Przegląd Statystyczny. Statistical Review, 2025, vol. 72, 2, 54–83 https://doi.org/10.59139/ps.2025.02.3
Received: 4.04.2025, revised: 20.10.2025, accepted: 21.10.2025

Application of tree ensemble methods to the two-asset portfolio selection problem – a case study

Krzysztof Kaczmarek,[a] Aleksandra Rutkowska[b]

Abstract. The aim of the study was to construct a two-asset optimal investment portfolio using machine learning and macroeconomic data at monthly and quarterly intervals. The auxiliary objective was to identify which macroeconomic variables significantly impact the estimation of the S&P 500 stock index and the USD/GBP currency pair. The framework included two steps: firstly, time series forecasts were conducted using tree ensemble methods, namely the random forest and XGBoost, and secondly, the forecasts were used as expected values to construct the portfolios. We analyze the extent to which the structure of a portfolio based on the estimated data differs from the one built using historical data. The results of the research showed that it was possible to use the macroeconomic data to efficiently forecast the considered time series and construct an optimal investment portfolio.

Keywords: random forest, ensemble model, XGBoost, portfolio optimization
JEL: G11, G17

1. Introduction

Investing requires skillful risk management and returns optimization. In a dynamic market environment, where asset volatility can be both an opportunity and a threat, constructing a well-balanced investment portfolio is crucial. This study explores the strategy of building a two-component portfolio consisting of a currency and a stock index. We focus on a long-term investment and simple diversification, in contrast to numerous publications that emphasize highly active investing and trading. A two-asset portfolio is easier to monitor, rebalance, and understand over the long term. With fewer components, there is less complexity in tracking performance and making strategic decisions. While equity/bond strategies are the most analyzed (Pham, 2025), we explore a slightly riskier yet potentially more rewarding combination, namely an equity index and a currency. This portfolio structure offers several key benefits. First, combining a currency with an index can enhance diversification: when the currency appreciates, stock indexes may react differently, helping to mitigate the overall risk.

[a] Student at Informatics and Econometrics, Poznan University of Economics and Business, al. Niepodległości 10, 61–875 Poznań, Poland, e-mail: 84890@student.ue.poznan.pl, ORCID: https://orcid.org/0000-0003-1469-646X.
[b] Poznan University of Economics and Business, Institute of Informatics and Quantitative Economics, Department of Applied Mathematics, al. Niepodległości 10, 61–875 Poznań, Poland, e-mail: aleksandra.rutkowska@ue.poznan.pl, ORCID: https://orcid.org/0000-0002-2111-7764.

Second, monetary policy and global economic trends often influence the relationship between exchange rates and stock markets, allowing investors to leverage correlations (or the lack thereof) to improve portfolio efficiency. Finally, selecting the right combination of these two components can enhance the risk-return profile, making this approach attractive for individual and institutional investors.

Forecasting financial markets remains a central concern for both investors and researchers, despite the challenges posed by the Efficient Market Hypothesis (EMH), which suggests that financial markets are largely unpredictable, as asset prices already reflect all the available information. According to Țițan (2015), who provides a comprehensive review of the empirical studies testing the EMH, the dilemma of whether the market is efficient or not remains unresolved. Recent advances in artificial intelligence and machine learning have reignited interest in improving financial forecasting. Modern computational techniques offer new possibilities for extracting patterns from complex and high-dimensional financial data, potentially enhancing the predictive power of models even in markets traditionally considered efficient. A particularly interesting overview of the capabilities of different machine learning models for time series forecasting can be found in Ahmed et al. (2010), who compare the performance of several models applied to the M3 competition data. Additional comparative studies include the review by Tang et al. (2022) and Maung and Swanson (2025), which focus specifically on machine learning approaches for financial time series forecasting. From a wide range of methods, we selected ensemble tree-based machine learning methods for our study. Tree-based ensemble methods, such as random forest (RF) and Extreme Gradient Boosting (XGBoost) have gained popularity in forecasting due to their robustness and accuracy. Moreover, they capture complex, non-linear relationships in time series data, which makes them suitable for real-world forecasting applications (Wong et al., 2023). These methods demonstrated competitive performance (they are fast and more efficient although with a slightly higher number of forecast errors) for more complex recurrent neural networks and LSTMs (Nabipour et al., 2020a). Moreover, RFs can be trained with a relatively small amount of data (Roßbach, 2018).

This case study aims to shed light on several aspects related to the broader research objective. In particular, it explores: (1) which macroeconomic variables may have a notable impact on the estimation of a stock market index and a currency pair; (2) how the structure of a portfolio based on estimated data compares with the one constructed using complete information on current returns; and (3) what implications these differences may have for the portfolio's value, return, and risk profile. By examining these areas, the study seeks to offer a more detailed view of portfolio behavior under model-based estimation, contributing to the ongoing discussion in the area of investment strategy analysis.

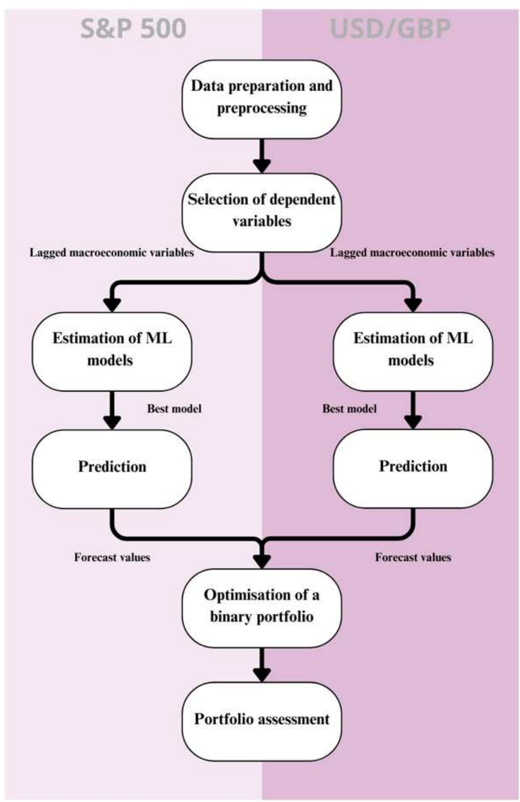
In the current literature, researchers have analyzed a variety of financial and macroeconomic indicators to model capital market behavior. The following variables have been examined in studies by Leippold et al. (2022): dividend yield, price-to-earnings (P/E) ratio, book-to-market ratio, net equity expansion, stock variance, term spread, and inflation. Additionally, money supply aggregates such as M2, trading volume, and monthly turnover were also considered. In the context of derivatives markets, futures contracts on silver, platinum, crude oil, and gold were analyzed by Shen et al. (2012) and Zhong and Enke (2019). They also used foreign exchange rates and stock indexes. Technical indicators such as the momentum indicator were studied by Choudhry and Garg (2008) and Reddy (2018), while %R Williams and the price volume trend were explored by Choudhry and Garg (2008). The stochastic oscillator was analyzed by Choudhry and Garg (2008), Patel et al. (2015), and Hegazy et al. (2014). Furthermore, technical indicators such as the relative strength index and moving average convergence divergence (Hegazy et al., 2014; Patel et al., 2015), moving averages (Hegazy et al., 2014; Patel et al., 2015; Zhong & Enke, 2019) were also investigated, as were indicators such as the Accumulation /Distribution (A/D) line and the Commodity Channel Index (CCI) (Patel et al., 2015); and finally, closing prices were described by Zhong & Enke (2019).

In the case of the currency pair, in articles where the authors applied ML models, the researchers used macroeconomic variables such as inflation or interest rates (Boyoukliev et al., 2022; Kaushik & Giri, 2020), money supply aggregates, government reserves, trade balance, and IIP (Kaushik & Giri, 2020). Boyoukliev et al. (2022) also used unemployment and, as Matuszewska-Janica & Witkowska (2008), interest rates on treasury bills. The reviewed studies also used precious metals and oil derivatives futures and stock indexes (Matuszewska-Janica & Witkowska, 2008). Other frequently used variables included technical analysis indicators, such as the Relative Strength Index and Rate of Change, which were examined by Abreu et al. (2018), Loh et al. (2022), Qi et al. (2020), and Baasher and Fakhr (2011). Moving averages (including WMA, EMA, SMA) were analyzed by Loh et al. (2022), Mabrouk et al. (2022) and Qi et al. (2020). Similarly, the moving average convergence/divergence (MACD) indicator was assessed by Baasher and Fakhr (2011), Mabrouk et al. (2022), and Loh et al. (2022). Other indicators such as the commodity channel index (Baasher & Fakhr, 2011) and the stochastic oscillator (Abreu et al., 2018; Baasher & Fakhr, 2011) were also considered, as were opening prices, highest and lowest prices during a session, closing prices, and trading volumes (Loh et al., 2022; Mabrouk et al., 2022).

