.51	-1.57%	-1.81%	8.29%	-0.66
2	-0.46%	-1.52%	7.57%	-0.21	-1.00%	-1.11%	7.91%	-0.44	-0.95%	-1.10%	8.48%	-0.39
3	-0.01%	-0.91%	6.27%	-0.01	-0.21%	-0.67%	6.61%	-0.11	-0.59%	-0.66%	7.02%	-0.29
4	-0.02%	-0.44%	6.04%	-0.01	0.06%	-0.31%	6.23%	0.03	0.30%	-0.31%	6.58%	0.16
5	0.27%	-0.03%	6.04%	0.15	-0.01%	-0.01%	6.83%	0.00	0.09%	0.00%	6.13%	0.05
6	0.31%	0.33%	5.70%	0.19	0.25%	0.27%	6.11%	0.14	0.23%	0.26%	6.02%	0.13
7	0.50%	0.70%	5.40%	0.32	0.20%	0.54%	5.49%	0.13	0.38%	0.53%	5.56%	0.24
8	0.31%	1.09%	5.33%	0.20	0.31%	0.82%	5.48%	0.20	0.42%	0.81%	5.36%	0.27
9	0.41%	1.56%	5.62%	0.25	0.36%	1.15%	5.67%	0.22	0.50%	1.14%	5.69%	0.30
High	0.42%	2.43%	5.75%	0.25	0.41%	1.71%	6.04%	0.24	0.49%	1.71%	5.63%	0.30
High-Low	**1.65%**	**4.97%**	**6.04%**	**0.95**	**1.60%**	**3.52%**	**6.57%**	**0.84**	**2.06%**	**3.52%**	**6.63%**	**1.08**
	ENet (α=0.5)				**RF**				**NN1**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	-1.15%	-1.79%	8.01%	-0.50	-0.56%	-1.74%	8.36%	-0.23	0.15%	-5.13%	6.22%	0.08
2	-0.99%	-1.10%	7.94%	-0.43	-0.22%	-0.65%	7.01%	-0.11	0.19%	-3.08%	5.86%	0.11
3	-0.15%	-0.67%	6.55%	-0.08	-0.09%	-0.04%	6.56%	-0.05	0.44%	-1.88%	6.04%	0.25
4	0.13%	-0.31%	6.21%	0.07	0.46%	0.40%	5.75%	0.28	0.40%	-1.00%	5.41%	0.26
5	-0.05%	-0.01%	6.59%	-0.02	0.35%	0.75%	6.11%	0.20	0.53%	-0.21%	5.66%	0.32
6	0.21%	0.27%	6.29%	0.12	0.29%	1.05%	5.70%	0.18	0.33%	0.55%	5.60%	0.20
7	0.17%	0.53%	5.50%	0.11	0.48%	1.33%	5.70%	0.29	0.54%	1.33%	5.72%	0.33
8	0.37%	0.81%	5.42%	0.24	0.49%	1.65%	5.15%	0.33	0.02%	2.23%	5.54%	0.01
9	0.37%	1.14%	5.70%	0.22	0.53%	2.12%	5.65%	0.33	0.18%	3.42%	5.34%	0.11
High	0.38%	1.70%	6.04%	0.22	0.45%	3.22%	6.61%	0.24	0.22%	5.65%	6.26%	0.12
High-Low	**1.52%**	**3.49%**	**6.52%**	**0.81**	**1.01%**	**4.95%**	**6.00%**	**0.59**	**0.07%**	**10.78%**	**4.50%**	**0.05**
	NN2				**NN3**				**NN4**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	0.08%	-6.99%	7.64%	0.04	-0.16%	-5.53%	8.34%	-0.07	-0.37%	-6.14%	7.57%	-0.17
2	0.60%	-3.85%	6.02%	0.35	-0.06%	-2.76%	6.40%	-0.03	0.20%	-3.18%	6.01%	0.12
3	0.11%	-2.32%	5.45%	0.07	0.09%	-1.50%	6.08%	0.05	0.48%	-1.84%	5.94%	0.28
4	0.34%	-1.20%	5.39%	0.22	0.31%	-0.63%	5.51%	0.20	0.42%	-0.93%	5.55%	0.26
5	0.20%	-0.25%	5.58%	0.13	0.24%	0.07%	5.30%	0.16	0.41%	-0.20%	5.18%	0.27
6	0.21%	0.65%	5.45%	0.13	0.43%	0.72%	5.56%	0.27	0.47%	0.44%	5.27%	0.31
7	0.34%	1.58%	5.36%	0.22	0.44%	1.38%	5.36%	0.28	0.36%	1.10%	5.55%	0.22
8	0.39%	2.69%	5.34%	0.25	0.47%	2.20%	5.36%	0.30	0.16%	1.87%	5.39%	0.10
9	0.32%	4.20%	5.58%	0.20	0.17%	3.33%	5.78%	0.10	0.21%	2.98%	5.96%	0.12
High	-0.04%	7.19%	6.79%	-0.02	0.48%	5.78%	6.45%	0.26	0.15%	5.36%	7.74%	0.07
High-Low	**-0.13%**	**14.18%**	**5.23%**	**-0.08**	**0.64%**	**11.31%**	**5.65%**	**0.39**	**0.52%**	**11.50%**	**5.70%**	**0.32**

