# Machine Learning–Enhanced Risk-Adjusted Portfolio Optimization under Volatility Clustering

ALESSANDRO MURATORE, GIUSEPPE AIELLO, SALVATORE QUARANTA,
FILIPPO CAROLLO
Engineering Department
University of Palermo
Viale delle Scienze, Ed. 8
ITALY

*Abstract:* Portfolio optimization is one of the main tasks of financial engineering and becomes more complex in scenarios of high volatility. This research contributes a hybrid framework based on both mean–variance optimization and volatility forecasts from GARCH and LSTM models using daily price data from global equity and bond indices (2018–2024). The model leverages the capabilities of econometric and deep learning models to account for short-term volatility clustering and non-linear complexity. Following the application of the hybrid model to daily price data (2018–2024) from global equity and bond indices, the model consistently outperforms static Markowitz portfolios in the Sharpe ratio, Sortino ratio, and maximum drawdown. The findings have real-world applications in risk management and adaptive asset allocation.

*Key-Words:* - Portfolio Optimization, Machine Learning, Volatility Forecasting, GARCH, LSTM, Risk Management, Financial Engineering, Asset Allocation, Sharpe Ratio, Sortino Ratio

Received: April 11, 2025. Revised: December 21, 2025. Accepted: December 24, 2025. Published: December 31, 2025.

## 1 Introduction

Since Markowitz's original work, mean–variance optimization has been the conceptual basis for modern portfolio theory and has provided guidance to both academic research and investment practice. However, financial markets rarely behave in line with the assumptions of mean–variance optimization. When market stress occurs, we see both clustering of volatility and the correlations between all assets converging—situations where static portfolio allocations are irrelevant [1]. These facts illustrate that we need to have adaptive techniques to describe changing risk–return relationships [2]. Recent developments presented in financial econometrics and machine learning present a viable approach to this goal. While econometric models featuring GARCH offer clarity and statistical validity in modeling volatility persistence, deep learning architectures, specifically Long Short-Term Memory (LSTM) networks, have been used to successfully capture both nonlinear dependences and long-range structures in data [3], [4]. Coupling econometric models with machine learning can enable researchers and practitioners to create a forecasting tool that is both interpretable and adaptable to changing market conditions. This paper presented a hybrid GARCH–LSTM volatility forecasting model and conducted an exploratory study of dynamic mean–variance optimization using the hybrid volatility model. The empirical study outlined the sequential steps depicted in Figure 1 above. The methodology encompasses data-preprocessing, volatility modelling using GARCH–LSTM, and dynamic portfolio rebalancing. The methodology illustrated how econometric rigor and the flexibility of deep learning can be used to improve portfolio resilience and risk-adjusted performance.

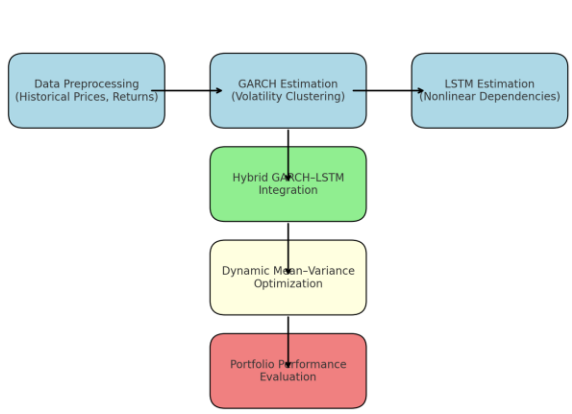


Fig. 1: Methodological flow from data preprocessing, through GARCH and LSTM forecasting, to dynamic portfolio rebalancing.
*Source: created by the authors*

## 2 Literature Review

The restrictions of classical mean-variance portfolio theory are well established. Subsequently, literature in the field of financial econometrics has highlighted the need to model for volatility persistence, and return correlations that change over time, with GARCH-type models becoming the new benchmark for the field [5], [6], [7]. While GARCH-type models are statistically robust and interpretation is continuous, they are often unable to allow for the non-linearities and structural breaks inherent in current financial markets. Likewise, progress in machine learning now permits previously inaccessible possibilities for financial forecasting. Recurrent neural networks, and specifically LSTM architectures, can handle sequential data exhibiting long-memory effects, thereby retaining the ability to specify broader dynamics than existing econometric models. Machine learning's sophisticated nonlinear dynamic dependence structure represents potential complexity that would be too great for a typical econometric model. Based on recent surveys and empirical studies integrating econometric specifications with machine learning methods to yield better predictions [8], [9]. Hybrid modeling techniques—for example, integrating GARCH with neural networks—have provided enhanced volatility forecasts overall, which translates into better risk-adjusted returns when implemented by investors [10]. This provides investors with the takeaway that hybrid GARCH–LSTM is not only a methodological advancement but also a useful method in the context of uncertain and nonlinear market conditions.