2. Methodology

The proposed solution combines machine learning with standard investment portfolio estimation methods. The proposed framework is presented in Figure 1. In the first step, the data used in the study were downloaded and prepared. The second step consisted in a two-stage selection of the variables used in the models. The first stage involved the literature review depicted in the introduction, aiming to identify the variables that should, in theory, significantly affect the estimated series. The second stage involved the selection of these variables considering their correlation with the dependent variable. In the third step, the lagged macroeconomic variables were put in the selected machine-learning models and estimated. Then, those models that achieved the smallest root mean squared errors were selected. Using these models, one-period estimations of the levels of the S&P 500 index and the USD/GBP currency pair were made. The resulting time series were applied to construct minimum variance and maximum Sharpe ratio portfolios.

Figure 1. Research flow



Source: authors' work.

The final step was the portfolio assessment. First, the structure of the portfolios built using predictions from machine learning models was compared to that of the portfolios based on the realized data, with the current return rate known. Then, to assess the optimal portfolios, a visualization of the capital changes over time was created and the portfolio evaluation metrics were calculated.

The portfolio optimization issue involves determining the optimal weights for different assets. This problem was confronted by the theory presented by Markowitz (1952) – a groundbreaking idea in the area of portfolio construction that has become the foundation of modern approaches to this issue. The key idea is to condition the selection of the assets' weight on the expected return and risk, which means maximizing return for a given level of risk, or minimizing risk for a given return. According to Markowitz, the return rate on an investment represents the income earned from it, with investors knowing the probability distribution of returns. Investors' risk estimation is proportional to the expected return distribution. They make decisions based solely on two parameters of the probability distribution of returns. Investors prefer to minimize risk for a given rate of return, and for a given level of risk, they choose the investment offering the highest return. In Markowitz's theory, asset returns are identified with random variables. If the distribution of such a variable is known, then it is possible to determine the parameters of this distribution from it. Otherwise, investors are forced to estimate the expected value and the variance on the basis of historical data. The estimation of these parameters may cause problems related to the selection of an appropriate estimator, as well as to the quality of the obtained result. Moreover, the choice of the period from which the data used for the estimation is drawn can significantly affect the results. Considering the classical approach, the distributions should be normal or close to normal, but this may not be the case (Kaszuba, 2011).

Therefore, assets with the smallest possible variance, the highest return and the lowest correlation should ultimately be selected for the portfolio. For this reason, it is worth considering portfolio diversification, which enables the reduction of portfolio risk (Łuniewska, 2012).

2.1. Time series forecasting

RF, which is a machine learning algorithm based on decision trees, is a popular method used in relation to classification and regression problems (Król-Nowak & Kotabra, 2022). Multiple decision trees are involved when creating an RF model. This results in the reduction of the negative effects of overfitting some of the trees that make up the forest. Classification and regression in this algorithm then consist of comparing the obtained results by the individual independent trees in the forest.

When comparing the results from the trees, most of the same outcomes finally shape the classification or regression value of the forest (Basak et al., 2019; Géron, 2022).

XGBoost is also a decision tree-based algorithm capable of solving regression and classification problems. It was developed for better performance compared to other tree-based models (Nabipour et al., 2020b). The approach differs significantly from that used in the RF, where the result is supposed to be the best possible partitioning verified by impurity coefficients. In this case, however, the method uses gradient boosting, which involves combining several weak predictors or classifiers into one. This is done by sequentially training the successive models, with each successive model attempting to correct the errors of its predecessor (Basak et al., 2019; Géron, 2022). Table 1 presents a comparison of both methods. One undoubted advantage of these methods is that tree-based ensemble models, such as RF and XGBoost, are generally robust to multicollinearity in terms of predictive accuracy (Cabrera Malik, 2024; Gregorutti et al., 2017; Roy & Larocque, 2012). However, it should not be forgotten that in the case of multicollinearity, measures of trait importance obtained from these models may become unreliable when strong correlations occur between the predictors.

Table 1. Comparison of RF and XGBoost features

Feature	RF	XGBoost
Ensemble type	Bagging	Boosting
Tree building	Parallel	Sequential
Focus	Diversity of trees	Correcting errors of the previous trees
Accuracy	Generally good, but may be slightly lower than that of XGBoost	High accuracy, often of a state-of-the-art level
Training time	Generally faster	Can be slower, but in many cases more accurate
Overfitting	Relatively robust to overfitting	Can be prone to overfitting, but regularization helps overcome its effects
Interpretability	More interpretable	Less interpretable

Source: authors' work.

In the forecasting stage, the explanatory variables are the closing prices of the S&P 500 and the USD/GBP currency pair. We use quarterly and monthly data downloaded from stooq.pl. The independent variables are from Q1 1985 to Q2 2023 for quarterly frequency and from January 1990 to September 2023 for the monthly interval. The total length of the time series is 154 and 405 periods, respectively. 20% of the observations are a test sample, not used for model training. Most of the downloaded data were complete, but any missing values were filled by adjusting the next or previous value using the average growth rate of the nearby observations.

2.2. Portfolio construction

This paper compares two portfolios: one optimized for minimal volatility (further designated as Portfolio 1: Min variance) and the other for the Sharpe ratio (Portfolio 2: Max Sharpe ratio).

In the case of a two-asset portfolio, the risk is determined by the variance of the portfolio return and is given by the formula:

$$V_p = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2cov_{12}, \tag{1}$$

where:

w_i – share of the i-th asset in the portfolio,

σ_i – standard deviation of return on the i-th asset,

cov_{ij} – covariance of the i-th and j-th return rates.

The Sharpe ratio is one of the possible indicators for assessing investment performance in terms of risk and was first presented by Sharpe in his work on the asset pricing model (Sharpe, 1966). It considers both return and risk, combining these two key factors into a single measure, as shown in Equation (2) (Sharpe, 1998):

$$S = (R_p - R_f)/\sigma_p, \tag{2}$$

where

R_p – average return of the portfolio,

R_f – risk-free rate of return,

σ_p – standard deviation of the portfolio, so $\sqrt{V_p}$, according to (1).

The higher the value, the better rate of return on investment to the risk taken. The outcome indicates how many units of return can be obtained for each of the incurred unit of risk that the investor is potentially exposed to.

Besides the metrics used to create the portfolio, other ones were used to evaluate it, i.e. the average return, tracking error and maximum drawdown. The tracking error measures how much an investment differs from its benchmark. It is often used for ETFs, which aim to closely follow an index or make certain performance assumptions (Charteris & McCullough, 2020). The maximum drawdown is a popular risk measure commonly used in the financial sector. It measures how severe a single investment loss can be for an investor. This indicator has been defined as the maximum cumulative loss occurring, beginning at a price peak and ending at a bottom (Choi, 2021; Magdon-Ismail & Atiya, 2004).

3. Empirical results

3.1. Prediction results

To estimate the expected value of the return rate, two different ML models were estimated with the one with a lower root mean square error (RMSE) chosen for the next step. The libraries used in the forecasts, along with the models' hyperparameters are presented in Table A1 in the Appendix. The independent variables for the models were chosen based on their correlation with the dependent variable. From this set, we selected those that also aligned with the macroeconomic variables identified in the literature review. For the results to truly reflect the predictive capabilities, the independent variables lagged when building the models, so the data from the *i*-th period were used to estimate the dependent variable in the i+1 period. The results are presented in Table 2.

Table 2. RMSE of return predictions of the RF and XGBoost with the naive method as a benchmark

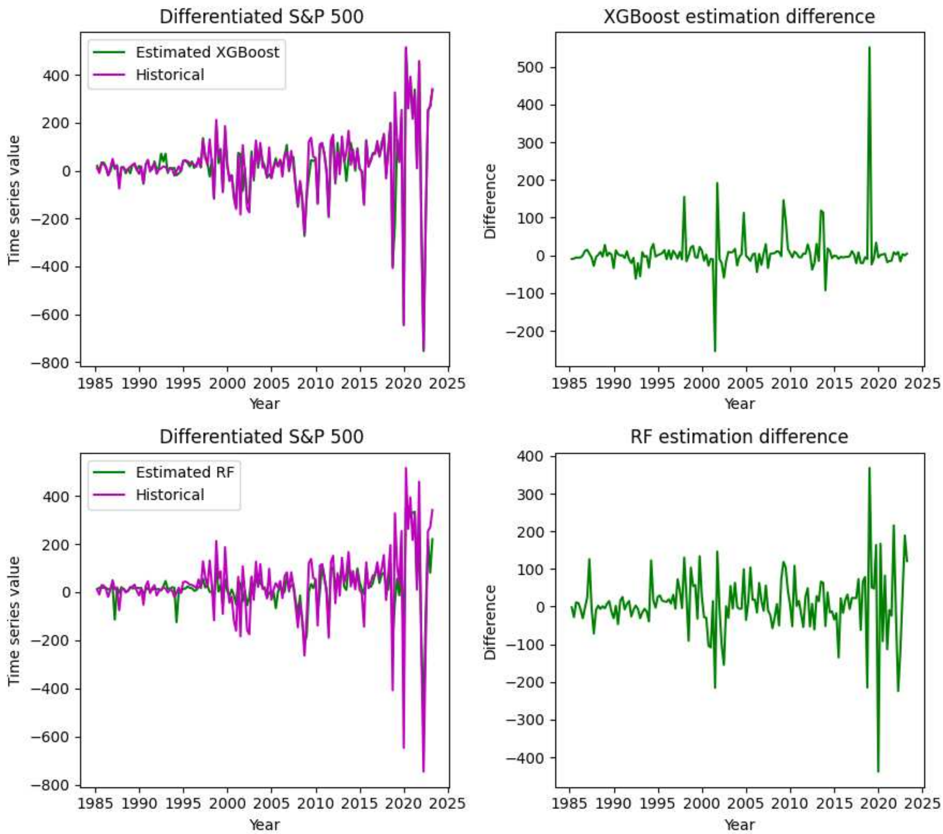
Assets	RF	XGBoost	Naive
Quarterly data			
S&P 500	103.07	130.49	204.51
USD/GBP	0.0377	0.0378	0.0429
Monthly data			
S&P 500	98.12	116.52	129.72
USD/GBP	0.0146	0.0170	0.0232

Note. The results from the test sample cover 20% of the observations.
Source: authors' work.

