OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Using ML Models to Compare Regression Models in Finance by Forecasting U.S. and European Equity Benchmarks

Matias Neves
matias.neves.perez@gmail.com

ABSTRACT

Stock performance is impossible to predict, but approximating it is. This is what financial analysts attempt every day. However, this remains a central challenge in finance that can be partially solved with the integration of machine learning models in finance, which can treat hundreds of data sets and variables while making minimal errors. However, while ML methods have been used increasingly in US equity markets, there is still hesitation regarding which regression methods are most efficient to analyse these stocks, and, moreover, there is even more uncertainty in foreign geographic markets, with minimal evidence of ML method use on European stocks. In this paper we used machine learning models to compare four different regression models – OLS, Ridge, LASSO, and ElasticNet – and to see which one performs best across US and European equity benchmarks: S&P 500 and STOXX Europe 600. We found that Ridge performed best for the US stocks with 216 out of the 471 stocks performing best under it, while LASSO dominated for the European stocks, taking 236 stocks out of the 393 total stocks. Additionally, when observing each models' RMSE and R^2 across both equity benchmarks, we saw that the models were nearly uniform for both the US and European stocks, but there were major differences between the results in S&P 500 and the STOXX 600. Specifically, the US models achieved an out-of-sample R^2 of around 0.14, when, in turn, the European models produced an out-of-sample R^2 of around -0.07, meaning that the Fama French factors had meaningful predicting power in the US, but could not explain Europe's out-of-sample return results. Ultimately, the three core factors from the Fama French model were an insufficient number of predictors to truly make a difference between the models, but that the different market structures had a major impact on the outcome of the results.

INTRODUCTION

If we knew what could happen in the future, everyone would become a millionaire. Nevertheless, predicting the future is impossible, but approximating it, with the use of the right data and models, is precisely what financial analysts attempt to achieve every day. However, acquiring this data and using the right models is not the easiest task: the financial world is full of unpredictable variables, news sentiment that can cause significant fluctuation for a stock within hours, and many more obstacles that impede financial analysts from creating accurate enough forecasts.

With the rise of artificial intelligence, many different forms of AI are integrating into many work industries. In this case, machine learning models (ML models) are increasingly being applied in financial forecasting. These purpose-built models open the possibilities of working with large data sets while producing results in a short amount of time, and working with all these variables and obstacles without making nearly as many errors as humans would make. Many of these ML methods use regression models, for instance OLS, to predict future stock returns. Moreover, more complex regression methods such as Ridge, LASSO, and ElasticNet are being used more to minimise the error between the predicted returns and the actual returns. Nevertheless, many financial analysts are unsure about which model to use, especially if they are working with different market structures.

Despite the growing popularity in the integration of ML methods in finance, most studies do not compare the regression models that these methods use to determine which model is most efficient. Moreover, not every market structure or stock performs uniformly under one regression method, which highlights the second issue: the geographic gap and the lack of research conducted on other market structures. This lack of research creates uncertainty whether model choice influences forecasting accuracy, and if the results vary across geographies.

To try to solve these unanswered issues and uncertainties, this study addresses the following research question: which regression method – out of OLS, Ridge, LASSO, and ElasticNet – best forecasts the cross-section of stock returns using the Fama French Factors, and do the results differ between US and European markets? We will first create two samples of stocks from the two equity benchmarks: S&P 500 and STOXX Europe 600, and gather the monthly stock returns from 2016-2026. Then, we will select the three Fama French factors that the regression models will use to create the most accurate predictions. Finally, we will feed all the data and information to the machine learning model to evaluate the performance of each model using RMSE and R^2 as indicators of this performance.

The paper will start off with a literature review illustrating the previous research done on the integration of ML models in finance as well as papers that have compared US stocks with European stocks. Thereafter we will present the data used during our research from the S&P 500, STOXX 600, and the Fama French model. Subsequently, the methodology will be shown with a description of all four models treated as well as their formulas. Then we will show the results given by the ML model and explain their significance. Lastly, before concluding, we will list out the limitations encountered throughout the research and what further research should be done.

LITERATURE REVIEW

The application of machine learning to stock return forecasting has grown substantially over the past decade. Sonkavde et al. (2023) provide one of the many instances of research being conducted on purpose built models used for financial forecasting. In their research they synthesize what has already been done previously in this growing trend. They conclude that ML methods consistently outperform traditional statistical models used by financial analysts, such as ARIMA. Additionally, they found that including factors like the news and social media can improve forecast accuracy by up to 18.6%. To go more into depth on actually testing these models, Sizan et al. (2023) ran models on real U.S. stock data using random forests, gradient boosting, and logistic regression, with the data they worked with consisting of historical prices, trading volumes, and economic indicators sourced from Yahoo Finance and Google Finance. Thereafter the researchers observed that, out of the three models tested, logistic regression came out on top with the highest accuracy, with the other two following slightly behind with no dramatic gap between the three. With the purpose of extending this research to international markets, Cakici et al. (2023) analysed 46 distinct stock markets to determine whether results vary depending on the market. Looking into this broader geographic pool, they found that methods like OLS outperforms more complex ML methods such as random forests and neural networks. Lehmann (2021), who has studied the evolution of the S&P 500 and STOXX 600 since 2010, further supports this claim as they observe that the two markets have grown at drastically distinct paces, mainly due to a difference in sector weighting: with the S&P being more tech heavy than the STOXX. This finding suggests that the structural difference between these two different geographic markets can influence how factor-based regression models view both markets. Factors and predictors like the ones found in the Fama French model may capture return dynamics more effectively in the US rather than Europe due to their tech heavy market structure.

