## Advanced Programming 2025

# Sector Rotation in Financial Markets with Machine Learning: A Forecasting-Based Model Comparison

Final Project Report

Lukas Tonkovic
lukas.tonkovic@unil.ch
Student ID: 22608855

January 9, 2026

**Abstract**

This paper investigates whether different machine learning models can predict both the magnitude and the sign of monthly stock returns of various financial sectors using historical data. Using monthly returns for selected sectors in the time span from 2005 to 2025, three machine learning models are compared based on their predictive robustness: Linear Regression, Ridge Regression, and Random Forest. The look-ahead bias is controlled by using a time-series-aware evaluation framework; therefore, predictions are always based only on past data. Evaluation of the mentioned models is done by using a fixed out-of-sample train–test split, where additionally the rolling and expanding window analyses are employed to the baseline Linear Regression model to capture robustness under more advanced training schemes. The important takeaway is that the higher complexity of ML models does not imply better performance; indeed, the paper's findings outline that they often underperform. Random Forest sometimes exhibits slightly lower forecast errors, but these improvements are not translated into real and consistent economic benefits. Moreover, none of the models demonstrates consistent or economically significant predictive performance that could be continuously exploited. The central findings are that the project simulates the real-investment environment using the machine learning techniques in building a sector-rotation strategy and provides results that are in line with the Efficient Market Hypothesis (EMH), where consistent abnormal excess returns generation is difficult to achieve and to identify.

**Keywords:** data science, Python, machine learning, sector rotation, financial markets, efficient market hypothesis, financial forecasting

# Contents

# 1 Introduction

Financial markets' nature captivates the researchers and investors due to the unpredictability of asset returns and the difficulty of correctly predicting the movements of market prices. One common strategy that deals with this challenge is known as sector rotation, where an investor alternates between sectors, depending on the cycles, in order to enhance the portfolio performance[1]. In recent years, the rapid development of data availability and machine learning has profoundly renewed interest in reexamining the stock market and its characteristics and possibly finding a path that would lead to evident improvements in capital allocation approaches.

However, machine learning has not yet demonstrated the ability to consistently anticipate the market's fluctuations and all factors that capture risk and shape prices. Moreover, the Efficient Market Hypothesis (EMH) is particularly connected with this topic. The EMH emphasizes the immediate adjustment of stock prices when new information emerges, therefore preventing investors from spotting undervalued stocks and, over the long term, earning excess returns for the same level of risk, despite having information and conducting analysis[2].

The research question of this project is to examine whether, or to what extent, the machine learning models can predict the magnitude and sign of monthly stock returns in specific sectors, utilizing only historical data. More precisely, the goal would be to discover whether the machine learning models can anticipate these metrics in a consistent and investment-beneficial manner, which could be economically meaningful for investment decision-making. The time-series nature of this problem is clearly addressed, thus preventing the look-ahead bias, which would lead to unreliable conclusions. Under realistic forecasting conditions, this project compares the performance of three machine learning models commonly used in similar areas of study: Linear Regression, Ridge Regression and Random Forest. By employing rolling and expanding window evaluation schemes to the Linear Regression, the project aims to assess whether the increased complexity of the model indeed increases the probability of more precise prediction, which would ultimately be robust over time.

This paper is organized as follows. The Literature Review section places the project in the context of existing literature and knowledge and compares insights and conclusions from other studies. The Methodology section describes the data relevance and its collection process and introduces models and specifies sectors that will be analyzed. The Results section shows empirical findings and further explains them graphically through tables and figures. In the Discussion section, the results will be interpreted and compared with other scientific findings on the same topic. Finally, the Conclusion summarizes the central findings of the project, discusses the possible limitations, and outlines directions for future research that could improve the field.

# 2 Literature Review / Related Work

Nowadays, machine learning's development plays a crucial role in enhancing and transforming many industries, including finance. Machine learning (ML) is described as a set of algorithms that learns the structure from the data and identifies and learns complex patterns, all in a high-dimensional environment[3]. Moreover, what differentiates ML from econometrics is its less ex-ante nature and more data-driven approach, resulting in a much weaker dependence on predefined distributional assumptions. In that way, ML can alleviate many limitations of econometrics or similar traditional approaches and capture some more advanced relationships[3]. Therefore, it is not surprising that the finance sector has witnessed a clear shift towards various artificial intelligence form implementations, which increasingly complement the traditional theoretical approaches in finance.