3.1.1. Quarterly interval

8 out of the 1,037 sampled variables were used to estimate the quarterly values of the S&P 500 index (see Figure A2 in the Appendix). Figure 2 shows the historical differentiated values of the index with the estimated values and the differences between the historical (real) and estimated differentiated values. Both models outperform the naive forecast: the XGBoost reduced the forecast error by 36.19%, and the RF by 49.60%.

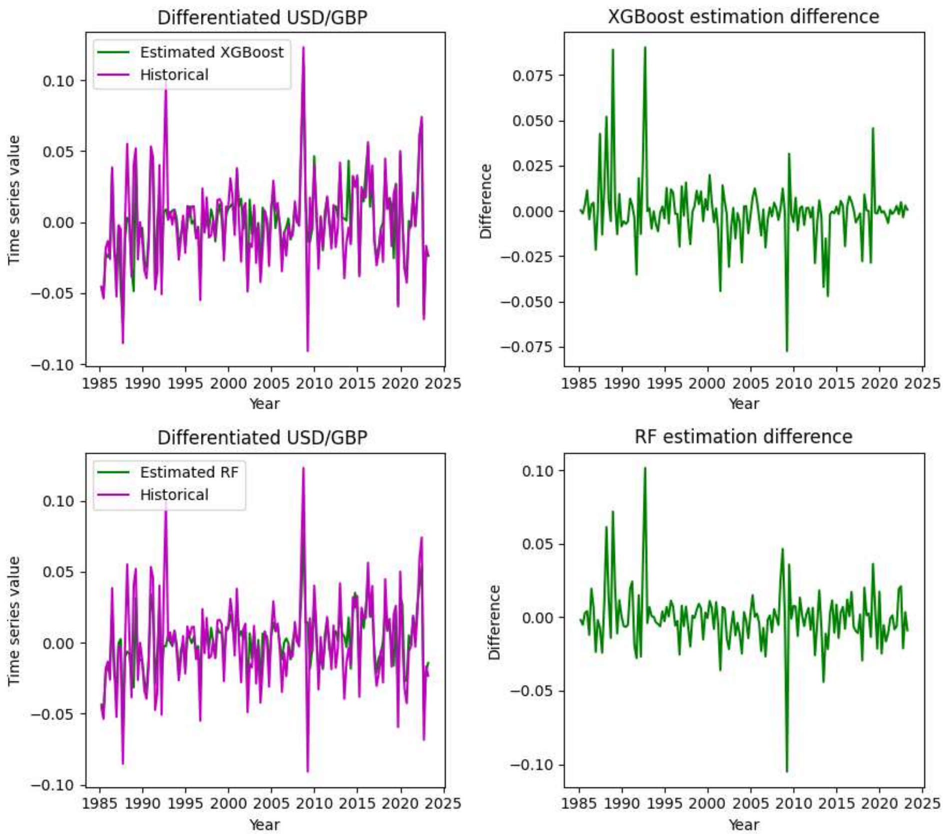
Figure 2. Estimated and historical quarterly differences of the S&P 500 index (left column) and estimation difference (right column)



Note. The results relate to the whole sample.
Source: authors' work.

The estimation of the quarterly differentiated closing prices of the USD/GBP currency pair used 8 variables from the 1,938 considered, of which both the US and the UK variables were among the sampled macroeconomic variables (for details, see Table A2 and Figure A2 in the Appendix) The results of the estimations, together with the differentiated historical values, are presented in Figure 3. Both models are characterized by a good fit to the data. Once again, both models corrected the forecast errors compared to the naive forecast by about 12% although there is no longer such a clear difference between the RF and the XGBoost.

Figure 3. Estimated and historical quarterly differences of the USD/GBP currency pair (left column) and estimation difference (right column)



Note. The results concern the whole sample.
Source: authors' work.

According to the results presented in Table 2 for the US stock exchange, the smallest error was observed in the RF regression model, and the same model proved to be the best for the currency pair.

The significance of the individual parameters was checked (see Figure A5 in the Appendix) for those models that achieved the best results (i.e. the RF for both estimated series). The most important variables for the stock market index are: the M3 aggregate money supply in the US, the US consumer price index (CPI), with the 2015 outcome as the baseline, and the net trade balance (the time series are presented in Figure A1 in the Appendix).

In the case of the model built to estimate the currency pair, the characteristics that have brought the greatest improvement in this case are: the UK public debt, the M1 aggregate money supply in the UK, the UK CPI and the US GDP (see Figure A2 in the Appendix). Despite the above distinction of features, it should be noted that the M1 monetary aggregate has the largest average contribution exceeding 20%, while the rest of the mentioned features exceed the threshold of 10% of the average total contribution to the model improvement.

Notably, three of the four most important variables used to estimate this currency pair are associated with the UK economy. This suggests that the US economy has a relatively minor impact on the UK.

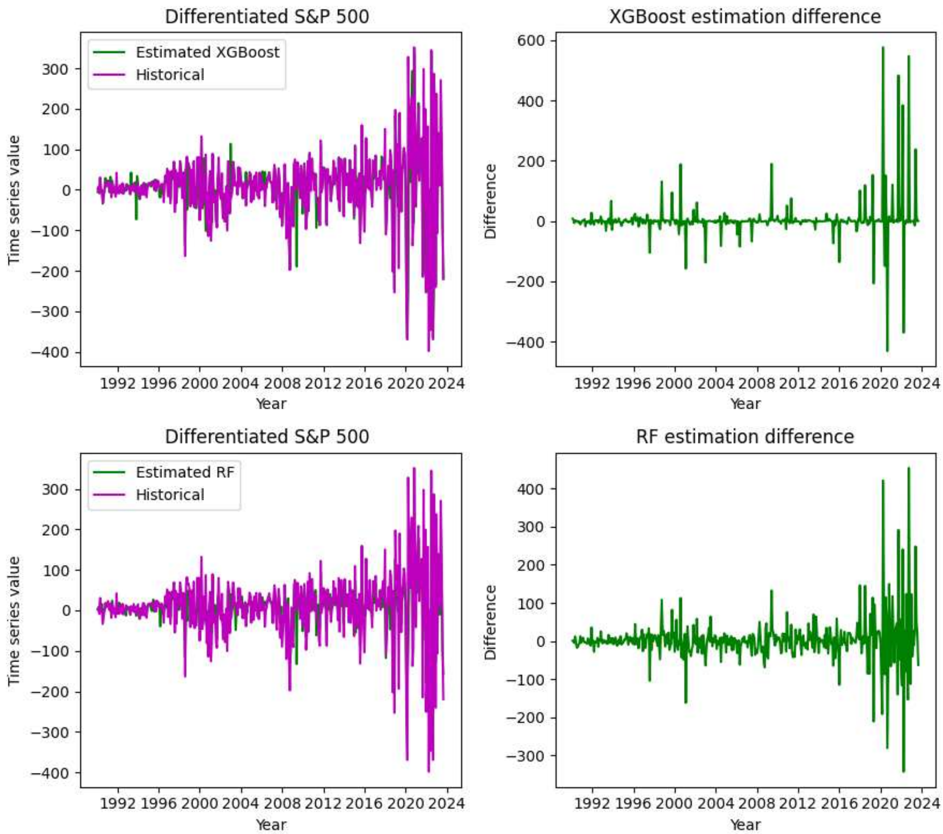
3.1.2. Monthly interval

In constructing models to estimate the differentiated stock market for the monthly interval, 6 out of the 780 independent variables were included (see Figure A3 in the Appendix). The significant difference in the number of the available data results from the fact that not all macroeconomic data are published monthly. The results are presented in Figure 4 and, once again, the models had a very good match. In the test sample, the XGBoost forecasts were better than the naive ones by 10% and RF by almost 25%.