Beyond purpose-built ML Models, generalist Large Language Models (LLMs) have also started to make their way into the financial industry, with researchers such as Batra et al. (2025) and Bi et al. (2024) showing that they are good tools, but require human supervision when dealing with high-stakes financial decisions. This reinforces the value of purpose-built quantitative regression approaches examined in this study, and the practicality of using AI-tools in financial forecasting.

In addition, there are not many direct comparisons between European and US stocks or stock markets, especially in recent times. Fraser & Oyefeso (2005) examine long term convergence between US, UK, and European stock markets. They discover that short-run diversification gains may occur; however, they usually are short-lived, and that this trend mainly occurs in European stock markets and less in US and UK markets. Essentially, the US market deviates from the long-term trends that can be found in European markets.

This paper is among the first to directly compare various models within machine learning, such as OLS, Ridge Regression, and more, using 9 years of data across both US and European markets. It provides a new layer in observing how ML Models can contribute to the finance industry through forecasting stock

returns. In comparing these four regression methods across two structurally different equity benchmarks, we can observe which one is the most efficient in financial forecasting for each benchmark, and if the result varies geographically. Moreover, using data from two major equity benchmarks from the U.S and Europe – S&P 500 and STOXX Europe 600 – adds another layer that will indicate if the accuracy of the models can vary depending on the market. These findings could therefore provide insight and guidance for investors looking into diversifying their portfolios, which could be beneficial to do across markets with different characteristics.

DATA

Table 1: Representative data summary statistics (Equity Indices)

Descriptive Statistics: S&P 500 vs STOXX 600 Monthly Returns (2016-2024)

Statistic	S&P 500	STOXX 600
Mean Monthly Return (%)	1.39%	1.08%
Median Monthly Return (%)	1.18%	0.94%
Standard Deviation (%)	9.34%	9.26%
Minimum Monthly Return (%)	-75.00%	-75.00%
Maximum Monthly Return (%)	75.00%	75.00%
Skewness	0.5092	0.3662

Note: Monthly returns winsorised at ±75% to remove extreme outliers. S&P 500: 471 stocks · STOXX 600: up to 432 stocks. Skewness is unitless.

Using data from Yahoo Finance and Google Finance

We can see a consistent pattern that emerges where the stocks in the S&P 500 perform better overall compared to the stocks in the STOXX 600. This is shown through a 0.21% higher mean monthly return, a 0.24% higher median monthly return, and around a higher skewness of 0.51 compared to 0.37 for the STOXX 600. This performance gap is not random, it is explained by the different market structures, which ties back to Lehmann's (2021) research where the tech heavy market structure outperforms the European stocks.

Table 2: Fama-French Three Factor Model Summary Statistics

Factor	Description	EU Mean	EU Std Dev	US Mean	US Std Dev
MKT-RF	Market Excess Return	0.61%	4.76%	0.81%	4.27%
SMB	Small Minus Big (Size)	-0.13%	1.70%	-0.29%	1.60%

HML	High Minus Low (Value)	0.42%	2.99%	0.15%	3.12%
RF	Risk-Free Rate	0.18%	0.16%	0.18%	0.16%

From Fama French 3 Factor Model

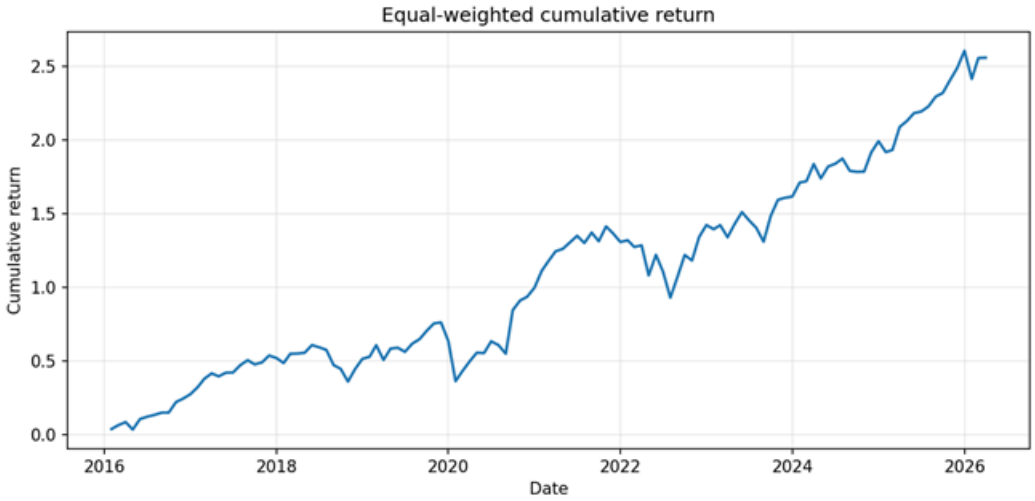
Market Excess Return (MKT-RF): It captures the amount earned by investors from being exposed to the broad market above the risk-free rate, and is also the most influential of the three factors as it accounts for the general movement of the stocks.