In this table, I report the performance of the prediction-sorted, value-weighted decile portfolios over the out-of-sample testing period. All stocks are sorted into deciles based on their predicted returns for the next month. Columns "Pred" provides the predicted monthly returns for each decile, "Avg" provides the average realized monthly returns, "SD" provides the standard deviations and "SR" reports the Sharpe ratios. This follows Gu et al.'s (2020) methodology.

4.5.4. Factor Regression Analysis

Under the FF3 model (Table 15), the long-short alphas for Ridge, (2.1%, t= 4.83), OLS (1.8%, t=4.45) and Lasso (1.7%, t=3.86) are positive and statistically significant, confirming that these strategies generate excess returns beyond standard market, size and value exposures. The exhibited R^2 values rise to 7.37% (Ridge), implying that a portion of the long-short portfolio performance is explained by the traditional risk factors, but a large share remains unexplained, underscoring that the models capture unique pricing signals.

Introducing the momentum in the Carhart model improves explanatory power: Ridge's R^2 value reaches 11.04%, with the other linear models also seeing gains. Alphas of the strategies decrease slightly (e.g., Ridge drops to 1.8% from 2.1%, t=3.98), reflecting that part of the strategies' excess returns depends on momentum exposures.

Finally, in the FF5 model Ridge's and Lasso's alphas stay 1.8% (t=3.74) and 1.4% (t=3.11), respectively, while OLS shows a slight decrease. The R^2 improves slightly (up to 11.64% for Ridge), indicating that these additional factors capture some variation. In returns, though their incremental explanatory power is limited.

Table 15 - Risk-adjusted performance of machine learning portfolios in the case of the UK

		OLS	Lasso	Ridge	ENet (α=0.5)	RF	NN1	NN2	NN3	NN4
FF3	mean_ret	1.65%	1.60%	2.06%	1.52%	1.01%	0.07%	-0.13%	0.64%	0.52%
	α	0.018	0.017	0.021	0.016	0.010	0.002	-0.001	0.006	0.005
	t (α)	4.45	3.86	4.83	3.72	2.77	0.67	-0.31	1.96	1.40
	R^2	6.69%	6.04%	7.37%	6.03%	0.26%	3.56%	0.55%	2.24%	0.01%
Carhart	mean_ret	1.65%	1.60%	2.06%	1.52%	1.01%	0.07%	-0.13%	0.64%	0.52%
	α	0.016	0.014	0.018	0.014	0.009	0.003	0.000	0.005	0.005
	t (α)	3.88	3.11	3.98	2.96	2.43	1.06	0.042	1.27	1.21
	R^2	7.27%	8.12%	11.04%	8.28%	0.51%	4.27%	1.00%	2.80%	0.01%
FF5	mean_ret	1.65%	1.60%	2.06%	1.52%	1.01%	0.07%	-0.13%	0.64%	0.52%
	α	0.015	0.014	0.018	0.013	0.010	0.001	-0.003	0.008	0.004
	t (α)	3.22	2.91	3.74	2.80	2.66	0.28	-0.81	2.29	1.04
	R^2	8.68%	10.18%	11.64%	10.21%	0.27%	4.02%	1.57%	4.97%	0.32%

In this table, I report the average monthly returns, the alphas and the R^2 values with respect to the Fama-French Three-Factor, the Carhart and the Fama-French Five-Factor models.