The AI component improves the finance field by providing a better process of information, easier data analysis, and more accurate predictions due to the capability of identifying relationships in large data[4]. Various studies have provided evidence that ML algorithms typically outperform the standard traditional methods of predicting data by being more accurate[5]. However, finance, and concretely the asset pricing field that we examine in this study, deviates in its characteristics from other areas where ML is usually employed. In particular, in standard applications the target of predictions is usually observable, while the expected returns forecasted in asset pricing implementations are unobservable and contain a lot of noise in data. Due to their availability and observability, realized returns serve as the proxies, but they are not completely accurate ones. This results in a significantly low signal-to-noise ratio, relative to other applications where ML is used, which typically does not improve even with larger datasets[6].

Other challenges ML faces in asset pricing are, besides the usual small data available, the weak forecasting strength due to market efficiency and the evolving nature of financial markets, where recognized patterns can vanish as the markets change[7].

All things considered, ML appears to be not just a current trend but a field at its early stage[8] and with a huge potential of evolving into an inevitable and widely used practice in finance[9].

# 3 Methodology

## 3.1 Data Description

Data used in this project is constructed at a monthly frequency and combines the sector-level financial metrics together with macroeconomic variables. Five financial sectors are examined in the project and obtained through the yfinance library. Namely, sectors GLD, VNQ, XLE, XLF, and XLK are exchange-traded funds (ETFs) used as proxies for commodities, real estate, energy, financials, and technology sectors, respectively. Macroeconomic metrics are gathered from the Federal Reserve Economic Data (FRED) database, utilizing the fredapi interface. The macroeconomic indicators describe inflation dynamics, the monetary policy environment, labor market strength, economic activity, and market uncertainty.

Table 1: Financial sectors included in the sector rotation strategy.

| Ticker | Sector |
|---|---|
| XLK | Technology |
| XLF | Financials |
| XLE | Energy |
| VNQ | Real Estate |
| GLD | Gold |

The raw datasets of financial sectors and macroeconomic metrics cover the period from January 2005 to July 2025 on a monthly basis. After reorganizing data and temporally aligning values of sector performance and macroeconomic figures, the final processed dataset spans from February 2005 to July 2025, also on a monthly basis. Therefore, the final tailored dataset results in 246 observations and systematically depicts in one place 9 macro variables and prediction targets for each sector separately. It is crucial to mention that although prediction targets are computed for all time periods, only the historical data is available at the time of forecast, due to the expanding and rolling windows that the project employs in the training and evaluation of models. All observations are in the form of a monthly time-series dataset, where explanatory variables are macro indicators that are defined by the market environment, and sector returns represent endogenous variables.

Inflation dynamics are observed in explanatory variables, the Consumer Price Index (CPI) and CPI year-over-year. Variables that reflect the monetary policy dimension and depict characteristics of the yield curve are the following: The Federal Funds Rate, 10-year Treasury yield, 2-year Treasury yield, and the 10Y–2Y term spread. The unemployment rate and industrial production are exogenous variables that concentrate on labor market conditions and demonstrate economic activity. The crucial element of risk and market uncertainty is embedded in the VIX (volatility) index. All of these variables provide a strong and comprehensive representation of the market setting that influences asset prices and therefore also the returns. The data was controlled and checked prior to the analysis, but none of the metrics were removed since the unstable and unpredictable nature of financial markets results in volatile and perhaps extreme values, which can contain important information for predictions and decision-making. Overall, the dataset is clean, reliable, and suitable for the experimental setup of the project.