Small Minus Big (SMB): A measurement of the return difference between small-capitalization and large-capitalization stocks within a given month. A positive SMB means small companies outperformed large ones, reflecting the historical tendency of smaller firms to generate higher returns due to their higher risk profile.

High Minus Low (HML): It measures the return difference between value stocks, for instance ones with a high book-to-market ratio, and growth stocks, which are companies with a low book-to-market ratio. A positive HML means value stocks outperformed growth stocks, capturing the value premium.

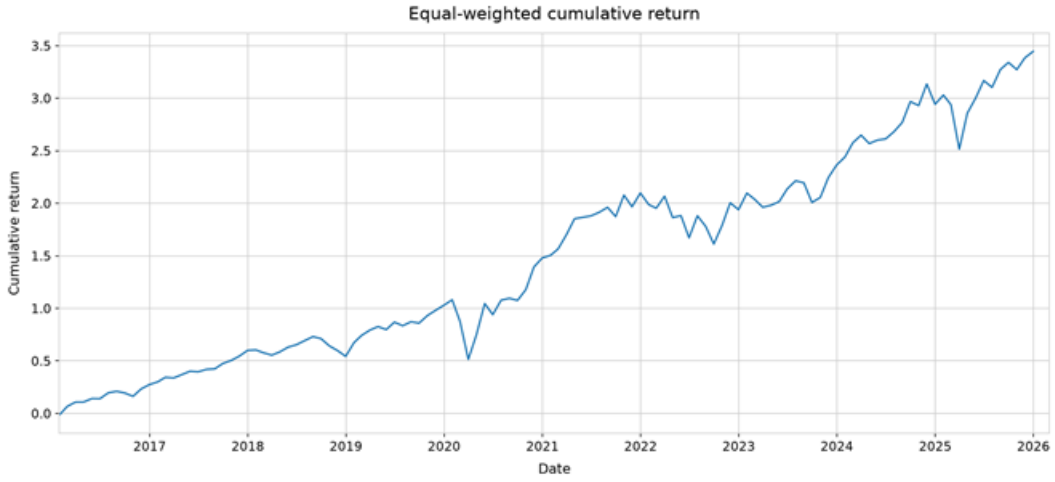
Risk Free Rate (RF): The return an investor can earn with zero risk, represented by the one-month US Treasury bill rate. It serves as the baseline return against which all investments are compared, and is subtracted from stock returns to calculate excess returns throughout this study.

Figure 1: Equal-Weighted Cumulative Returns from 2016-2026 (STOXX Europe 600)



This graph shows that the STOXX 600 universe returned roughly 250% throughout our 2016-2026 time period, with the COVID crash clearly visible in early 2020.

Figure 2: Equal-Weighted Cumulative Returns from 2016-2026 (S&P 500)



This chart shows the evolution of the equal weighted cumulative return of the S&P 500 stocks since 2016. In the graph we can see that the S&P 500 equal-weighted portfolio grew a cumulative return of over 350% from 2016-2026.

METHODOLOGY

Fama French Model

The Fama French three factor model is a model that explains stock returns using market risk, company size, and value factors, which extends the traditional CAPM model. Eugene Fama and Kenneth French developed this model in 1992 and deduced that these three factors, out of the millions of predictors that could be used to justify a stock's return, are the most significant for determining this. An important discovery Fama and French found in creating their models is that small companies and value stocks tend to outperform larger companies and growth stocks. If we only used the CAPM model to evaluate this conclusion, the results would be different and less detailed.