4.6. France

After excluding variables with more than 50% of missing observations and imputing missing values with cross-sectional medians I obtained a final sample with 246,381 monthly firm-level observations and 129 characteristics from the JKP variable set.

4.6.1 Model Performance Evaluation

The out-of-sample performance of the models (presented in Table 16) for France show that regularized linear approaches, such as Lasso, Elastic Net and Ridge achieved the most stable results, with RMSEs around 0.1366 and negative but relatively small R^2 values of -

1.89%, -1.89% and -1.91%, respectively. OLS performed slightly worse with reaching an R^2 of -2.11%.

Random Forest's R^2 value is even worse than OLS's, with achieving only -3.02% and an RMSE of 0.1373, underlining that tree-based methods struggle to capture meaningful structure in the French equity market. This may be tied to their tendency to fit local patterns that do not generalize well across time, especially when cross-sectional return variation is modest.

Table 16 - Out-of-Sample R-Squared and RMSE values for the French equity market, data collected from WRDS database

Machine Learning Method	R^2	RMSE
OLS	-2.11%	0.1367
Lasso	-1.89%	0.1366
Ridge	-1.91%	0.1366
ENet (α=0.5)	-1.89%	0.1366
RF	-3.02%	0.1373
NN1	-14.08%	0.1439
NN2	-20.10%	0.1474
NN3	-10.99%	0.1422
NN4	-11.21%	0.1423

The biggest underperformance came from Neural Networks (NN1-NN4), which showed negative R^2 values between -10.99% and -20.10% and higher RMSEs (up to 0.1474). This suggests significant overfitting and poor generalization capabilities when applied out-of-sample.

4.6.2. Variable Importance Analysis

The most influential predictors are shown on Figure 6. *Ret_1_0* (1-month momentum) stands out as the strongest predictor, highlighting the relevance of short-term momentum effects in the French equity market. The predictor *age* suggests that the company's maturity plays an important role in shaping returns. Medium-horizon momentum factors (e.g., *ret_3_1*, *ret_6_1*, *ret_9_1*) remain significant, showing the persistence of momentum signals across multiple time windows. The presence of *seas_1_1an* (seasonality) and *ret_12_7* underscore the contribution of seasonal and long-term momentum components and therefore further emphasize the impact of calendar effects. Variables such as *mispricing_perf* and *emp_gr1* (employee growth) point out the role of fundamental mispricing proxies and operational metrics, indicating that firm-specific efficiency and labour dynamics matter in French equities. Lastly, liquidity and trading cost characteristics (e.g., *bidaskhl_21d* and *highprc252d*) also

appear in the top predictors, emphasizing the influence of trading behaviour on asset pricing in this market.

To sum up, the diversity of top predictors, such as momentum, seasonality, fundamental metrics and liquidity, suggests that excess returns in the French market are impacted by a mix of behavioural (momentum and seasonality) and structural (firm fundamentals and trading costs) factors.