Table 2: Macroeconomic variables used as predictive features.

| Variable | Description |
|---|---|
| cpi_index | Consumer Price Index (price level). |
| cpi_yoy | Year-over-year CPI inflation rate. |
| fed_funds_rate | Federal Funds Rate. |
| yield_2y | 2-year U.S. Treasury yield. |
| yield_10y | 10-year U.S. Treasury yield. |
| term_spread_10y_2y | Term spread between 10-year and 2-year Treasury yields. |
| unemployment_rate | Unemployment rate. |
| industrial_production | Industrial production index. |
| vix | Market volatility index (VIX). |

### 3.2 Approach

The research question is approached by employing three ML models and evaluating them to conclude whether the more complex models indeed lead to more robust and significant results. Through the lens of predictive performance and statistical metrics, Linear Regression, Ridge Regression, and Random Forest are compared. The question that is central to answer is whether the consistent and strong relationship between the past macroeconomic variables and future asset movements can be detected. Linear Regression is the simplest model present in this study, upon which, exclusively, the rolling-expanding training schemes will be employed in order to improve the analysis. To enhance the linear features and to reduce overfitting, Ridge Regression is introduced. With the L2 regularization, all of these challenges, together with the multicollinearity, are removed. The potential non-linear association is examined by the Random Forest model, which represents the most complex model in this study.

Data preprocessing was a crucial step prior to the analysis. For instance, monthly adjusted closed prices for each sector, in raw datasets, were translated into asset returns, which can be seen in processed datasets at Github. The same was done with the macroeconomic metrics, which were all aligned time-wise to correspond to sectoral monthly asset returns. Additional modifications made to improve the explanatory power were the computations of the CPI year-over-year rate (to better explain inflation dynamics) and the construction of the term spread, defined as the difference between the 10-year and 2-year Treasury yields.

The data was constructed with no ad hoc cleaning or removing values since, for the chosen timespan, all of the variables were fully contained. The out-of-sample nature of this analysis is central for proper conduct and reliable findings. The out-of-sample error therefore presents the difference between the predicted and actual sector returns, based on data that is known but not observed at the time of analysis. The prevention of too optimistic and overfitted conclusions is done by using the rolling and expanding window methods while continuously, in all 3 ML models, strongly adhering to the out-of-sample performance.

### 3.3 Implementation

Python was used for the programming aspect of this project. In the process of data manipulation and data collection, many libraries were implemented. Besides numpy and pandas, yfinance was a central library for retrieving sector monthly adjusted closed prices, which were then transformed to asset returns. The FRED database was used for the collection of macroeconomic variables. Libraries such as fredapi and scikit-learn were crucial for the implementation and evaluation of all ML models. More precise information on the dependencies present in this project can be found in the requirements.txt file, at Github repository.

The GitHub repository follows mainly the same structure as the conduct of the study. Starting with the raw data (data collection) and advancing to data preprocessing, it reflects the realistic stages of the research. Ending with data modeling and evaluation, one can find all intermediate and final findings in the results folder. With the execution of the central script (main.py), one gets the information on the performance of all three ML models, together with their comparison. Moreover, the results connected to the rolling and expanding window schemes of the Linear Regression model are displayed.

## 4 Results

### 4.1 Experimental Setup

All of the computations were done in the Python environment, more precisely VS Code. The emphasis is put on forecasting monthly returns of financial sectors, strictly adhering to realistic conditions following an out-of-sample manner. The 3 machine learning models (Linear Regression, Ridge Regression and Random Forest) use the fixed interval from February 2005 to December 2018 as a training period and the interval from January 2019 to July 2025 as a test period, based on which the model performance is evaluated. Along with the fixed training and test dataset, in order to increase robustness, the rolling forecasting schemes are applied to the Linear Regression model. Two main strategies were incorporated. An expanding window is the first strategy, where simultaneously the training dataset grows as the new information becomes available. A rolling window is the second approach, which constantly reevaluates models with the most recent information. The results of the model performance are reviewed using various criteria. From the statistical perspective, the directional accuracy and forecast error are criteria used for the evaluation of the model. From the economic perspective, forecasted returns are used to form a sector-rotation strategy whose performance is compared to the passive market investment, represented by the benchmark (SPY).

## 4.2 Performance Evaluation

All models show a very limited predictive strength. Regarding each sector individually, they display small and even negative values of $R^2$, which suggests a poor forecasting power in comparison with the simple historical mean approach. Random forest does occasionally have smaller negative values of $R^2$ in contrast with the other two models, but nothing significant or relevant for investment decision-making. Detailed sector-level $R^2$ values are found in Table 4 in the Appendix.