Ordinary Least Squares (OLS)

Ordinary Least Squares (OLS), also known as linear regression, is the simplest model out of the 4 models examined in this study. The goal of OLS is to minimize the sum of the residuals, which is the difference between the observed values – the data plotted on the graph – and the predicted value – the line of best fit. Creating this line of best fit that aims to minimise this distance between the line and the data allows people to forecast and predict future outcomes by extending the line and reading the data as a simple linear equation. However, this is only the case if there is one factor that we are observing. Since we are using the Fama-French factors, there are three different factors that must be accounted for; therefore, we must use the following equation to create this line.

$$\beta OLS = \text{argmin}\beta\ \Sigma(Y - \beta \cdot X)^2$$

Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO regression is very similar to OLS; however, it solves a crucial problem that OLS has: each of the factors it uses to make predictions has the same value and importance. In the real world, nothing is equal and there is always going to be one thing that is more significant than the other depending on the situation. LASSO solves this problem by giving a punishment on the less important factors but canceling them out completely.

$$\beta LASSO = \text{argmin}\beta\ \Sigma(Y - \beta \cdot X)^2 + \lambda\ \Sigma|\beta i|$$

Ridge Regression

Ridge regression does the exact same thing as LASSO, but instead reducing the importance of the non-significant factors all the way to 0, it minimises it close to 0 but never absolutely. In the equation, the difference can be observed with LASSO using the absolute value of the punishment, and Ridge using the square of the punishment.

$$\beta Ridge = \text{argmin}\beta\ \Sigma(Y - \beta \cdot X)^2 + \lambda\ \Sigma\beta i^2$$

ElasticNet Regression

ElasticNet Regression is the most complicated out of the 4, and is a regression method that combines the penalties from both Ridge and LASSO regression. In doing so it provides solutions to the limitations found in both methods.

$$\beta ElasticNet = \text{argmin}\beta\ \Sigma(Y - \beta \cdot X)^2 + \lambda 1\ \Sigma\beta i^2 + \lambda 2\ \Sigma|\beta i|$$

RMSE

RMSE indicates how far the predicted value on average deviates from the actual result; ultimately, allowing us to assess the average size of the prediction error; therefore, the model's accuracy. It is calculated as the square root of the average squared prediction errors. A lower RMSE value indicates a better-fitting model, while a higher one means that there are larger deviations between the prediction and the actual result.

R^2

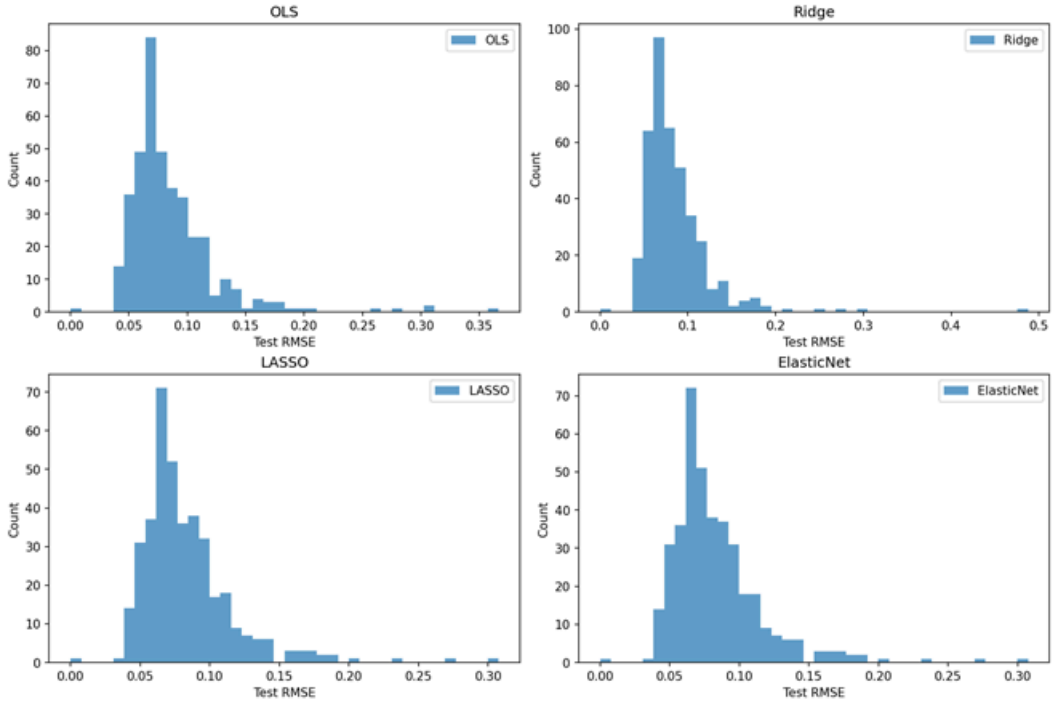
R^2, also known as the coefficient of determination, shows how well a regression model explains the variation in the dependent variable compared to a baseline that always predicts the mean.