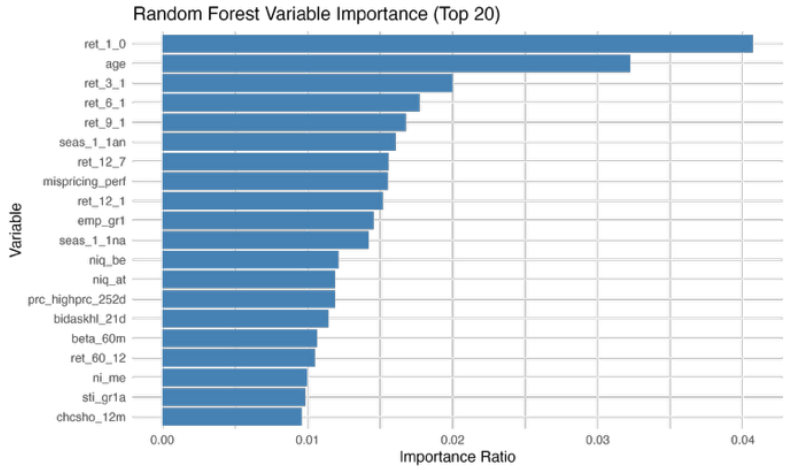

Figure 6 - Random Forest normalized feature importance ratios - the 20 highest values in the case of France

4.6.3. Portfolio Analysis

Table 17 shows the performance of machine learning portfolios constructed on the French equity market. We can observe that the best performing portfolio was OLS, achieving a long-short spread of 1.57% with an OOS annualized Sharpe ratio of 0.95. This is followed by Lasso at 1.06% (SR=0.55), Ridge at 1.01% (SR=0.50) and Elastic Net at 0.98% (SR=0.52). Tree-based and Neural Network models show more varying results: Random Forests generated a moderate High-Low spread of 0.28% with as low of a Sharpe ratio as 0.13. Neural Network models (NN1-NN4) produced extreme values for predicted spreads (e.g., 20.03% for NN2 and 16.74% for NN1) but their realized returns were much smaller, ranging between 0.15% (NN2) and 0.34% (NN3). The overestimation underlines the tendency of deep learning models to overfit the in-sample data without translating those forecasts into meaningful out-of-sample returns.

Overall, these results indicate that linear models remain the most reliable for portfolio construction in the French equity market, providing a superior performance compared to tree-based and neural network approaches on realized spreads and risk-adjusted performance. The large difference between predicted and realized returns in neural networks further supports the

overfitting appearing in the mentioned models. This is consistent with the model evaluation results, where the NN1-NN4 models exhibited the weakest R^2 values and highest RMSEs.