From among a total of 1,379 variables considered, 11 independent variables were used in the construction of the models for the currency pair. Figure 5 illustrates the obtained results. The XGBoost improved the forecast results by 27% compared to the naive forecast, and the RF by 37%.

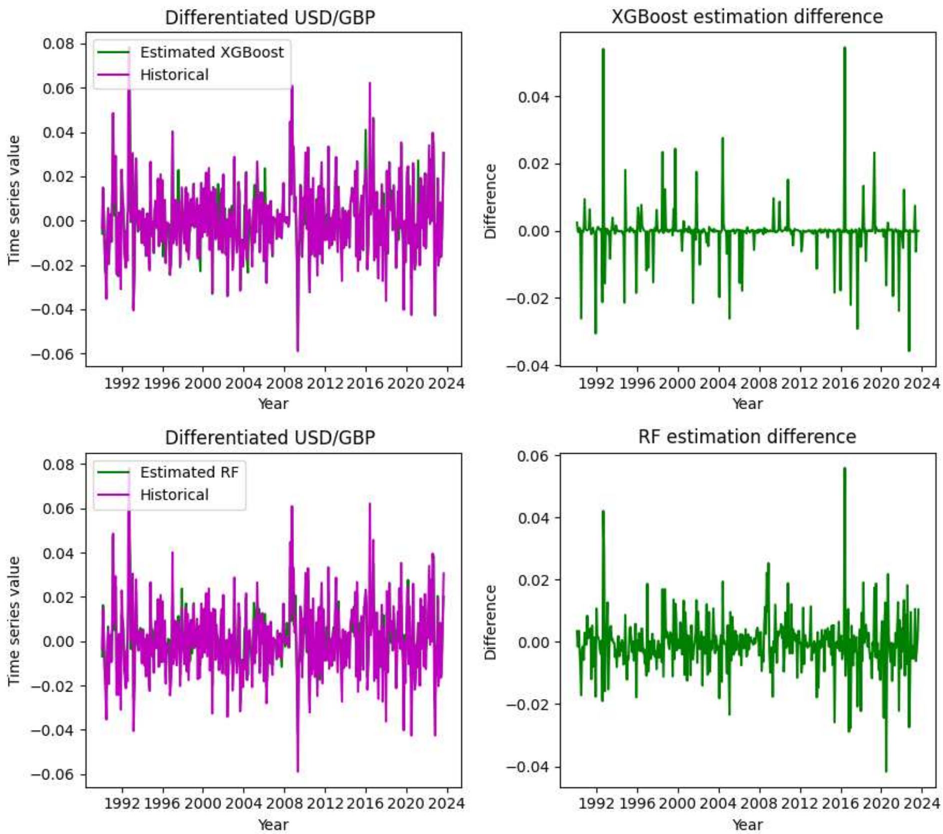
The RF models achieve the smallest RMSE for both assets, so the structure of the monthly portfolio is determined by their results.

Figure 4. Estimated and historical monthly differences of the S&P 500 index and estimation difference



Note. The results relate to the whole sample.
Source: authors' work.

Figure 5. Estimated and historical monthly differences of the USD/GBP currency pair and estimation difference



Note. The results relate to the whole sample.
Source: authors' work.

For the stock market index, again three of the five variables played a significant role in improving the quality of the model, and these were: the M3 aggregate money supply in the US, the US CPI (with the 2015 outcome as the baseline) and the net trade value. It is worth noting that both the CPI and the M3 aggregate played a crucial role in both time intervals. Although all the variables put in the model should, according to the literature review, play an important role in the closing prices, the relevance of the interest rates was below 10% in the conducted study.

The second model shows a lower diversification in the average impact of the variables, which may be indicative of their inferior selection for the model. The variables that brought the greatest improvement were the US and UK real effective exchange rates calculated as weighted average two-sided exchange rates adjusted for relative consumer prices with 2020 as the base period (see Figure A4 in the

Appendix). The first variable played twice as an important role in the model as the second variable, and in addition, the rest of the variables did not exceed 10% of the average contribution (for importance, see Figure A6 in the Appendix).

3.2. Investment portfolio results

To build the portfolio, the expected value of each instrument was based on the return rate from the one-period prediction of the selected model. Asset shares were calculated using an algorithm that either minimized variance (1) or maximized the Sharpe ratio (2). A minimum asset weighting of 10% was set to maintain portfolio stability.

These initial portfolio weights were considered 'raw' since constantly adjusting them for minor changes would have been impractical. To account for transaction costs and frequent fluctuations, weight adjustments were only made when an asset's share changed by more than 2%. Additionally, portfolios calculated using ex-post data were used as benchmarks (here: xxx_hist). The deviation from these benchmarks was measured by calculating the root of the sum of the squared differences between the estimated and benchmark weights. We calculated the portfolio for the following cases:

A – raw and recalculated weights (here: xxx_new) of the minimum variance portfolio,

B – raw and recalculated weights (here: xxx_new) of the maximum Sharpe ratio portfolio,

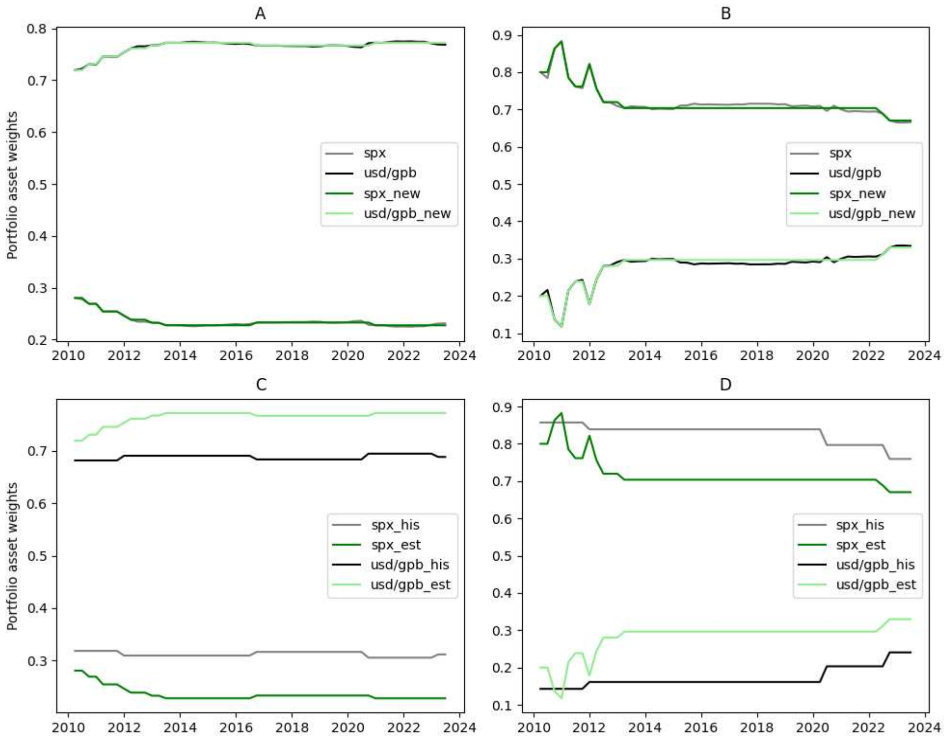
C – recalculated weights (here: xxx_est) and weights based on the realized values (here: xxx_hist) of the minimum variance,

D – recalculated weights (here: xxx_est) and weights based on the realized values (here: xxx_hist) of the maximum Sharpe ratio portfolio.

The graphs in the first row of Figure 6 illustrate the structure of the portfolios determined by variance minimization (A) and Sharpe ratio maximization (B) for the quarterly interval. Portfolio (A) maintains a quite stable structure remaining in a similar ratio of 2:8 for the currency pair. A similar stable structure was obtained for (C), where the ratio was 3:7.

For portfolio (B), the structure is slightly more unstable at the beginning, and stabilizes at a ratio of 3:7, but the asset weights reversed. The weights in (D) exhibit a similar behavior. The obtained results indicate a higher volatility of the stock index.