Interestingly, the Linear and Ridge Regression do achieve modest metrics when it comes to directional accuracy, lying typically between 50 and 62 percent, depending on the sectors. Despite these results and the ability to predict a return direction to some extent, the improvement from the random guessing remains mostly unachieved. On the other side, Random Forest exhibits lower directional accuracy of 45 to 53 percent, signaling that increased model complexity does not always result in more confident and robust performance. The directional accuracy metrics for all models can be found in the Appendix, Table 5.

Analyzing the performance from the economic context, all three models perform poorly when evaluating the sector-rotation strategy. The hit rate is at approximately 18 percent, highlighting that the forecast of the best-performing sector is hard to get correctly and consistently. In terms of generating excess returns, Linear Regression stands out as having the best results, with generating near-zero excess returns in comparison with the market benchmark. Ridge regression performs slightly inferior, while the Random Forest demonstrates the worst results. More precisely, it generates major negative excess returns relative to the market benchmark (SPY) and a negative Sharpe ratio. Moreover, the Linear Regression obtains the highest (positive) Sharpe ratio. All of these results are presented in Table 3.

Results derived from the rolling-window evaluations under the baseline Linear Regression model further underline the sensitivity of model strength and robustness and its dependence on the chosen training scheme. The most favorable results are seen in the expanding-window approach, which demonstrates a small but positive excess return and a positive Sharpe ratio. What is more, its hit rate of 21 percent is the highest derived hit rate in this study. Conversely, the fixed 60-month rolling window substantially underperforms in comparison with the market SPY benchmark, indicating that shorter training horizons even intensify the noise in data and its estimation, together with the regime unpredictability. The mentioned findings are reported in Table 6, in the Appendix.

Table 3: Out-of-sample performance of sector rotation models relative to the SPY benchmark (2019–2025).

| **Model** | **Hit Rate (%)** | **Cumulative Excess Return (%)** | **Annualized Sharpe (Excess)** |
|---|---|---|---|
| Linear Regression | 17.72 | -0.08 | 0.07 |
| Ridge Regression | 17.72 | -3.78 | 0.03 |
| Random Forest | 17.72 | -38.63 | -0.28 |

## 4.3 Visualizations

Figure 1 depicts the low forecasting performance of all three models when compared to the SPY benchmark. More precisely, the cumulative excess returns derived from the out-of-sample test period are negligible or even negative when compared to the SPY. The significant underperformance of the Random Forest model is in accordance with the results provided in Table 3.

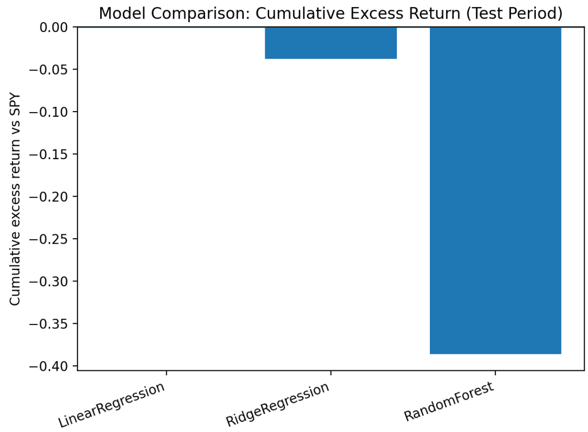


Figure 1: Cumulative excess returns relative to SPY for each model over the out-of-sample test period.

Figure 2 compares the cumulative excess returns of two rolling training schemes under the baseline Linear Regression: the expanding-window and the fixed 60-month rolling-window. As can be observed, the expanding-window forecasting scheme presents modest but positive excess returns over time. On the other hand, the fixed rolling-window strategy displays high volatility and unpredictability, which signals constant underperformance. This figure reconfirms that shorter time periods amplify noise in data and make it harder to correctly forecast financial data. Furthermore, these observations align with the findings in Table 6, located in the Appendix.