Table 17 - Performance of the machine learning portfolios in the case of France

	OLS				Lasso				Ridge			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	-0.59%	-2.47%	7.93%	-0.26	-0.12%	-1.40%	8.09%	-0.05	-0.14%	-1.33%	8.31%	-0.06
2	0.52%	-1.27%	6.98%	0.26	0.58%	-0.62%	7.37%	0.27	0.21%	-0.58%	8.15%	0.09
3	0.39%	-0.57%	6.24%	0.21	0.36%	-0.17%	6.65%	0.19	0.24%	-0.15%	7.42%	0.11
4	0.73%	-0.04%	6.25%	0.41	0.80%	0.18%	6.54%	0.42	0.61%	0.19%	6.64%	0.32
5	0.59%	0.42%	6.29%	0.33	0.21%	0.48%	6.40%	0.11	0.51%	0.48%	6.59%	0.27
6	0.68%	0.85%	5.84%	0.40	0.62%	0.75%	6.08%	0.35	0.62%	0.74%	6.08%	0.35
7	0.74%	1.28%	5.78%	0.44	0.59%	1.03%	6.21%	0.33	0.85%	1.00%	6.17%	0.48
8	0.97%	1.76%	6.62%	0.51	0.89%	1.33%	6.23%	0.49	0.82%	1.29%	6.05%	0.47
9	1.01%	2.34%	6.41%	0.54	1.10%	1.69%	6.23%	0.61	0.99%	1.64%	6.12%	0.56
High	0.98%	3.39%	7.23%	0.47	0.94%	2.29%	6.71%	0.49	0.87%	2.26%	6.75%	0.45
High-Low	**1.57%**	**5.87%**	**5.74%**	**0.95**	**1.06%**	**3.69%**	**6.68%**	**0.55**	**1.01%**	**3.59%**	**6.99%**	**0.50**
	ENet (α=0.5)				**RF**				**NN1**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	-0.03%	-1.41%	8.05%	-0.01	0.23%	-2.06%	9.48%	0.08	0.43%	-7.67%	6.74%	0.22
2	0.50%	-0.62%	7.39%	0.24	0.58%	-0.91%	7.51%	0.27	0.60%	-4.30%	6.62%	0.31
3	0.37%	-0.17%	6.64%	0.19	0.21%	-0.34%	6.56%	0.11	0.80%	-2.51%	6.63%	0.42
4	0.81%	0.17%	6.49%	0.43	0.71%	0.05%	6.25%	0.40	0.53%	-1.08%	6.07%	0.30
5	0.22%	0.48%	6.20%	0.13	0.65%	0.38%	6.08%	0.37	0.84%	0.15%	6.27%	0.47
6	0.59%	0.75%	6.17%	0.33	0.81%	0.70%	5.87%	0.48	0.87%	1.35%	6.32%	0.48
7	0.66%	1.03%	6.17%	0.37	0.95%	1.03%	5.91%	0.56	0.70%	2.56%	6.51%	0.37
8	0.89%	1.33%	6.22%	0.50	0.86%	1.44%	6.66%	0.45	0.90%	3.94%	6.16%	0.51
9	1.05%	1.70%	6.35%	0.57	0.53%	2.09%	6.60%	0.28	0.68%	5.68%	6.35%	0.37
High	0.95%	2.30%	6.71%	0.49	0.51%	3.51%	7.21%	0.25	0.61%	9.07%	6.61%	0.32
High-Low	**0.98%**	**3.71%**	**6.57%**	**0.52**	**0.28%**	**5.58%**	**7.58%**	**0.13**	**0.18%**	**16.74%**	**4.63%**	**0.13**
	NN2				**NN3**				**NN4**			
	Avg	Pred	SD	SR	Avg	Pred	SD	SR	Avg	Pred	SD	SR
Low	0.59%	-9.40%	6.62%	0.31	0.50%	-6.46%	7.26%	0.24	0.60%	-6.69%	7.90%	0.26
2	0.79%	-5.15%	6.55%	0.42	0.67%	-3.20%	6.73%	0.34	0.15%	-3.42%	6.95%	0.07
3	0.74%	-3.04%	6.11%	0.42	0.90%	-1.63%	6.09%	0.51	0.71%	-1.85%	6.35%	0.39
4	0.66%	-1.40%	6.05%	0.38	0.64%	-0.57%	6.04%	0.37	0.73%	-0.72%	6.16%	0.41
5	0.67%	-0.03%	5.94%	0.39	0.93%	0.32%	6.25%	0.52	0.99%	0.23%	6.79%	0.51
6	0.91%	1.31%	6.21%	0.51	0.56%	1.18%	6.17%	0.32	0.64%	1.14%	5.90%	0.37
7	0.80%	2.68%	6.50%	0.43	0.49%	2.08%	6.17%	0.27	0.62%	2.10%	5.98%	0.36
8	0.40%	4.27%	6.23%	0.22	0.92%	3.18%	6.19%	0.52	1.02%	3.24%	6.12%	0.58
9	0.96%	6.35%	7.05%	0.47	0.83%	4.67%	7.07%	0.41	0.67%	4.87%	6.38%	0.37
High	0.74%	10.63%	7.36%	0.35	0.84%	8.02%	7.58%	0.38	0.79%	8.14%	7.56%	0.36
High-Low	**0.15%**	**20.03%**	**4.76%**	**0.11**	**0.34%**	**14.47%**	**5.73%**	**0.21**	**0.19%**	**14.83%**	**5.69%**	**0.12**

In this table, I report the performance of the prediction-sorted, value-weighted decile portfolios over the out-of-sample testing period. All stocks are sorted into deciles based on their predicted returns for the next month. Columns "Pred" provides the predicted monthly returns for each decile, "Avg" provides the average realized monthly returns, "SD" provides the standard deviations and "SR" reports the Sharpe ratios. This follows Gu et al.'s (2020) methodology.