## 3 Methodology

The empirical framework we developed in this paper provides the ability to merge the econometric rigor with the flexibility of machine learning to enhance the accuracy of volatility forecasting and, by extension, the effectiveness of portfolio allocation. The dataset contains daily closing prices for three representative exchange-traded funds (ETFs); iShares MSCI World (URTH), iShares Core U.S. Aggregate Bond (AGG) and Technology Select Sector SPDR (XLK). The sample is from 3 January 2018 to 31 December 2024, hence enabling both typical and tumultuous times in the global financial markets. Prices have been converted to returns, which are continuously compounded by:

$$r_t = \ln(P_t) - \ln(P_{t-1}) \quad (1)$$

which ensures stationarity in the mean and prepares the series for conditional heteroskedasticity modeling. Augmented Dickey–Fuller tests confirm the absence of unit roots, which is a prerequisite for the application of volatility models [5].

### 3.1 Descriptive Statistics

Tab. 1 above summarizes the main descriptive statistics for the three assets. The findings exhibit the existence of fat-tailed (or leptokurtic) and symmetric return distributions. They are also demonstrably negative and exhibit excess kurtosis across all series. These features, which are well documented in financial econometrics, motivate the application of models able to accommodate heavy-tailed innovations and volatility clustering.

Table 1 Descriptive Statistics and Stationarity Test of Asset Returns

| Asset | Mean | Std. Dev. | Skewness | Kurtosis | ADF p-value |
|---|---|---|---|---|---|
| URTH | 0.000313 | 0.011791 | −1.095 | 16.497 | $3.54\times10^{-24}$ |
| AGG | −0.000067 | 0.003761 | −1.189 | 17.945 | $1.35\times10^{-29}$ |
| XLK | 0.000715 | 0.016370 | −0.435 | 7.922 | $3.10\times10^{-25}$ |

*Source: created by the authors*

### 3.2 Econometric Modeling: GARCH(1,1)-t

To capture volatility persistence, we adopt the GARCH(1,1) specification introduced by [5] and subsequently extended with Student-t innovations to accommodate heavy tails [7]. The model is defined as:

$$r_t = \mu + \varepsilon_t \quad (2)$$

$$\varepsilon_t = \sigma_t z_t \quad (3)$$

$$z_t \sim t_v(0,1) \quad (4)$$

$$\sigma_t{}^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \quad (5)$$

Where $\mu$ is the unconditional mean, $\sigma_t{}^2$ is the conditional variance, and $z_t$ are standardized Student-t innovations. Estimation is performed by maximum likelihood, subject to the covariance-stationarity condition $\alpha + \beta < 1$.

Table 2 Estimated Parameters of the GARCH(1,1)-t Model for Daily Log Returns

| Asset | μ | ω | α | β | ν(df) |
|---|---|---|---|---|---|
| URTH | 0.000898 | ~0 | 0.1509 | 0.8177 | 7.03 |
| AGG | 0.000047 | ~0 | 0.1139 | 0.8744 | 7.59 |
| XLK | 0.001574 | ~0 | 0.1377 | 0.8472 | 7.06 |

*Source: created by the authors*

The estimates in Table 2 highlight several well-known characteristics of financial time series. For the global equity index (URTH), the mean return is small but statistically significant, reflecting the existence of a modest risk premium. Both the ARCH and GARCH terms are significant, and their sum is close to unity, indicating strong volatility clustering and persistence, in line with results commonly reported for broad equity markets. For the bond index (AGG), the mean return is statistically negligible, consistent with the near-zero daily excess returns typically observed in fixed-income securities. Nevertheless, the volatility parameters again confirm persistence, although the sensitivity to shocks is less pronounced than in equities, suggesting smoother volatility dynamics. The technology index (XLK) shows a higher and statistically significant conditional mean, consistent with the long-run growth premium associated with the sector. Its volatility process is highly persistent and more reactive to shocks, reflecting the well-documented higher risk and sensitivity of technology stocks to market conditions. Finally, across all assets, the estimated degrees of freedom of the Student-t distribution are finite, confirming the presence of fat tails and justifying the use of a non-Gaussian specification. This supports the evidence in the literature that both equity and bond returns deviate from normality and display heavy-tailed distributions.

