Forecast Combination for Asset Classes ETFs: Insights on Market Efficiency and Arbitrage

Rodrigo Baggi Prieto Alvarez

NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal; ISCTE-IUL, Lisbon; IPS, Setúbal, Portugal

[0009-0008-1682-9653]
rodrigo.alvarez@novaims.unl.pt

Jorge Miguel Ventura Bravo

NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal; ISCTE-IUL BRU, Lisbon, Portugal; CEFAGE-UE, Évora, Portugal; Université Paris-Dauphine PSL & LEDa, Paris, France

[0000-0002-7389-5103]

This is the final Author Peer Reviewed version of:

Alvarez, R. B. P., & Bravo, J. M. (2026). Forecast Combination for Asset Classes ETFs: Insights on Market Efficiency and Arbitrage. In Á. Rocha, F. García Peñalvo, C. J. Costa, & R. Gonçalves (Eds.), Proceedings of 20th Iberian Conference on Information Systems and Technologies (CISTI 2025) (Vol. 2, pp. 274-286). (Lecture Notes in Networks and Systems; Vol. 1717). Springer. https://doi.org/10.1007/978-3-032-10721-3_25

Funding/Acknowledgments: This work was supported by national funds through FCT (Fundação para a Ciência e a Tecnologia), under the project - UID/04152/2025 - Centro de Investigação em Gestão de Informação, MagIC, NOVA IMS, Universidade NOVA de Lisboa - https://doi.org/10.54499/UID/04152/2025 (2025-01-01/2028-12-31), UID/PRR/04152/2025 https://doi.org/10.54499/UID/PRR/04152/2025 (2025-01-01/ 2026-06-30) and UIDB/00315/2020 - Instituto Universitário de Lisboa (ISCTE-IUL).

Forecast Combination for Asset Classes ETFs: Insights on Market Efficiency and Arbitrage

Rodrigo Baggi Prieto Alvarez [1] [0009-0008-1682-9653]

Jorge Miguel Ventura Bravo [2] [0000-0002-7389-5103]

[1] PhD Student at NOVA-IMS, Lisbon, Portugal, and Assistant Professor at ISCTE-IUL
rodrigo.alvarez@novaims.unl.pt
[2] Associate Professor at NOVA-IMS (MagIC, Lisbon, Portugal), Université Paris-Dauphine, BRU-ISCTE-IUL and CEFAGE-UE
jbravo@novaims.unl.pt

Abstract. The Exchange-Traded Funds (ETFs) have transformed asset allocation, allowing investors to gain exposure to diverse asset classes with a single instrument. In turn, forecast combination models have emerged as advantageous methods for improving prediction accuracy. While the Efficient Market Hypothesis (EMH) posits that prices fluctuate randomly, making abnormal returns unattainable, empirical evidence reveals autocorrelation in stock returns, challenging the EMH's strict interpretation. This raises the question of whether new econometric and machine learning methods can predict asset values more effectively. We investigate the effectiveness of forecast combination in predicting financial asset classes. Analyzing ETFs across equity, fixed income, commodities, and cryptocurrency markets, we test the predictive accuracy of ten econometric and machine learning models, and preliminary results suggest that ensemble methods can indeed outperform simple models. Comparisons between forecasts and futures market prices reveal potential inefficiencies, suggesting opportunities for spot-futures index arbitrage. These findings contribute to discussions on market efficiency and highlight the role of forecast combinations in improving asset predictability and portfolio management.

Keywords: ETF; Asset Prediction; Forecast Combination; EMH; Spot-Futures Index Arbitrage.

1 Introduction

Exchange-Traded Funds (ETFs) have transformed modern finance by offering investors diversified, low-cost, and liquid instruments to gain exposure to a wide array of asset classes, including equities, bonds, commodities, and alternative investments

(Madhavan, 2016; DiLellio & Stanley, 2011). This transparency and liquidity (ETFs are traded on exchanges like individual stocks) have made them an important tool for institutional and retail investors seeking to implement strategic and tactical asset allocation. Their scope involving sector-specific, regional, and several asset classes has also encouraged a growing interest in understanding predictive patterns and implications for efficiency. The Efficient Market Hypothesis (EMH) posits that asset prices reflect all available information, leaving little room for predictable patterns. However, studies by Marshall et al. (2013) and Box et al. (2021) highlight that ETF mispricing, while short-lived, provide opportunities for intraday arbitrage.

The predictability observed in ETF flows and price deviations, as discussed by Brown et al. (2021), challenges the strict form of the EMH, suggesting an adaptive market where inefficiencies arise intermittently. By integrating findings from Sheng & Ma (2022), who demonstrate the profitability of machine learning-driven arbitrage strategies, and Saha et al. (2021), who emphasize the episodic nature of ETF efficiency, we intend to evaluate how forecasting models would really reflect market efficiency or allow arbitrage opportunities. This raises the question of whether econometric and machine learning methods and forecast combinations can predict asset prices more effectively (e.g. Bravo, 2023). Also, comparing forecasts with futures market prices may reveal potential arbitrage opportunities, offering insights into market inefficiencies.