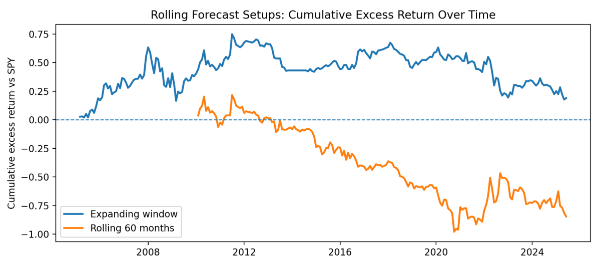


Figure 2: Cumulative excess returns over time for expanding and fixed 60-month rolling window forecasting schemes.

Figure 3 displays annualized excess Sharpe ratios estimated from three ML models. The figure reasserts the already stated findings. Sharpe ratio values are positive and close to zero for the Linear and Ridge Regression, while the Random Forest model exhibits large and negative ones. This suggests that none of the models present an enhanced risk-adjusted performance over the benchmark. Therefore, a sector-rotation strategy hardly yields consistent and positive excess returns.

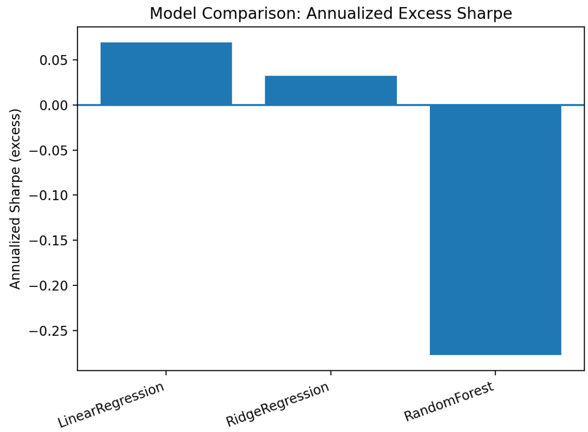


Figure 3: Annualized excess Sharpe ratios of the evaluated models during the test period.

## 5 Discussion

Findings of this study highlight the inherently unpredictable nature of the asset price movements, which is in line with the postulates of the EMH. Moreover, the study shed light on the effectiveness and robustness of machine learning models in forecasting financial developments under realistic conditions.

First, the observed predictability performance is notably limited due to the fact that in the majority of models, the out-of-sample $R^2$ values are predominantly negative. Furthermore, in the best models, directional accuracy metrics are just slightly outperforming the historical mean approach. These empirical findings are in line with other finance literature, outlining the challenge of anticipating future stock movements, particularly over the short time horizon. That suggests that the noise in historical data even impedes advanced machine learning models from extracting valuable insights from the data and recognizing stable patterns.

Second, the comparison of linear and nonlinear models gives rise to captivating insights. The linear models have consistently demonstrated better statistical metrics, especially when it comes to directional accuracy values. Random Forest has exhibited slightly better error-based measures, but its implication on the real investment setting is ambiguous and rather limited. Therefore, the poor signal quality has led to weaker performance of more complex models, concluding that more sophisticated models are not automatically indicative of better performance.

Third, the economic value of the models has a very limited significance as well. Despite that some models demonstrate better forecasting features and directional accuracy above chance, the consistent prediction of the best-performing sector is where the models fail to excel. Hit rates across models exhibit too low values to be valuable, and excess returns are almost of negligible magnitude when compared to the market benchmark. These results imply that the modest statistical enhancement of the models is not sufficient to translate into meaningful economic gains in the real-investment setting.

The rolling-window schemes, connected to the Linear Regression model, also outline interesting conclusions regarding the financial market behavior. The expanding-window setup distinguishes itself as the best performing, generating small but positive excess returns and obtaining a positive Sharpe ratio. However, the real-world implications remain modest and not completely stable. The fixed 60-month rolling window, on the other hand, underperforms and produces negative excess return, positioning itself as impractical. Primarily, the cause is its short-horizon characteristic, which is particularly sensitive to the estimation errors. All of these suggest that in order to preclude the overinterpretation of certain results, the analysis should be methodologically diverse to capture as many different points of view as possible to get the most correct and complete picture possible.