4.6.4. Factor Regression Analysis

The results of the FF3 regressions are presented in Table 18. We can observe that the linear models generate positive alphas, with OLS achieving an alpha of 1.6% (t=4.95) and Lasso following at 1.1% (t=2.81). The alphas are statistically significant and their R^2 values remain relatively low, 0.04% and 0.75%, respectively. Ridge and Elastic Net both provide

alphas of 1.0%, with t-values of 2.57 and 2.74, respectively. Random Forest and the deep learning models (NN1-NN4) generate near-zero alphas which are not statistically significant.

In the Carhart model (Table 18), the alpha of OLS drops to 1.5% (t=4.34), while Ridge and Lasso remain stable. This implies that part of the previously unexplained return is related to momentum exposure. However, the R^2 values remain low, suggesting that the model doesn't explain the variation in excess returns successfully.

Ridge and Lasso both experience slight increases in their alphas under the FF5 model (see Table 18), with Ridge achieving an alpha of 1.1% (t=2.59) and Lasso an alpha of 1.3% (t=2.93). Since R^2 values are still modest, we can claim that a rather big part of the long-short portfolio performance remains unexplained.

Table 18 - Risk-adjusted performance of machine learning portfolios in the case of France

		OLS	Lasso	Ridge	ENet (α=0.5)	RF	NN1	NN2	NN3	NN4
FF3	mean_ret	1.57%	1.06%	1.01%	0.98%	0.28%	0.18%	0.15%	0.34%	0.19%
	α	0.016	0.011	0.010	0.010	0.003	0.001	0.000	0.004	0.002
	t (α)	4.95	2.81	2.57	2.74	0.71	0.40	0.14	0.99	0.50
	R^2	0.04%	0.75%	0.26%	1.00%	1.02%	2.17%	2.38%	0.11%	0.53%
Carhart	mean_ret	1.57%	1.06%	1.01%	0.98%	0.28%	0.18%	0.15%	0.34%	0.19%
	α	0.015	0.010	0.010	0.010	0.004	0.001	0.001	0.005	0.002
	t (α)	4.34	2.60	2.39	2.54	0.90	0.21	0.39	1.44	0.54
	R^2	0.26%	0.88%	0.28%	1.10%	1.08%	2.29%	2.64%	1.04	0.53%
FF5	mean_ret	1.57%	1.06%	1.01%	0.98%	0.28%	0.18%	0.15%	0.34%	0.19%
	α	0.015	0.013	0.011	0.012	0.004	0.001	-0.001	0.004	0.004
	t (α)	4.12	2.93	2.59	2.76	0.82	0.32	-0.27	0.87	1.86
	R^2	0.09%	2.16%	0.35%	2.02%	1.33%	2.24%	3.31%	0.11%	2.27%

In this table, I report the average monthly returns, the alphas and the R^2 values with respect to the Fama-French Three-Factor, the Carhart and the Fama-French Five-Factor models.

5.Conclusions

This thesis examined the predictive and economic performance of machine learning (ML) methods in constructing cross-sectional stock selection strategies across six major equity markets: the United States, Germany, the United Kingdom, Japan, China, and France. Building on the framework of Gu, Kelly, and Xiu (2020), the study applied a rolling-window design to 153 firm-level predictors from the Jensen, Kelly, and Pedersen (2023) global factor library, employing models ranging from Ordinary Least Squares (OLS) and regularized linear regressions (Lasso, Ridge, Elastic Net) to nonlinear methods (Random Forests and Feedforward Neural Networks with one to four hidden layers). The aim was to assess whether these models, when restricted to firm characteristics alone, could generate economically meaningful, factor-robust excess returns in a variety of market structures.

From a predictive standpoint, regularized linear models consistently outperformed both OLS and nonlinear alternatives in terms of out-of-sample R^2 and RMSE. Ridge regression was the most robust performer overall, delivering the lowest RMSE values and the least negative R^2 statistics in the United States, the United Kingdom, China, and France. Lasso proved more effective in Germany, while no model achieved strong statistical accuracy in Japan, reflecting the difficulty of predicting returns in a market characterised by high institutional ownership, unique governance practices, and potentially greater pricing efficiency. Nonlinear approaches—particularly deep neural networks—underperformed in all markets, suggesting that their flexibility leads to overfitting in the low-signal, high-noise environment of monthly return prediction.