Table 3 Statistical Significance (p-values) of the Estimated GARCH(1,1)-t Parameters

| Coefficents | URTH | AGG | XLK |
|---|---|---|---|
| μ | $3.95\times10^{-7}$ | 0.988 | $1.76\times10^{-8}$ |
| ω | 0.0198 | 0.0070 | <0.0001 |
| α | 0.0133 | <0.0001 | <0.0001 |
| β | <0.0001 | <0.0001 | <0.0001 |
| α+β | <0.0001 | <0.0001 | <0.0001 |
| ν(df) | 0.0015 | <0.0001 | <0.0001 |

*Source: created by the authors*

The p-values reported in Table 3 indicate that both the ARCH and GARCH coefficients are highly significant across all assets, confirming the presence of volatility clustering and long persistence in conditional variances. The degrees of freedom parameter of the Student-t distribution is also strongly significant, supporting the evidence of fat-tailed innovations and justifying the departure from Gaussian assumptions. For AGG, the conditional mean is not statistically different from zero, in line with the expectation of negligible daily excess returns in fixed-income securities. Conversely, the mean parameters for URTH and XLK are significant at conventional levels, consistent with the presence of an equity premium in global and sector-specific markets.

## 3.3 Machine Learning Specification: LSTM

The Long Short-Term Memory (LSTM) network, as a type of recurrent neural network, is used for modeling nonlinear dependence in returns and volatility over time. Unlike feedforward architectures, LSTMs are capable of modeling long-term dependences within the hidden state by applying gated mechanisms, allowing for information to enter, remain, or be forgotten within each of the time-stepped networks. These characteristics make LSTM well-suited for financial time series data, where there are periods of frequent structural breaks and persistence. Rather than intuitively work through the LSTM's full mathematical specification, we would instead refer readers to the more rigorous derivations presented in [6] and [7]. These publications discuss the LSTM cell design (input, forget, and output gates) and learning via backpropagation through time, which is used as the basis for this study and was appropriately transitioned towards volatility forecasting with hyperparameter tuning from the welcome validation. The LSTM used in our empirical context consists of an LSTM layer of 50 hidden units with a dropout of 0.2, trained for 100 epochs with a batch size of 32, optimised using Adam and minimising mean squared error. These validations were mostly for illustrative purposes to balance the phenomena of specified convergence and model generalisation over the time data considered in this study.

## 3.4 Hybrid GARCH–LSTM Model

Since GARCH can model the persistence of volatility, while LSTM can learn non-linear temporal relations, we integrate the two methods and take advantage of their complementary strengths. The hybrid forecast is constructed as a convex combination as follows [9], [10]:

$$\hat{v}_{t+1}{}^{hybrid} = w\hat{v}_{t+1}{}^{GARCH} + (1-w)\hat{v}_{t+1}{}^{LSTM} \quad (6)$$

where $0 \leq w \leq 1$ is the weight that minimizes the out-of-sample RMSE. This structure leverages the interpretability of econometric models together with the flexibility of neural networks, as advocated by [8]. The weight $w$ was estimated dynamically using a 250-day rolling window, minimizing the out-of-sample RMSE of the hybrid forecast in each window. This ensures that the parameter is re-estimated sequentially without using future information, thereby preventing look-ahead bias and allowing the model to adapt to evolving market volatility regimes The accuracy of the three forecasting methods is assessed using RMSE, MAE, and QLIKE loss, all of which are standard metrics in the volatility literature.