1 – Model explains variation
0 – Model is no better than predicting the mean
Negative – Model performs worse than the mean predictor

RESULTS

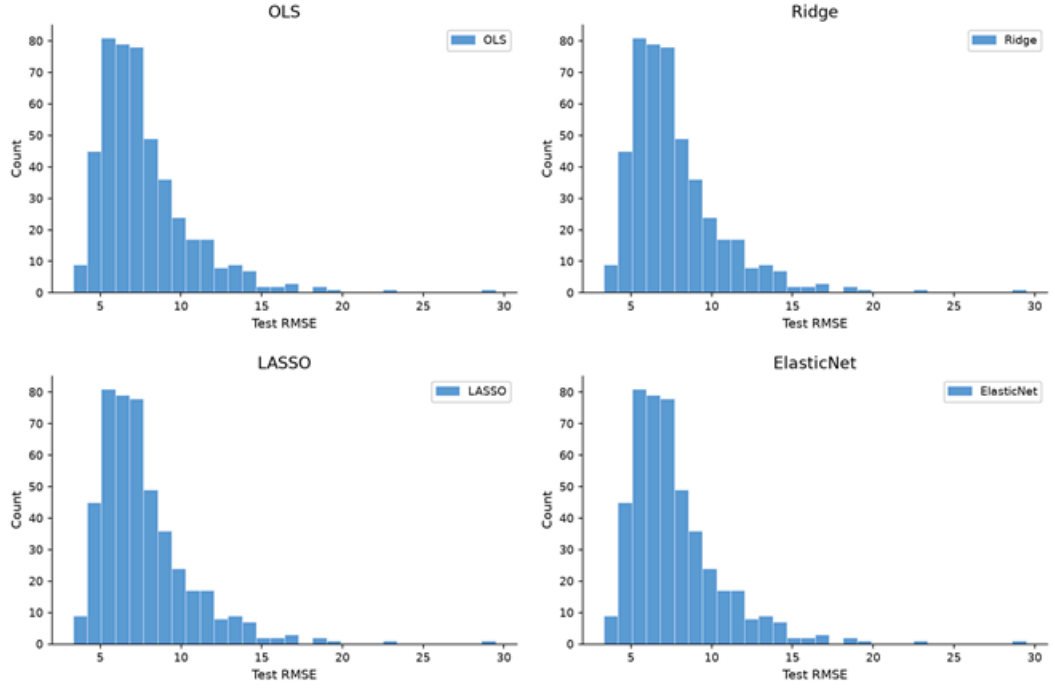
RMSE (Root Mean Squared Error) Results

Figure 3: Per Model RMSE Histograms (STOXX Europe 600)



Note: The values in the x-axis are expressed in decimals (0.50 ⇔ 50%)

Figure 4: Per Model RMSE Histograms (S&P 500)



Note: The values in the x-axis are directly expressed in %

We can see that while the EU OLS histogram peaks at around 6-7% and has a wider spread going up to 37%, the U.S OLS histogram peaks at 5-6% and has a slightly tighter distribution with most stocks clustered around 4-10%. Setting itself apart from the other models in the EU, the EU Ridge histogram has a tighter peak around 7-9% and a longer tail that extends to 50%. This indicates that the Ridge model in Europe performs poorly when forecasting individual European stocks, even more than LASSO or ElasticNet. However, the US histograms seem quite uniform across all models, meaning that the model choice has little impact on the distribution of stock errors in the US market.

Ultimately, in both markets, the four models are almost identical and similar to one another, confirming that the model choice has minimal impact on the prediction errors, and that no model dramatically outperforms the others at the stock level distribution. However, it is important to recognise the fact that whatever small differences there are between the results for each model, it is due to the choice in different regression methods.

Regression Models Performance Results

Figure 5: Best Model Per Stock (STOXX Europe 600)