The portfolio analysis revealed that these differences in statistical accuracy translated into substantial variation in economic outcomes. In the United States, Ridge regression achieved the highest average monthly long-short spread of 1.55% with an annualised Sharpe ratio of 0.84, producing results broadly comparable to those in Gu et al. (2020) despite the narrower predictor set. In Germany, Lasso generated a spread of 1.98% with a Sharpe ratio of 0.82, outperforming Ridge and reversing the model ranking found by Gu et al. in the U.S., a divergence likely driven by Germany's lower liquidity and more concentrated ownership structures, where sparsity in variable selection may be advantageous. The United Kingdom also favoured Ridge, with spreads above 2.0% and Sharpe ratios exceeding 0.80, a result that aligns with Gu et al.'s conclusion that Ridge performs well in deep and liquid markets. China presented a more challenging environment. Ridge again led, with a long-short spread of around 1.98%, with Sharpe ratios all higher than 0.60. France displayed patterns similar to the U.S., with OLS producing the highest spread of 1.57% and a Sharpe ratio of 0.95. Japan remained an outlier, with all models generating spreads below 1.0% and Sharpe ratios under 0.50, corroborating Gu et al.'s observation that some market environments offer little exploitable cross-sectional predictability.

Comparisons with Gu et al. (2020) highlight both parallels and differences. The relative ordering of models was broadly consistent, with Ridge and Lasso outperforming other approaches, and nonlinear methods failing to produce competitive results. However, the spreads achieved in this thesis were generally smaller than those in Gu et al.'s U.S. setting, which can be attributed to the absence of macroeconomic variables, sectoral interactions, and certain microstructure predictors in the present study. More importantly, while Gu et al.

documented stable model rankings over time within a single market, this thesis reveals pronounced cross-country variation in the optimal choice of model, underscoring the importance of tailoring model selection to local market characteristics.

Crucially, the long-short portfolio returns observed here were not simply the product of exposures to established risk premia. Factor regressions using the Fama-French three-factor, Carhart four-factor, and Fama-French five-factor models consistently produced low R^2 values, typically below 5–7%, and left significant alphas intact. For example, the U.S. Ridge portfolio retained an alpha of 1.6% per month ($t = 4.58$) when considering the FF3 model, while the German Lasso portfolio achieved 2.0% ($t = 3.21$), even after controlling for momentum, profitability, and investment factors. These results echo Gu et al.'s conclusion that ML-based portfolios capture signals not subsumed by traditional factor structures, offering genuine alpha potential.

Overall, this thesis extends the evidence base for the application of machine learning in global equity selection, showing that well-regularised linear models can deliver economically significant, factor-robust returns across diverse markets, albeit with varying degrees of success. The results suggest that Ridge is particularly effective in liquid, diversified markets, while Lasso can outperform in less liquid, more concentrated environments. The generally weak performance of nonlinear models reinforces the importance of regularisation and parsimony in the design of predictive systems for equity returns. While the scope of predictors in this study was deliberately restricted, the strong alphas achieved indicate that firm-level characteristics alone can underpin profitable investment strategies. Future research could investigate whether incorporating macroeconomic conditions, sectoral effects, and more sophisticated cross-sectional interactions could further improve portfolio performance, particularly in markets such as Japan where predictive signals appear weakest.