## 6 Conclusion and Future Work

This paper unveils the effectiveness of machine learning models in forecasting monthly returns of financial sectors, built on historical macroeconomics knowledge. Specifically, the sector-rotation strategy is tested under three ML models: Linear Regression, Ridge Regression and Random Forest. The forecasting environment was constructed on realistic conditions, incorporating out-of-sample performance and avoiding look-ahead bias. The central objective was to assess whether the ML models are capable of providing superior results both statistically and economically, in terms of valuable insights and identifying patterns that would challenge the implications of market efficiency and consistently generate excess returns. Overall, the study's findings align with the EMH, confirming the stated limitations of ML implementation in finance mentioned in other academic works. Certainly, the specific patterns exist in historical data, but they are hard to identify, they evolve fast, and they are hard to exploit in a consistent manner to earn high excess returns. This is best observable in the poor performance of ML models when it comes to hit rates, directional accuracies of sector returns, and generating excess returns in comparison with the SPY benchmark. Robust economic improvements are practically non-existent. Furthermore, in most cases, complex models do not provide more reliable and meaningful results. Future work could alleviate the limitations of low signal-to-noise ratio by employing regime-dependent modeling frameworks or other more flexible machine learning approaches.

## References

1. Chong, J., & Phillips, G. M. (2015). *Sector Rotation with Macroeconomic Factors.* Financial Analysts Journal, 71(4), 114–129.

2. Malkiel, B. G. (2003). *The Efficient Market Hypothesis and Its Critics.* Journal of Economic Perspectives, 17(1), 59–82.

3. López de Prado, M. (2020). *Machine Learning for Asset Managers.* Cambridge University Press.

4. Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V., & Stein, M. (2024). *Intelligent financial system: How AI is transforming finance.* BIS Working Papers No. 1194.

5. Rundo, F., Trenta, F., di Stallo, A. L., & Battiato, S. (2019). *Machine Learning for Quantitative Finance Applications: A Survey.* Applied Sciences, 9(24), 5574. `https://doi.org/10.3390/app9245574`

6. Nagel, S. (2021). *Machine Learning in Asset Pricing.* Princeton University Press.

7. Gu, S., Kelly, B., & Xiu, D. (2020). *Machine Learning for Asset Pricing.* The Review of Financial Studies, 33(5), 2223–2273.

8. Weigand, A. (2019). *Machine Learning in Empirical Asset Pricing.* Financial Markets and Portfolio Management, 33(1), 93–104.

9. Hoang, D., & Wiegratz, K. (2023). *Machine learning methods in finance: Recent applications and prospects.* European Financial Management, 29(5), 1657–1701. `https://doi.org/10.1111/eufm.12408`

# A Additional Figures

Table 4: Out-of-sample $R^2$ values by sector and model over the test period (2019–2025). Negative values indicate forecasting performance inferior to the historical mean benchmark.

| Sector | Linear Regression | Ridge Regression | Random Forest |
|---|---|---|---|
| GLD | -0.2015 | -0.1768 | -0.0783 |
| VNQ | -0.2374 | -0.1575 | -0.3149 |
| XLE | -0.3933 | -0.3885 | -0.0715 |
| XLF | -0.2989 | -0.2753 | -0.1417 |
| XLK | -0.3031 | -0.2954 | -0.1453 |

Table 5: Directional accuracy of monthly sector return predictions by model and sector over the out-of-sample test period (2019–2025).

| Sector | Linear Regression | Ridge Regression | Random Forest |
|---|---|---|---|
| GLD | 54.43% | 55.70% | 53.16% |
| VNQ | 62.03% | 62.03% | 41.77% |
| XLE | 51.90% | 50.63% | 45.57% |
| XLF | 55.70% | 55.70% | 50.63% |
| XLK | 55.70% | 56.96% | 46.84% |

Table 6: Performance comparison of rolling forecast strategies relative to the SPY benchmark.

| Training Scheme | Hit Rate (%) | Cumulative Excess Return (%) | Annualized Sharpe (Excess) |
|---|---|---|---|
| Expanding window | 21.31 | 5.13 | 0.09 |
| Rolling 60 months | 19.46 | -67.38 | -0.33 |

# B Code Repository

**GitHub Repository:** `https://github.com/LukasTonkovic/Sector-Rotation-Strategy`

# C Use of aids

ChatGPT and Grammarly were used for enhancing the text style and correcting grammar mistakes.