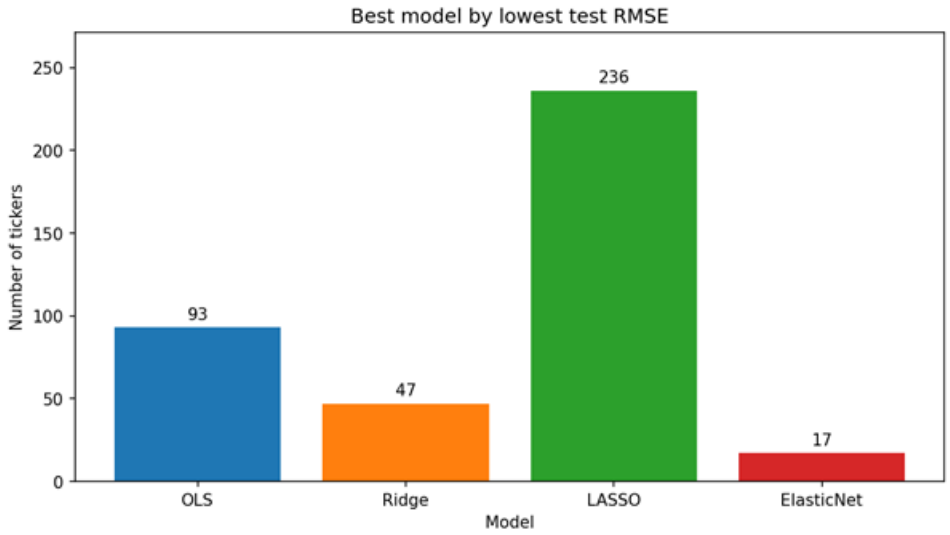
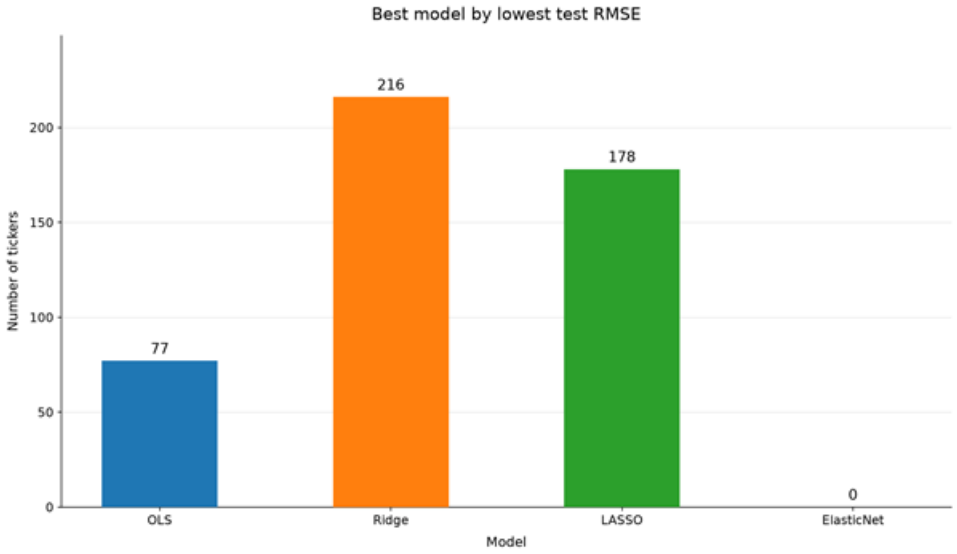


Figure 6: Best Model Per Stock (S&P 500)



In the EU chart, LASSO dominates all other models with a staggering 236 stocks out of the total 393 (around 60%). OLS comes in second with 93, Ridge in third with 47, and ElasticNet in last with only 17.

When looking at the US chart, we can see a clear distinction between the results of the EU stocks and the US stocks. Unlike the EU results, Ridge dominates the other 3 models with 216 stocks out of the 471 total. Following closely behind, LASSO comes in second with 178 stocks, then OLS with 77 stocks, and lastly ElasticNet with a surprising 0 stocks.

The first distinction that should be flagged between the US and European stocks is that the "winner" reverses between markets, with Ridge being the winner for the US stocks, and LASSO for the European stocks. The dominance of Ridge among the US stocks points out that the best strategy would be to keep all the Fama-French factors but shrink the coefficients of the less important ones specific to the stock. With the superiority of LASSO among these four models in the EU stocks, we can deduce that the best method is to eliminate one or more of the three Fama-French factors entirely and rely solely on the factors that matter for the specific stock being forecasted. Essentially, this reversal tells us that, in the US, the three factors – market risk, size, and value – are relevant among the majority of US stocks, no matter if their level of influence is reduced; however, for the EU stocks, not all the factors are necessary and, therefore, can be eliminated entirely based on their relevance with the stock.

An eye-catching observation is that, for the US stocks, ElasticNet disappears entirely. Compared to Europe's 17 stocks, there is no US stock that has the smallest forecast error when using ElasticNet. This is

most likely due to the fact that we are only forecasting these stocks with three factors, and the middle ground between Ridge and LASSO turns out to be unnecessary with such a small number of factors to account for.

R^2 by Model Results

Figure 7: R-Squared by Model (STOXX Europe 600)