Figure 6. Structure of optimal portfolios for the quarterly interval



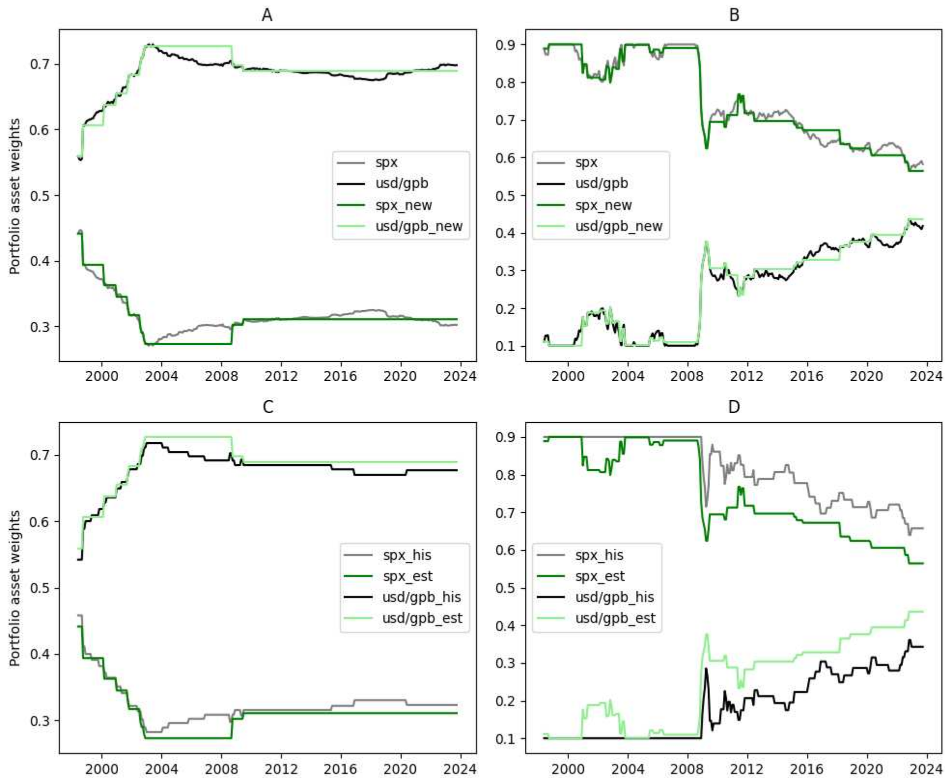
Note. Minimum variance portfolio – left panels, maximum Sharpe ratio portfolio – right panels, raw weights – upper panels, recalculated weights – lower panels. The results relate to the whole sample.
Source: authors' work.

Furthermore, the graphs in the second row of Figure 6 show the adjusted portfolio structure based on the estimated closing prices, overlaid with the structure based on historical data. Due to the criteria applied, the weights of the optimal portfolios largely overlap with the weights resulting from the historical values. The root sum of the squares of the differences in the weights calculated based on price prediction and historical values is 8% and 11%, respectively. These are low values, which indicate correct estimation results and fairly small deviations from the benchmark.

Figure 7 illustrates the structure of the portfolios established for monthly intervals. As the interval is shortened, a significantly higher volatility of the structure is observable in both portfolios compared to the quarterly portfolios.

Minimum variance portfolio A was at first characterized by an almost equal distribution of assets which tended to assign more weight to the currency pair and maintain that level until the end of the considered period. In this case, the portfolio is slightly less diversified and the weights move in ranges close to the 7:3 ratio for the currency pair.

Figure 7. Structure of optimal portfolios for the monthly interval



Note. Minimum variance portfolio – left panels, maximum Sharpe ratio portfolio – right panels, raw weights – upper panels, recalculated weights – lower panels. The results relate to the whole sample.
Source: authors' work.

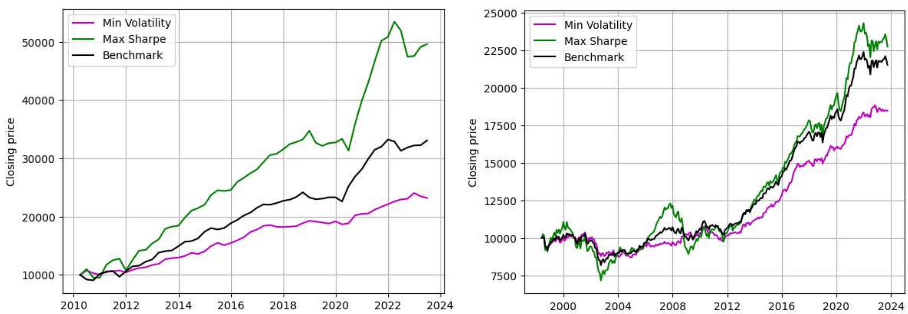
For portfolio (B), and similarly for (D), a higher stock market index weight was observed again, which held the maximum possible weight for almost a quarter of the examined period. In the monthly time interval, there was much more variance in the portfolio structures than in the longer period. This underestimation is noticeable in the minimum variance portfolio and it occurs throughout the whole studied period. In the case of portfolio B, however, the significant changes in the structure are more likely to be temporary and occur towards the end of the period.

In this case, the root sum of the squares of the differences between the weights calculated based on price prediction and historical prices was 2% and 8%. The minimum variance portfolio shows a significant improvement of this parameter.

In contrast, the portfolio maximizing the Sharpe ratio, despite the more abrupt changes in the weights, slightly adjusts the level of variance compared to the previous interval.

The resulting portfolios were also compared to a benchmark, which was assumed to be a portfolio with equal instrument weights throughout the period under study. Therefore, a kind of simulation was performed assuming the investment of 10,000 monetary units in both portfolios and the benchmark. The results are presented in Figure 8.

Figure 8. Capital value over time



Note. Quarterly – left panel, monthly – right panel. The results concern the whole sample.
Source: authors' work.

The portfolio results are shown in Table 3. The smallest portfolio variance occurred in Portfolio 1 and the largest in Portfolio 2. Portfolio 3 represents a kind of arithmetic average in this respect, which is related to the structure of this portfolio. Surprisingly, the largest Sharpe ratio was achieved by Portfolio 1, which should theoretically be the third-best portfolio in this metric. Given that this metric combines a portfolio's return with its risk, there can be some kind of anomaly or discrepancy between the expected and received result when making comparisons and in terms of extreme values. In this case, this is probably due to the three times lower variance in relation to Portfolio 2, which was in the denominator when calculating this ratio.

The highest overall rate of return of approximately 500% was achieved by Portfolio 2, thus achieving double the performance of Portfolio 1. The Sharpe ratio, maximizing the portfolio also recorded the highest average return of 3.27% from period to period. This seems a logical consequence of achieving such a high overall return. This portfolio, however, did not perform as well as the others in terms of portfolio variance. This is where the variance minimizing portfolio performed best, while the benchmark underperformed by just 7 percentage points.

Table 3. Portfolio assessment metrics

Metrics	Portfolio 1 Min variance	Portfolio 2 Max Sharpe ratio	Portfolio 3 Benchmark (50:50)
quarterly			
Portfolio variance	9.43%	27.43%	16.17%
Sharpe ratio	59.62	49.73	56.77
Rate of return	231.96%	496.18%	330.89%
Average rate of return (quarterly)	1.64%	3.27%	2.37%
Maximum drawdown	6.63%	15.87%	9.47%
Tracking error	0.36%	0.37%	–
monthly			
Portfolio variance	1.77%	2.95%	2.36%
Sharpe ratio	14.87	12.12	15.91
Rate of return	184.80%	232.42%	215.01%
Average rate of return (monthly)	0.21%	0.30%	0.26%
Maximum drawdown	16.13%	35.25%	20.61%
Tracking error	0.04%	0.6%	–

Note. The results relate to the whole sample.

Source: authors' work.

The smallest maximum drawdown was achieved by Portfolio 1, but it should be noted that there were no large divergences in relation to Portfolio 3. The variance minimizing portfolio (Portfolio 1) is the closest to the benchmark, where on average its one-period returns deviated from the benchmark rates by 0.36%. In contrast, Portfolio 2 outperforms both the benchmark and the variance minimizing portfolio.

Comparing the graphs in Figure 8 and focusing on the period from 2010, it becomes apparent that there was a definite smoothing of any price peaks and lows by the quarterly portfolios and a greater divergence between the portfolios compared to the shorter interval.

In contrast to the previous interval, in this case, the portfolio maximizing the Sharpe ratio significantly outperformed the other two portfolios only near the years 2000 and 2008. The higher capital value in these periods was due to the large price fluctuations of both assets. Around 2002, these were caused by the 9/11 attacks and the bursting of the speculative internet bubble (involving the overvaluation of IT companies). In the second period mentioned, stock market falls and the weakening of the dollar after the subprime crisis were the reasons. During the recession following these events, only two periods occurred where Portfolio 2 was outperformed by both the minimizing variance portfolio and the benchmark.

In the monthly time interval, there were no major differences in terms of which portfolio performed best on a given metric compared to the quarterly interval. On the other hand, despite the shorter intervals, portfolio variance improved significantly by up to nine times the variance of the quarterly portfolios. Due to the shortening of the time interval, it is natural for the one-period portfolio return to decrease as well. The largest change between the considered intervals was in the metric of the average rate of return, where a decrease of up to ten times was observed for Portfolio 2. Despite the lower volatility, the tracking error of the optimized portfolios increased. In addition, Figure 8 indicates much smaller divergences between portfolios in the quarterly interval than in the monthly one.

The above graphs of capital change over time and the calculated evaluation metrics show that the portfolio maximizing the Sharpe ratio offers the highest return but at the price of the highest volatility. The minimizing variance portfolio instead offers more stable but slower capital growth. However, the benchmark set places itself between the two portfolios, combining and averaging the advantages and disadvantages of both investment portfolios. Ultimately, the benchmark in the monthly interval considerably outperformed the minimizing variance portfolio.