6.Bibliography

Bianchi, D., Büchner, M., & Tamoni, A. (2021). Bond Risk Premiums with Machine Learning. *The Review of Financial Studies*, *34*(2), 1046–1089. https://doi.org/10.1093/rfs/hhaa062

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, *52*(1), 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Chen, C., & Chen, X. (2007). The information content of rights offerings in China. *Research in International Business and Finance*, *21*(3), 414–427. https://doi.org/10.1016/j.ribaf.2006.12.006

Dirkx, P., & Peter, F. J. (2020). The Fama-French Five-Factor Model Plus Momentum: Evidence for the German Market. *Schmalenbach Business Review*, *72*(4), 661–684. https://doi.org/10.1007/s41464-020-00105-y

Drobetz, W., & Otto, T. (2021). Empirical asset pricing via machine learning: Evidence from the European stock market. *Journal of Asset Management*, *22*(7), 507–538. https://doi.org/10.1057/s41260-021-00237-x

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, *47*(2), 427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (2013). A Four-Factor Model for the Size, Value, and Profitability Patterns in Stock Returns. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2287202

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, *128*(2), 234–252. https://doi.org/10.1016/j.jfineco.2018.02.012

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the Factor Zoo: A Test of New Factors. *The Journal of Finance*, *75*(3), 1327–1370. https://doi.org/10.1111/jofi.12883

Feng, G., He, X., Wang, Y., & Wu, C. (2025). Predicting individual corporate bond returns. *Journal of Banking & Finance*, *171*, 107372. https://doi.org/10.1016/j.jbankfin.2024.107372

Green, J., Hand, J. R. M., & Zhang, X. F. (2017). The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *The Review of Financial Studies*, *30*(12), 4389–4436. https://doi.org/10.1093/rfs/hhx019

Giglio, S. W. (2016). Inference on Risk Premia in the Presence of Omitted Factors. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2865922

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Harvey, C. R., Liu, Y., & Zhu, H. (2016). … and the Cross-Section of Expected Returns. *The Review of Financial Studies*, *29*(1), 5–68. https://doi.org/10.1093/rfs/hhv059

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.2307/1267351

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, *48*(1), 65–91. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

Jensen, T. I., Kelly, B., & Pedersen, L. H. (2023). Is There a Replication Crisis in Finance? *The Journal of Finance*, *78*(5), 2465–2518. https://doi.org/10.1111/jofi.13249

Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, *134*(3), 501–524. https://doi.org/10.1016/j.jfineco.2019.05.001

Lincoln, J. R., & Gerlach, M. L. (2004). *Japan's network economy: A structure, persistence, and change*. Cambridge University Press.

Lintner, J. (1956). Distribution of Incomes of Corporations Among Dividends, Retained Earnings, and Taxes. *The American Economic Review*, *46*(2), 97–113. JSTOR.

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, *47*(1), 13. https://doi.org/10.2307/1924119

Messmer, M. (2017). Deep Learning and the Cross-Section of Expected Returns. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3081555

Moritz, B., & Zimmermann, T. (2016). *Tree-Based Conditional Portfolio Sorts: The Relation between Past and Future Stock Returns* (SSRN Scholarly Paper No. 2740751). https://doi.org/10.2139/ssrn.2740751

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, *108*(1), 1–28. https://doi.org/10.1016/j.jfineco.2013.01.003

Pelger, M. (2019). *Understanding Systematic Risk: A High-Frequency Approach* (SSRN Scholarly Paper No. 2647040). https://doi.org/10.2139/ssrn.2647040

Rapach, D. E., Strauss, J. K., & Zhou, G. (2013). International Stock Return Predictability: What Is the Role of the United States? *The Journal of Finance*, *68*(4), 1633–1662. https://doi.org/10.1111/jofi.12041

Roy, R., & Shijin, S. (2018). A six-factor asset pricing model. *Borsa Istanbul Review*, *18*(3), Article 3. https://doi.org/10.1016/j.bir.2018.02.001

Sharpe, W. F. (1964). CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK*. *The Journal of Finance*, *19*(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288.

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

6.1. The use of AI

Artificial intelligence was primarily employed to support the data analysis process in R. This included diagnosing and resolving coding issues, refining scripts, identifying suitable packages for optimal efficiency, and verifying the appropriateness of the applied framework while suggesting possible improvements.

In addition, large language models were utilised for proofreading and refining the text, ensuring grammatical accuracy, enhancing coherence, and promoting the academically appropriate use of terminology and style.