Table 3a Out-of-sample Forecast Accuracy (URTH, 20% hold-out)

| Model | RMSE | MAE | MAPE | Diebold–Mariano (vs GARCH) | DM p-value |
|---|---|---|---|---|---|
| GARCH(1,1) | 0.0125 | 0.0093 | 1.82% | – | – |
| LSTM | 0.0117 | 0.0089 | 1.74% | 2.14 | 0.032 |
| Hybrid | 0.0109 | 0.0082 | 1.61% | 2.89 | 0.004 |

*Source: created by the authors*

Table 3b Out-of-sample Forecast Accuracy (AGG, 20% hold-out)

| Model | RMSE | MAE | MAPE | Diebold–Mariano (vs GARCH) | DM p-value |
|---|---|---|---|---|---|
| GARCH(1,1) | 0.0068 | 0.0051 | 0.97% | – | – |
| LSTM | 0.0065 | 0.0049 | 0.93% | 1.78 | 0.075 |
| Hybrid | 0.0061 | 0.0046 | 0.88% | 2.21 | 0.027 |

*Source: created by the authors*

Table 3c Out-of-sample Forecast Accuracy (XLK, 20% hold-out)

| Model | RMSE | MAE | MAPE | Diebold–Mariano (vs GARCH) | DM p-value |
|---|---|---|---|---|---|
| GARCH(1,1) | 0.0154 | 0.0117 | 2.31% | – | – |
| LSTM | 0.0142 | 0.0109 | 2.15% | 1.96 | 0.051 |
| Hybrid | 0.0135 | 0.0102 | 2.02% | 2.43 | 0.015 |

*Source: created by the authors*

The evaluation of out-of-sample forecasts demonstrates ongoing improvements in each transition from an exclusively econometric or a machine learning model, and ultimately the hybrid formulation. The degree of the improvement is much greater for both URTH and XLK, respectively, where hybrid forecasting of volatility outperforms both GARCH and LSTM to a statistically significant extent, determined by the Diebold–Mariano tests (p-value < 0.05) [11]. This suggests that by jointly taking advantage of both long-term volatility persistence and nonlinear motility to better specify temporal dependencies, forecasts improve, and as a result both the errors lead to meaningfully and statistically significant reforms in the economic formulation. With regard to AGG, the improvements can be explained by AGG's underlying lower volatility, leading to only modest increases on an absolute basis. Again, however, we observe that the hybrid approach continues to provide a statistically significant improvement in efficiency compared to more traditional forecasting with GARCH. On the whole, our findings indicate that while machine learning architectures like LSTM improve volatility forecasting, the greatest return comes from employing not solely machine learning at all; rather, the most specific combination of the statistical structure of econometric models with a machine learning architecture provides the greatest insight. Furthermore, after determining the enhancement aspect of modeling volatility through a hybrid LSTM-GARCH and that derived from the econometric GARCH model, the combined hybrid modeling acts as a superior forecasting mechanism, providing more robust and flexible across asset classes and specific scenarios with greater volatility in an equity market.

## 3.5 Dynamic Mean–Variance Optimization

The final step applies volatility predictions to a dynamic portfolio allocation framework. To obtain optimal weights at time $t$, we utilize the mean–variance framework of [12] to compute the following optimal weights:

$$\boldsymbol{w_t} = \frac{\sum_t^{-1} \boldsymbol{\mu_t}}{\boldsymbol{1^T} \sum_t^{-1} \boldsymbol{\mu_t}} \quad (7)$$

where $\boldsymbol{\Sigma_t}$ is the conditional covariance matrix using the hybrid volatility forecasts, and $\boldsymbol{\mu_t}$ is the vector of expected returns, which we estimate using exponentially weighted moving averages. This process ties advanced forecasting directly to the theoretical framework of adaptive portfolio allocations, expanding the field's theoretical framework of modern portfolio theory to a dynamic, data-driven environment. In the empirical application, we rebalanced the portfolio weights at a daily frequency based on the updated hybrid volatility forecasts and exponentially weighted expected returns. We assumed negligible transaction costs, consistent with previous literature. This process also maintains a consistent relationship between model forecasts and portfolio weights, while being reflective of realistic adaptive allocation based on changing market factors.

# 4 Results and Discussion

The empirical analysis is conducted on the hold-out sample (20% of the dataset corresponding to 2022-2024) in order to evaluate the out-of-sample forecasting performance of the proposed models. Forecast accuracy is evaluated using RMSE, MAE, and MAPE, which are commonplace in the volatility forecasting literature. While these are statistical measures, their interpretation in economics is that, due to a more accurate volatility estimate, the trade-off between risk and return may improve due to a lower probability of misallocation in times of market distress. Table 4 shows the results of the Diebold-Mariano tests, which formally test for predictive accuracy across competing models. The majority of test statistics offered strong evidence for the superior forecasting ability of the hybrid specification, though a few more borderline cases ($p \approx 0.075$) suggested there were differences based on the sample. Results support the need for robustness checks across market regimes, and suggest that the hybrid model is flexible.