4. Conclusions

In this study, we proposed a framework to build a two-asset portfolio. For this purpose, we combined machine learning algorithms based on trees (RF and XGBoost) with methods optimizing portfolio performance. The framework was illustrated in the periods from Q1 1985 to Q2 2023 (quarterly intervals) and from January 1990 to September 2023 (monthly intervals), showing promising results.

Based on our case study, for the selected assets and time frame, the analysis of the research questions led to several conclusions.

The macroeconomic variables used in the modeling allowed the models to estimate the time series efficiently. The assessment of the significance of these variables confirms that some of the variables derived from the economic theory and the literature review have a particularly significant influence on the estimation results, especially money supply aggregates and inflation rates.

The structure of the constructed portfolio based on the estimated data does not differ significantly from the structure of the portfolio based on the realized return rates. These deviations were measured by the mean square error between the received and historical structure. The obtained result involved deviations ranging from 2 to 12 percentage points.

The return of the resulting portfolios, which is also the most important portfolio evaluation metric, ranged from around 200% to just over 500% depending on the optimized values in the portfolio. The portfolio maximizing the Sharpe ratio obtained, on average, a better return than the portfolio minimizing variances by about one-third of the portfolio results. Of course, this was associated with a significantly higher risk, as measured by the portfolio variance, which was on average about three times higher. It came as a surprise that in the conducted study, the portfolio maximizing the Sharpe ratio did not score the best values in this metric. This is related to the higher contribution of capital to a riskier asset which outweighed the achieved high return and resulted in lower scores in this metric. It is also worth mentioning that, despite the highest return, this portfolio did not always produce the best results. Indeed, in the monthly interval, there were two periods in which the value of the evenly spread assets portfolio was the highest.

Finally, summarizing the obtained results, this case study shows that it is possible to efficiently construct a two-component portfolio using macroeconomic data and machine learning methods. The research problem explored in this study should be further developed in future work by expanding the scope to include a greater number of markets and investment portfolio components, as well as by employing a broader range of ML methods.

References

Abreu, G., Neves, R., & Horta, N. (2018). *Currency exchange prediction using machine learning, genetic algorithms and technical analysis*. https://doi.org/10.48550/arXiv.1805.11232.

Ahmed, N. K., Atiya, A. F., El Gayar, N., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, *29*(5–6), 594–621. https://doi.org/10.1080/07474938.2010.481556.

Baasher, A. A., & Fakhr, M. W. (2011). FOREX Trend Classification using Machine Learning Techniques. In A. Zaharim, K. Sopian, N. Mastorakis & V. Mladenov (Eds.), *ACS'11: Proceedings of the 11th WSEAS International Conference on Applied Computer Science* (pp. 41–47). World Scientific and Engineering Academy and Society. https://www.wseas.us/e-library/conferences/2011/Penang/ACRE/ACRE-05.pdf.

Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, *47*, 552–567. https://doi.org/10.1016/j.najef.2018.06.013.

Boyoukliev, I. V., Kulina, H. N., & Gocheva-Ilieva, S. G. (2022). Modelling and Forecasting of EUR/USD Exchange Rate Using Ensemble Learning Approach. *Cybernetics and Information Technologies*, *22*(4), 142–151. https://doi.org/10.2478/cait-2022-0044.

Cabrera Malik, S. (2024). *Is xgboost immune to multicolinearity?*. https://medium.com/@sebastian.cabrera-malik/is-xgboost-inmune-to-multicolinearity-4dd9978605b7.

Charteris, A., & McCullough, K. (2020). Tracking error vs tracking difference: Does it matter? *Investment Analysts Journal*, *49*(3), 269–287. https://doi.org/10.1080/10293523.2020.1806480.

Choi, J. (2021). Maximum Drawdown, Recovery, and Momentum. *Journal of Risk and Financial Management*, *14*(11), 1–25. https://doi.org/10.3390/jrfm14110542.

Choudhry, R., & Garg, K. (2008). A Hybrid Machine Learning System for Stock Market Forecasting. *Proceedings of World Academy of Science, Engineering and Technology*, *29*, 315–318. https://par.cse.nsysu.edu.tw/resource/paper/2008/080820/A%20Hybrid%20Machine%20Learning%20System%20for%20Stock%20Market%20Forecasting.pdf.

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*. O'Reilly Media.

Gregorutti, B., Michel, B., & Saint-Pierre, P. (2017). Correlation and variable importance in random forests. *Statistics and Computing*, *27*(3), 659–678. https://doi.org/10.1007/s11222-016-9646-1.

Hegazy, O., Soliman, O. S., & Salam, M. A. (2014). *A Machine Learning Model for Stock Market Prediction*. https://doi.org/10.48550/arXiv.1402.7351.

Kaszuba, B. (2011). Praktyczne problemy w teorii portfela. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (174), 381–393.

Kaushik, M., & Giri, A. K. (2020). *Forecasting Foreign Exchange Rate: A Multivariate Comparative Analysis between Traditional Econometric, Contemporary Machine Learning & Deep Learning Techniques*. https://doi.org/10.48550/arXiv.2002.10247.

Król-Nowak, A., & Kotarba, K. (2022). *Podstawy uczenia maszynowego*. Wydawnictwa Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie. https://winntbg.bg.agh.edu.pl/skrypty4/0599/podstawy_uczenia.pdf.

Leippold, M., Wang, Q., & Zhou, W. (2022). Machine learning in the Chinese stock market. *Journal of Financial Economics*, *145*(2A), 64–82. https://doi.org/10.1016/j.jfineco.2021.08.017.

Liu, Y. (2023). Government debt and risk premia. *Journal of Monetary Economics*, *136*, 18–34. https://doi.org/10.1016/j.jmoneco.2023.01.009.

Loh, L. K. Y., Kueh, H. K., Parikh, N. J., Chan, H., Ho, N. J. H., & Chua, M. C. H. (2022). An Ensembling Architecture Incorporating Machine Learning Models and Genetic Algorithm Optimization for Forex Trading. *FinTech*, *1*(2), 100–124. https://doi.org/10.3390/fintech1020008.

Łuniewska, M. (2012). *Ekonometria finansowa. Analiza rynku kapitałowego*. Wydawnictwo Naukowe PWN.

Mabrouk, N., Chihab, M., & Chihab, Y. (2022). *A Trading Strategy in the Forex Market based on Linear and Non-linear Machine Learning Algorithms*. https://www.scitepress.org/Papers/2021/107288/107288.pdf.

Magdon-Ismail, M., & Atiya, A. F. (2004). Maximum Drawdown. *Risk Magazine*, *17*(10), 99–102.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91. https://doi.org/10.2307/2975974.

Matuszewska-Janica, A., & Witkowska, D. (2008). Modelowanie kursu euro/dolar: dynamiczne modele ekonometryczne i sztuczne sieci neuronowe. *Zeszyty Naukowe SGGW – Ekonomika i Organizacja Gospodarki Żywnościowej*, (69), 55–75. https://doi.org/10.22630/EIOGZ.2008.69.92.

Maung, K., & Swanson, N. R. (2025). A survey of models and methods used for forecasting when investing in financial markets. *International Journal of Forecasting*, *41*(4), 1355–1382. https://doi.org/10.1016/j.ijforecast.2025.03.002.

Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., Salwana, E., & Shahab, S. (2020a). Deep Learning for Stock Market Prediction. *Entropy*, *22*(8), 1–23. https://doi.org/10.3390/e22080840.

Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., & Mosavi, A. (2020b). Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *IEEE Access*, *8*, 150199–150212. https://doi.org/10.1109/ACCESS.2020.3015966.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, *42*(4), 2162–2172. https://doi.org/10.1016/j.eswa.2014.10.031.

Pham, N., Cui, B., & Ruthbah, U. (2025). *The Performance of the 60/40 Portfolio: A Historical Perspective*. https://rpc.cfainstitute.org/research/reports/2025/performance-of-the-60-40-portfolio.

Qi, L., Khushi, M., & Poon, J. (2020). Event-Driven LSTM for Forex Price Prediction. *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, Australia, 1–6. https://doi.org/10.1109/CSDE50874.2020.9411540.

Reddy, V. K. S. (2018). Stock Market Prediction Using Machine Learning. *International Research Journal of Engineering and Technology*, *5*(10), 1032–1035. https://www.irjet.net/archives/V5/i10/IRJET-V5I10193.pdf.

Roßbach, P. (2018). *Neural Networks vs. Random Forests – Does it always have to be Deep Learning*. https://blog.frankfurt-school.de/wp-content/uploads/2018/10/Neural-Networks-vs-Random-Forests.pdf.

Roy, M.-H., & Larocque, D. (2012). Robustness of random forests for regression. *Journal of Nonparametric Statistics*, *24*(4), 993–1006. https://doi.org/10.1080/10485252.2012.715161.

Sharpe, W. F. (1966). Mutual Fund Performance. *The Journal of Business*, *39*(1), 119–138.

Sharpe, W. F. (1998). The Sharpe ratio. In P. L. Bernstein & F. J. Fabozzi (Eds.), *Streetwise. The Best of The Journal of Portfolio Management* (pp. 169–178). Princeton University Press. https://doi.org/10.2307/j.ctv1mjqtwg.24.