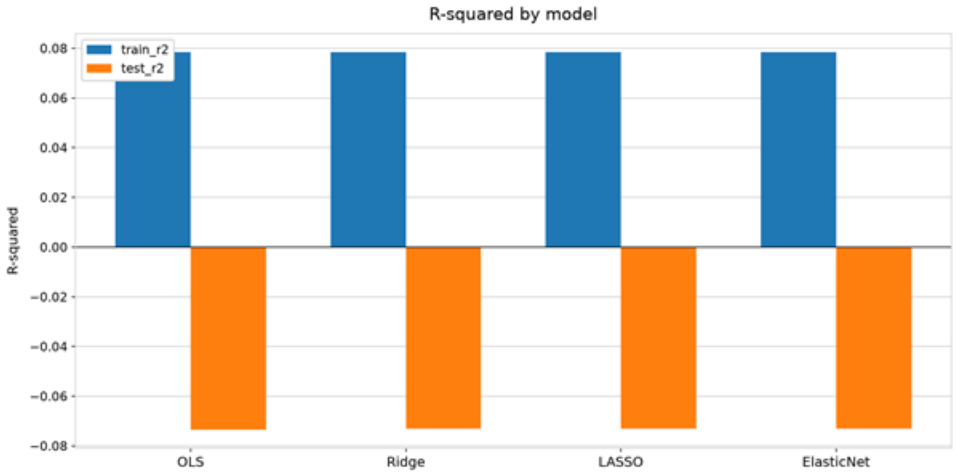
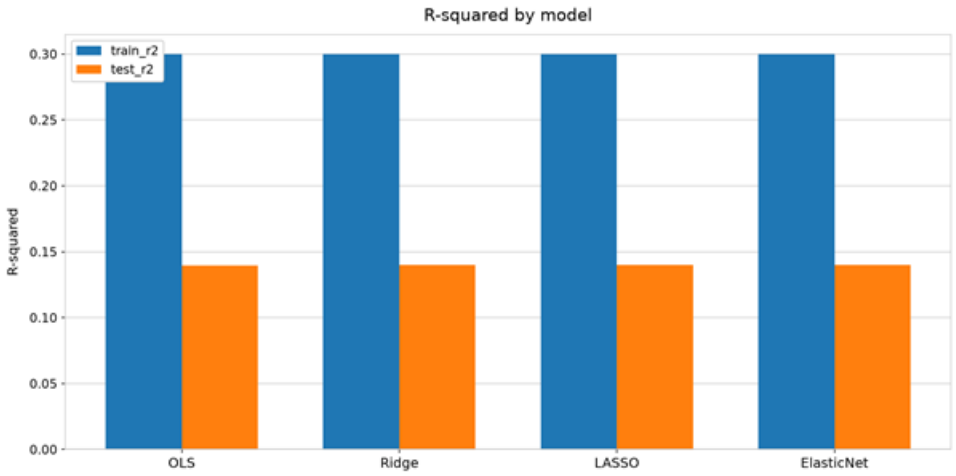


Figure 8: R-Squared by Model (S&P 500)



In both the US and EU stocks, all four models have identical training R^2. For the EU, the training R^2 is 0.08, and for the US, the training R^2 is 0.30. An important observation is that the US training R^2 is nearly four times higher than the European R^2, indicating that the Fama French factors have substantially more power in the US stock forecast. Another remark is that, since there are only 3 factors taken into account,

the regularization penalties have minimal impact on the stock forecasts, which is why all four models are identical in both the US and European in-sample stocks.

For context, this low-data environment created by using monthly observations in a nine-year period (roughly 108 observations) illustrates a constraint on the results as this dataset is relatively small for ML applications. Regression models like Ridge, LASSO, ElasticNet have limited opportunity to show their advantages over OLS due to this constrained time frame and relatively small dataset. The scarcity of observations restrains the models from learning the relationships between the three core factors from the Fama French model. Consequently, this is why the regularization models were unable to outperform the historical baseline in the European test set.

The clear distinction is made when comparing the test R^2 between the US and European stocks. In the US the R^2 test shows approximately 0.14 across all four models. However, the European stocks' R^2 holds at -0.07 across all four models, meaning that the models perform worse than the benchmark that predicts the historical average return every month. This means that the factors are counterproductive in the test period; consequently, causing the predictions to be worse than no model at all.

Additionally, we can also see that, in both markets, all four models produce almost identical R^2 results for both the training and test sets. This allows us to determine the difference in R^2 across all models in both markets is not driven by model choice, but by market structure. Fundamentally, this comparison between the R^2 results from the European stocks and US stocks indicate that model performance varies systematically across US and European markets.

LIMITATIONS AND FURTHER RESEARCH

The first and most important limitation that should be flagged is the lack of factors being used – MKT-RF, SMB, and HML – while using complex models such as Ridge, LASSO, and ElasticNet. Regularization models such as Ridge, LASSO, and ElasticNet are designed to handle these high-dimensional environments where the number of predictors and factors is large relative to the number of observations. These three regression models are used to give power to more important factors, and reduce or eliminate the coefficients of less relevant factors. However, with this small number of factors to work with, the regularization penalties in Ridge, LASSO, and ElasticNet have little to no variance to restrict, which causes the four models to converge towards the same results. Ultimately, this explains why in most of the core findings all four models produced virtually identical results, including the European and US RMSE histograms and the R-squared charts for both market structures.

A further limitation is regarding accuracy and a scarce number of observations: using monthly data rather than daily, and using only roughly 400-500 stocks for each equity benchmark. Using monthly data since 2016 only gives us roughly 108 observations per stock. While this is sufficient for basic estimations, it will not allow us to observe possible spikes or crashes within a single month, which is why we lose accuracy in choosing the monthly data. The same goes for the 400-500 stocks chosen for each equity benchmark. While these stocks provide a general insight on the distribution of results for the performance

of each model, using more stocks in this analysis, especially smaller-capitalisation and more recently listed firms, would allow us to see which model performs best across many different types of stocks. Furthermore, including more observations would allow these more developed regression models to establish more reliable statistical relationships between the three factors and individual stock returns.