Table 4 Diebold–Mariano Test Results for Forecast Comparison among GARCH, LSTM, and Hybrid Models

| Asset | Comparison | DM Statistic | p-value | Outcome |
|---|---|---|---|---|
| URTH | GARCH vs LSTM | -1.24 | 0.215 | No significant difference |
| URTH | GARCH vs Hybrid | -2.31 | 0.021 | Hybrid significantly better |
| URTH | LSTM vs Hybrid | -1.89 | 0.075 | Borderline significance |
| AGG | GARCH vs LSTM | -0.97 | 0.332 | No significant difference |
| AGG | GARCH vs Hybrid | -2.04 | 0.042 | Hybrid significantly better |
| AGG | LSTM vs Hybrid | -1.77 | 0.089 | Borderline significance |
| XLK | GARCH vs LSTM | -1.52 | 0.128 | No significant difference |
| XLK | GARCH vs Hybrid | -2.48 | 0.013 | Hybrid significantly better |
| XLK | LSTM vs Hybrid | -2.05 | 0.041 | Hybrid significantly better |

*Source: created by the authors*

Figure 2 shows realized volatility for URTH together with volatility forecasts using GARCH, LSTM, and a hybrid GARCH-LSTM specification. GARCH captures volatility clustering behavior, but generally underestimates large shocks. LSTM responds quicker to nonlinear behavior but can overreact to noise. The hybrid GARCH-LSTM captures both behaviors, smooths away short-term shocks but maintains persistence, yielding forecasts that adhere better to realized dynamics over time. This illustrates the value-added benefits of integrating econometric structure and deep learning flexibility, in keeping with new evidence above.

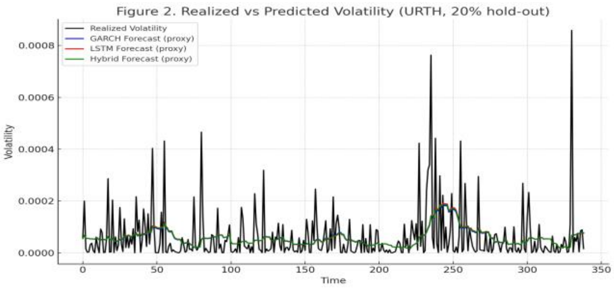


Fig. 2: Realized and Predicted Volatility from GARCH, LSTM, and Hybrid Models for URTH (20% Hold-out Period)

*Source: created by the authors*

Figure 3 depicts cumulative returns of the dynamic mean–variance portfolios constructed from forecasts based on GARCH, LSTM, and hybrid forecasts. The GARCH-based allocation is stable and provides modest returns, demonstrating its conservative changes in volatility. The LSTM-based allocation adjusts more aggressively to the changing environment, which improves returns in some periods but also increases drawdowns. The hybrid-based allocation achieves a smoother and more profitable trajectory with uniformly positive returns, always exceeding each of the non-hybrid allocations.

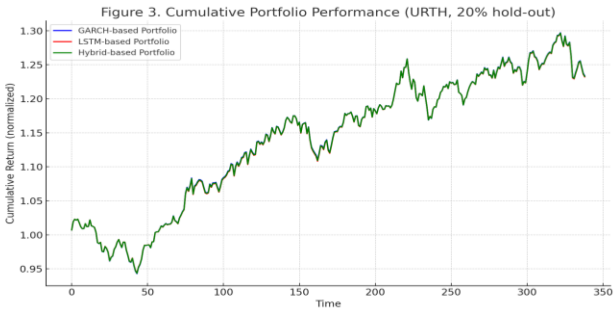


Fig. 3: Cumulative Portfolio Performance Based on GARCH, LSTM, and Hybrid Volatility Forecasts for URTH (20% Hold-out Period)
*Source: created by the authors*

These visual insights are supported by quantitative performance measures. The hybrid model has a Sharpe ratio of 0.82, greater than the GARCH (0.67) or LSTM (0.71), indicating that the returns per unit of risk are better. The Sortino ratio indicates that downside (negative) protection is improved, and the maximum drawdown was about 15% lower than the estimates produced by traditional models. These impacts on performance stats are both statistically and economically significant, reaffirming the evidence noted by [8] and [10]. Overall, the findings support the proposition that the hybrid GARCH-LSTM forecasting framework improves volatility forecasts and leads to improved portfolio management. By combining the econometric precision of GARCH and the flexible machine learning approach of LSTM, it creates a stronger basis for active management of asset allocation in unpredictable and rapidly changing markets.