Shen, S., Jiang, H., & Zhang, T. (2012). *Stock Market Forecasting Using Machine Learning Algorithms*. Stanford University.

Tang, Y., Song, Z., Zhu, Y., Yuan, H., Hou, M., Ji, J., Tnag, C., & Li, J. (2022). A survey on machine learning models for financial time series forecasting. *Neurocomputing*, *512*, 363–380. https://doi.org/10.1016/j.neucom.2022.09.003.

Țițan, A. G. (2015). The Efficient Market Hypothesis: review of specialized literature and empirical research. *Procedia Economics and Finance*, *32*, 442–449. https://doi.org/10.1016/S2212-5671(15)01416-1.

Wong, A., Whang, S., Sagre, E., Sachin, N., Dutra, G., Lim, Y. W., Hains, G, Khmelevsky, Y., & Zhang, F. C. (2023, December). *Short-term stock price forecasting using exogenous variables and machine learning algorithms* [paper presentation]. 3rd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), Denpasar, Indonesia. https://doi.org/10.1109/ICICyTA60173.2023.10428814.

Zhong, X., & Enke, D. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation*, *5*(1), 1–20. https://doi.org/10.1186/s40854-019-0138-0.

Appendix

Table A1. Models' hyperparameters

	Model	Monthly interval		Quarterly interval	
		Parameter	Value	Parameter	Value
SPX	XGBoost	Objective	Reg: squarederror	Objective	Reg: squarederror
		max_depth learning_rate n_estimators num_boost_round	10 0.66 100 10	max_depth learning_rate n_estimators	4 1 100
	RF	n_estimators random_state	100 24	max_depth min_samples_leaf min_samples_split n_estimators random_state	10 1 8 300 24
USD/GBP	XGBoost	Objective	Reg:squarederror	Objective	Reg:squarederror
		max_depth learning_rate n_estimators	10 0.9 100	max_depth learning_rate n_estimators	4 0.6 100
	RF	n_estimators random_state	14 25	n_estimators random_state	135 24

Note. The algorithm implementations used in this study are provided by the following Python libraries: XGBoost, sklearn.ensemble, sklearn.metrics, sklearn.preprocessing, and sklearn.model_selection.

Source: authors' work.

Table A2. List of independent variables

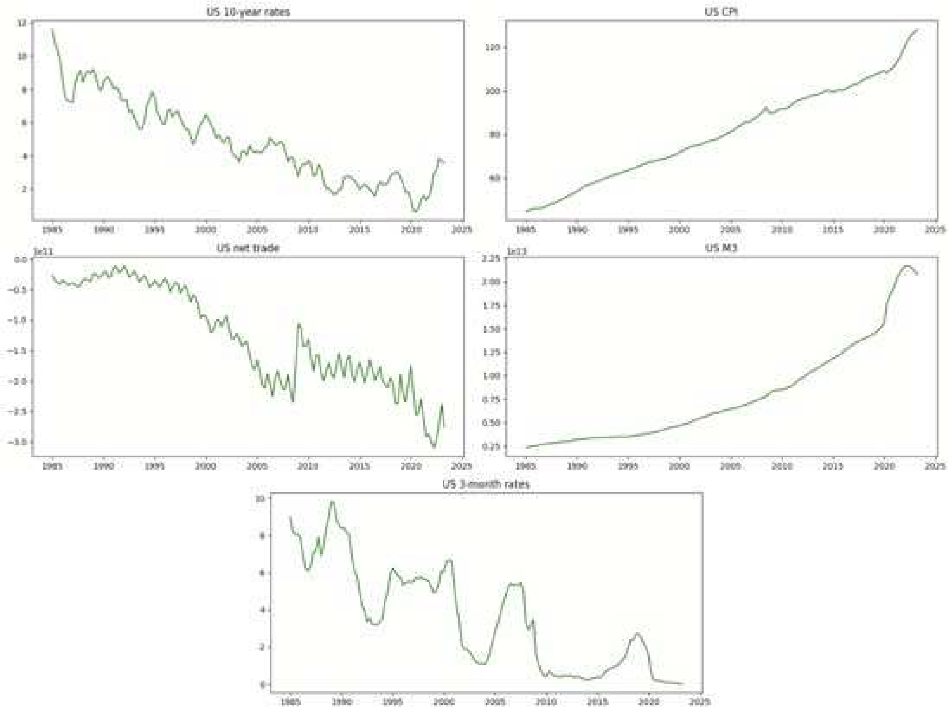
Short name	Description	Source study
	Quarterly S&P 500	
US 10-year rates	10-year government bond yield for the United States	(Boyoukliev et al., 2022), (Kaushik & Giri, 2020), (Matuszewska-Janica & Witkowska, 2008)
US CPI	Consumer price index for the United States, with the 2015 outcome as the baseline	(Leippold et al., 2022), (Boyoukliev et al., 2022), (Kaushik & Giri, 2020)
US net trade	Commodities Trade Balance for United States	(Kaushik & Giri, 2020)
US M3	M3 aggregate money supply in the United States	(Leippold et al., 2022), (Kaushik & Giri, 2020)
US 3-month rates	3-month interest rate for United States	(Boyoukliev et al., 2022), (Kaushik & Giri, 2020), (Matuszewska-Janica & Witkowska, 2008)
	Quarterly USD/GBP	
US labor inactivity rate	Labor inactivity rate for persons aged 15 and older in the United States	(Boyoukliev et al., 2022)
US import	Real Imports of Goods and Services for the United States	(Kaushik & Giri, 2020)
US export	Real Exports of Goods and Services for the United States	(Kaushik & Giri, 2020)
US GBP	Real Gross Domestic Product for the United States	(Liu, 2023)
WB labor inactivity rate	Labor inactivity rate for persons aged between 25 and 54 in the United Kingdom	(Boyoukliev et al., 2022)
WB CPI	Consumer price index for the United Kingdom, with the 2015 outcome as the baseline	(Leippold et al., 2022), (Boyoukliev et al., 2022), (Kaushik & Giri, 2020)
WB M1	M1 aggregate money supply in the United Kingdom	(Leippold et al., 2022), (Kaushik & Giri, 2020)
WB public debt	Total credit to general government for the United Kingdom (credit covers loans and debt securities)	(Liu, 2023)
	Monthly S&P 500	
US 10-year rates	10-year government bond yield for the United States	(Boyoukliev et al., 2022), (Kaushik & Giri, 2020), (Matuszewska-Janica & Witkowska, 2008)
US CPI	Consumer price index for the United States, with the 2015 outcome as the baseline	(Leippold et al., 2022), (Boyoukliev et al., 2022), (Kaushik & Giri, 2020)
US net trade	Commodities Trade Balance for the United States	(Kaushik & Giri, 2020)

Table A2. List of independent variables (cont.)

Short name	Description	Source study
US M3	M3 aggregate money supply in the United States	(Leippold et al., 2022), (Kaushik & Giri, 2020)
US annual rates	Annual interest rate for the United States	(Boyoukliev et al., 2022), (Kaushik & Giri, 2020), (Matuszewska-Janica & Witkowska, 2008)
Monthly USD/GBP		
US effective exchange rate	Real effective exchange rates are calculated as weighted averages of bilateral exchange rates adjusted by the relative consumer prices	(Shen et al., 2012), (Zhong & Enke, 2019)
US labor participation rate	Labor force participation rate for persons aged 15 and older in the United States	(Boyoukliev et al., 2022)
US reserves, excluding gold	Total reserves, excluding gold for the United States	(Kaushik & Giri, 2020)
US retail sales	Total retail trade value in the United States	(Kaushik & Giri, 2020)
US CPI	Consumer price index for the United States, with the 2015 outcome as the baseline	(Leippold et al., 2022), (Boyoukliev et al., 2022), (Kaushik & Giri, 2020)
US M3	M3 aggregate money supply in the United States	(Leippold et al., 2022), (Kaushik & Giri, 2020)
WB M1	M1 aggregate money supply in the United Kingdom	(Leippold et al., 2022), (Kaushik & Giri, 2020)
WB reserves, excluding gold	Total reserves, excluding gold for the United Kingdom	(Kaushik & Giri, 2020)
WB effective exchange rate	Real effective exchange rates are calculated as weighted averages of bilateral exchange rates adjusted by the relative consumer prices	(Shen et al., 2012), (Zhong & Enke, 2019)
WB CPI	Consumer price index for the United Kingdom, with the 2015 outcome as the baseline	(Leippold et al., 2022), (Boyoukliev et al., 2022), (Kaushik & Giri, 2020)
WB net trade	Commodities trade balance for the United Kingdom	(Kaushik & Giri, 2020)