In the future, to address these limitations, further research should be done with an expanded factor set, and a higher data frequency. The immediate extension would naturally be to include richer predictors and factors so that the regression models such as Ridge, LASSO, and ElasticNet can be used to their full potential and punish or eliminate the necessary factors. Some factors outside of the three Fama French could be as simple as stock-level momentum, volatility, and news sentiment, or even more complex predictions such as Robust Minus Weak (RMW), which measures the return difference between profitable and non-profitable firms, and Conservative Minus Aggressive (CMA), which calculates the difference between firms that invest aggressively or conservatively. Subsequently, adding a higher data frequency, for instance from monthly to daily, would open the door to more detailed observations that could not have been seen otherwise in the monthly data.

CONCLUSION

Throughout the study we have set out to answer the following question: which regression method – out of OLS, Ridge, LASSO, and ElasticNet – best forecasts the cross-section of stock returns using the Fama French Factors, and do the results differ between US and European markets? When forecasting stock performance, the best regression method varies depending on the market structure and the factors chosen to predict the stock performance. As we saw from the results, Ridge regression performed best for the stocks deriving from the S&P 500, and LASSO regression performed best for the stocks deriving from the STOXX Europe 600. We observed this through the RMSE histograms and model performance charts which showed that Ridge achieved the lowest RMSE for 216 out of the 471 S&P 500 stocks, and LASSO achieved the same for 236 out of the 393 STOXX 600 stocks, showing that the results vary because of these different market structures. This conclusion is confirmed with the two R^2 charts that show major differences in both the training period of the models and, principally, the testing period. Along with the informative results provided by these charts, they also highlight one of the main limitations flagged during the research: the lack of factors and predictors to forecast these stocks across different models. In short, Ridge, LASSO, and ElasticNet are all regression models that punish and give power to different factors based on their importance relative to each stock. Since we are using only the three factors found in the Fama French model, these three models are not used to their full potential. This is the main reason why most of the results across all models are uniform or very similar.

This paper is among the first to directly compare different regression models being used in finance and compare the results across different market structures. In doing so, it provides insight on how AI, and more specifically machine learning models, can be used in finance and financial forecasting. The results gathered in this research can be used to inform financial analysts on what regression model is most efficient across US stocks and European stocks; ultimately, opening the door to investors diversifying

their portfolios to other stock markets like Europe. In the future, it would be useful to determine why there are different causes to these different results and why different regression models perform inconsistently across different markets. Along with this, this future research should be conducted with additional factors and predictors taken into account to maximise the potential of these more complex models such as Ridge, LASSO, and ElasticNet. By demonstrating that the optimal regression method varies systematically across different market structures – with Ridge dominating in the US and LASSO in Europe – this paper provides evidence that factor-based machine learning models must be selected and calibrated according to the specific market in which they are applied, rather than treating any single model as universally transferable across international equity environments.

REFERENCES

Batra, Devesh, et al. "A Review of LLM Agent Applications in Finance and Banking." SSRN, 2025, papers.ssrn.com/sol3/papers.cfm?abstract_id=5381584.

Bi, Shuochen, Jue Xiao, and Tingting Deng. "The Role of AI in Financial Forecasting: ChatGPT's Potential and Challenges." Proceedings of the 4th Asia-Pacific Artificial Intelligence and Big Data Forum, 2024.

Cakici, Nusret, et al. "Machine Learning Goes Global: Cross-Sectional Return Predictability in International Stock Markets." Journal of Economic Dynamics and Control, vol. 155, 2023, p. 104725.

Fraser, P., and O. Oyefeso. "US, UK and European Stock Market Integration." Journal of Business Finance and Accounting, vol. 32, 2005, pp. 161–181, doi.org/10.1111/j.0306-686X.2005.00591.x.

French, Kenneth R. Kenneth R. French Data Library. Tuck School of Business at Dartmouth, mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Accessed 31 May 2026.

Lehmann, Kai. "S&P 500 vs. Stoxx 600: Why the Leading US Equity Index Has Outperformed Its European Counterpart." Flossbach von Storch Research Institute, 2021.

Sizan, Mir Mohtasam Hossain, et al. "AI-Enhanced Stock Market Prediction: Evaluating Machine Learning Models for Financial Forecasting in the USA." Journal of Business and Management Studies, vol. 5, no. 4, 2023, pp. 152–166.

Sonkavde, Gaurang, et al. "Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications." International Journal of Financial Studies, vol. 11, no. 3, 2023, p. 94.

ACKNOWLEDGEMENTS

The author would like to express their gratitude to Vicente Garcia for their guidance and mentorship throughout this project. Their expertise in quantitative finance and machine learning gave valuable direction on how to write and structure the paper.