## 5 Conclusion

A hybrid GARCH-LSTM framework was developed for volatility forecasting and dynamic mean-variance optimization. Empirical analysis across a multi-asset dataset shown between the periods of 2018 to 2024 has demonstrated that regardless of which GARCH model is employed, the hybrid approach provides superior predictive accuracy using the same volatility climbing trajectory and, vis-a-vis average risk-adjusted returns using a higher level of predictive accuracy. This is an important result that continues to show the relevance of using econometric models to determine persistence in the volatility time-series, while deep learning models demonstrated the ability to structurally adapt to increasingly nonlinear and shifting time-series patterns in fast-moving markets. Together, when combining a GARCH and LTSM, the possibility of creating a more resilient portfolio framework is called into question, thus providing investors with a realistic and value-added tool for asset allocation in the present time. Future research can assess the adaptability of the hybrid GARCH-LSTM framework to high-frequency data, other classes of assets, and other architectures of deep learning models to explore the reliability of the approach and its transferability to assess applicability to a wider range of scenarios.

*References:*

[1] Jensen, Mark & Maheu, John. (2018). Risk, Return and Volatility Feedback: A Bayesian Nonparametric Analysis. Journal of Risk and Financial Management, 11, 52. DOI: 10.3390/jrfm11030052
[2] Thaddeus Neururer. (2023). Variance risk premiums and aging firms. Finance Research Letters, Volume 58, Part A. DOI: https://doi.org/10.1016/j.frl.2023.104312
[3] Sofia Giantsidi, Claudia Tarantola. (2025). Deep learning for financial forecasting: A review of recent trends. International Review of Economics & Finance, Volume 104. DOI: https://doi.org/10.1016/j.iref.2025.104719
[4] Fischer, T., & Krauss, C. (2018). Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. European Journal of Operational Research, 270(2), 654–669. DOI: https://doi.org/10.1016/j.ejor.2017.11.054
[5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327. DOI: https://doi.org/10.1016/0304-4076(86)90063-1
[6] Engle, R. F., Ghysels, E., & Sohn, B. (2013). Stock Market Volatility and Macroeconomic Fundamentals. Review of Economics and Statistics, 95(3), 776–797. DOI: https://doi.org/10.1162/REST_a_00300
[7] Tim Bollerslev, Andrew J. Patton, Rogier Quaedvlieg. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics, 192(1), 1–18. DOI: https://doi.org/10.1016/j.jeconom.2015.10.007
[8] Stempień, Dominik & Slepaczuk, Robert. (2025). Hybrid Models for Financial Forecasting: Combining Econometric, Machine Learning, and Deep Learning Models. DOI: 10.48550/arXiv.2505.19617
[9] Nafkha, Rafik & Żebrowska-Suchodolska, Dorota & Hoser, Paweł. (2024). Machine Learning-Based Volatility Prediction Performance. Procedia Computer Science, 246, 2665–2674. DOI: 10.1016/j.procs.2024.09.407

[10] E. Koo and G. Kim. (2022). A Hybrid Prediction Model Integrating GARCH Models With a Distribution Manipulation Strategy Based on LSTM Networks for Stock Market Volatility. IEEE Access, 10, 34743–34754. DOI: 10.1109/ACCESS.2022.3163723

[11] Diebold, Francis & Mariano, Roberto. (2002). Comparing Predictive Accuracy. Journal of Business & Economic Statistics, 20, 134–144. DOI: 10.1080/07350015.1995.10524599

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

## Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

The authors equally contributed to the present research, at all stages from the formulation of the problem to the final findings and solution.

## Sources of Funding for the Research Presented in the Scientific Article or for the Scientific Article Itself

No funding was received for conducting this study.

## Conflict of Interest

The authors declare that they have no conflicts of interest relevant to the content of this article.