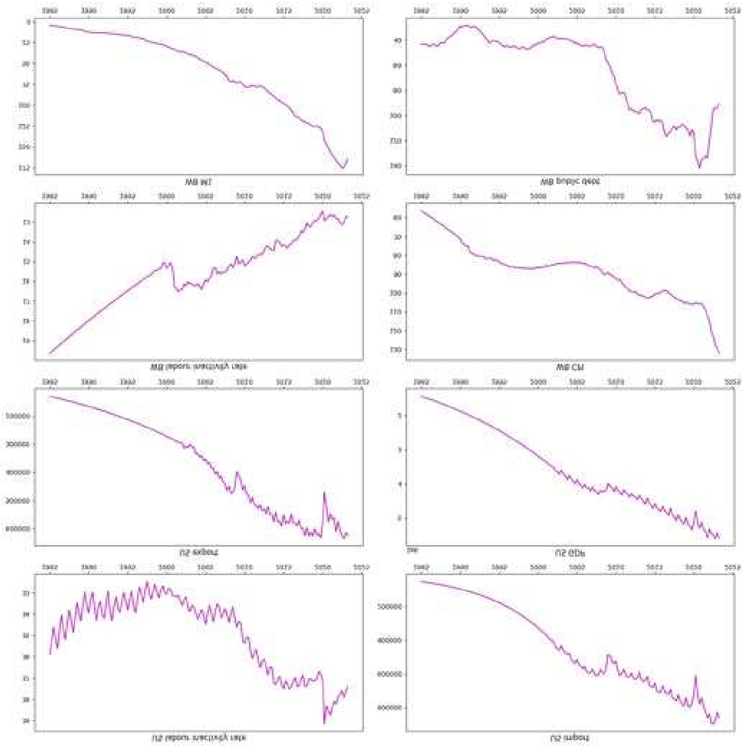
Source: authors' work.

Figure A1. Independent variables used in quarterly S&P 500 estimation



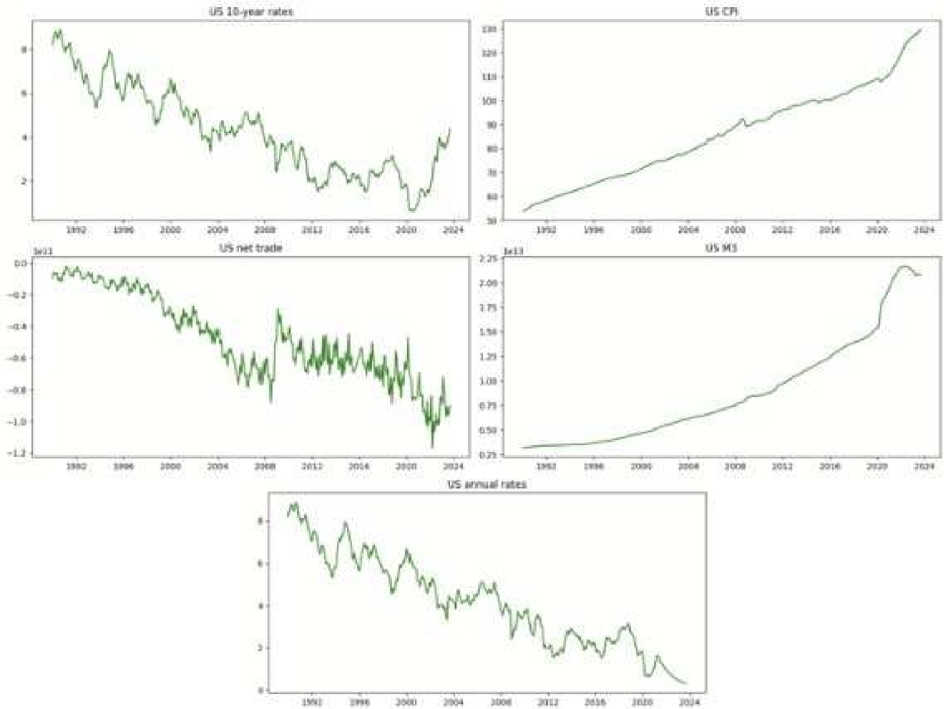
Source: authors' work.

Figure A2. Independent variables used in quarterly USD/GBP estimation



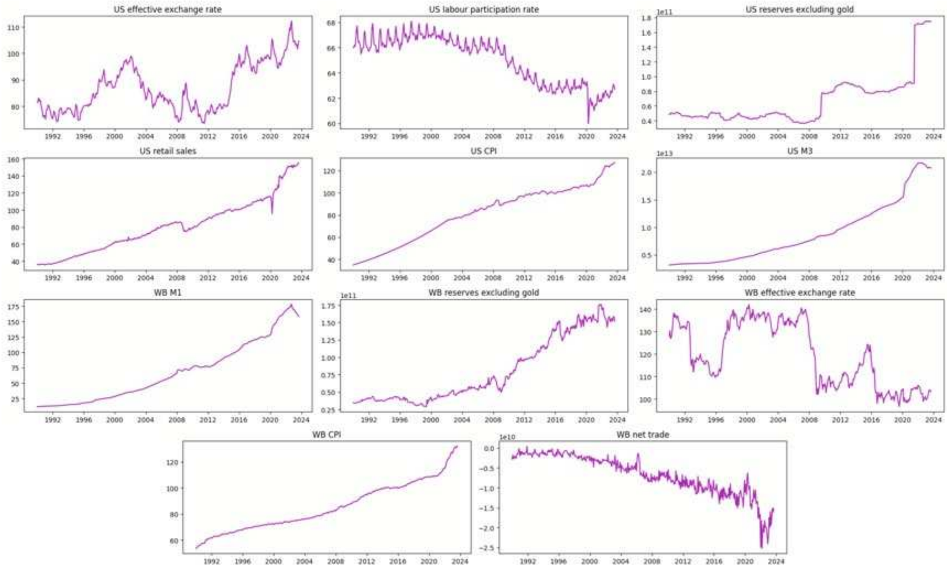
Source: authors' work.

Figure A3. Independent variables used in monthly S&P 500 estimation



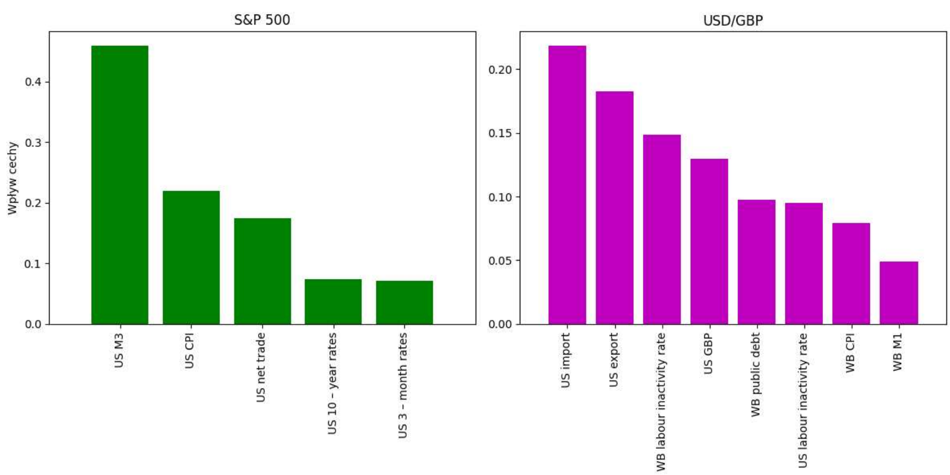
Source: authors' work.

Figure A4. Independent variables used in monthly USD/GBP estimation



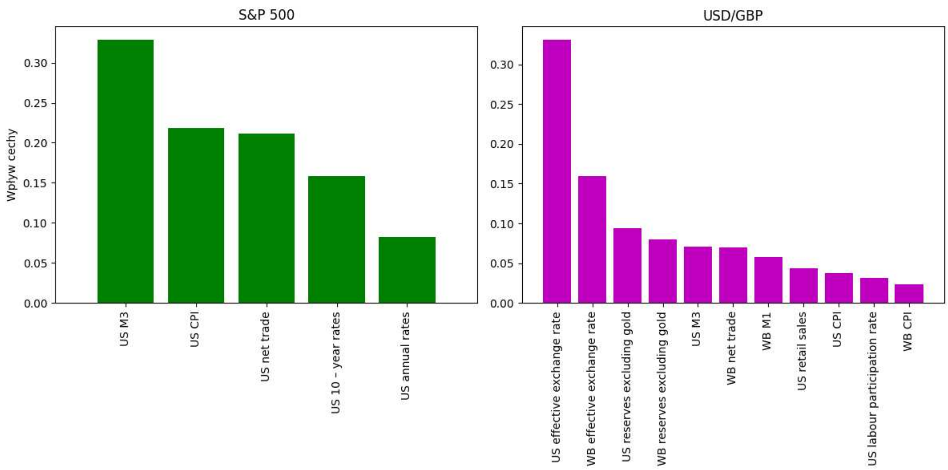
Source: authors' work.

Figure A5. Feature importance of variables used in the quarterly estimation



Source: authors' work.

Figure A6. Feature importance of variables used in the monthly estimation



Source: authors' work.