Therefore, this study aims to address the following questions: which forecasting model is best suited to predict main ETFs asset classes? For each ETF, can a single model individually (or a simple combination) be more accurate than complex algorithms based on ML techniques? And what do the results may reveal regarding the predictability of these assets in terms of EMH and the spot–futures arbitrage opportunities? Accurate predictions are vital for portfolio management, and understanding the predictive behavior of ETFs across different asset classes and examining the efficiency of these markets thus becomes essential for both academics and practitioners in finance. Recent studies provide some critical insights into these questions. For instance, the role of machine learning in forecasting financial time series, as demonstrated by Sheng & Ma (2022), highlights the effectiveness of models like LSTM and XGBoost in predicting arbitrage opportunities in ETFs and related derivatives. Similarly, Piovezan et al. (2024) underscore the utility of classification and regression models in enhancing forecast precision, often outperforming traditional benchmarks like buy-and-hold strategies. These findings underscore the evolving landscape of financial forecasting and the need to explore model combinations and their implications for market efficiency.

Furthermore, forecast combination models have emerged as advantageous methods for improving prediction accuracy. Ensemble learning approach, notably through advanced machine leaning methods, has demonstrated superiority over individual models by integrating different predictive insights to minimize errors, besides positioning as a potential solution to the so-called Forecast Combination Puzzle. Empirical studies support that combining models can capture diverse market signals more effectively, providing more robust forecasts. This study contrasts diverse forecasting models, from traditional econometric approaches to advanced ML algorithms. Econometric methods such as SARIMA-GARCH models and their extensions provide a foundational framework, but studies by Shih et al. (2024) and Day & Lin (2019) also indicate that NN and

LSTM algorithms outperform the linear models in ETF prediction accuracy. Machine learning models like XGBoost and Random Forest are particularly effective for identifying patterns in high-dimensional datasets, as shown by Sheng & Ma (2022). However, Gowani & Kanjiani (2024) illustrate that sector-specific ETFs benefit from specialized LSTM architectures capable of predicting directional changes with high accuracy across various market conditions.

Nevertheless, mixing methods could be the best approach at capturing linear and nonlinear long-term dependencies in ETF time-series. Brown et al. (2021) suggest that combining models can mitigate individual biases, particularly in volatile market conditions. Forecast combination, as noted by Piovezan et al. (2024), could further enhance predictive power, leveraging the strengths of different algorithms for more robust forecasts. While prior studies highlight the superior predictive accuracy of machine learning models, to our knowledge, there is still limited exploration in the literature of the integration between econometric and ML approaches and the use of forecast combination for predicting ETF index. For instance, combining the interpretability of SARIMA models with the predictive power of LSTM could bridge a gap between traditional finance theories and data-driven methods. Key variables influencing ETF forecasting include their own historical indexes and sector-specific trends. Another unexplored avenue involves understanding the influence of macroeconomic shocks on model performance, offering an additional opportunity for comparison in this research.

We explore forecast combination models in predicting ETF values across multiple asset classes, particularly equities, fixed income, commodities, and cryptocurrency, with implications for market efficiency and arbitrage opportunities. By evaluating accuracy across different forecasting horizons, this study contributes to the understanding of asset predictability within the framework of EMH and more broadly to the literature on financial econometrics by providing insights into the interplay between forecast combination, market efficiency, and arbitrage. Emekter et al. (2022) highlight the importance of ETFs for portfolio optimization, providing a basis for integrating forecasting output into actionable investment strategies. In other words, the findings will inform practitioners on strategies through model-driven insights, bridging the gap between academic research and real-world applications.

The remainder of this paper is divided as follows: in section 2 we briefly present the theoretical and empirical literature framework, in section 3 we describe the data and estimated models, in section 4 we highlight the main results found, and finally in section 5 we present some concluding remarks and possible further improvements based on this exploratory study.

2 Related literature

Research on efficient markets and the behaviour of stock prices has evolved significantly, exploring both random walk hypotheses and the possibility of predictable patterns in asset prices. In his seminal work, Samuelson (1965) had demonstrated that anticipated prices should fluctuate randomly under rational expectations and the EMH predicted that must be impossible for investors to consistently achieve abnormal returns through market timing or stock picking (e.g. Fama, 1970, 1991). On the other hand, posterior empirical evidence challenged the random walk model by uncovering significant autocorrelation in stock returns, suggesting the presence of mean-reverting or non-stationary components that contradict the EMH's strict interpretation (Lo & MacKinlay, 1988). Additionally, under this critical perspective, markets could not be fully efficient because, paradoxically, if they were, there would be no incentive for investors to acquire costly information (Grossman & Stiglitz, 1980).

The analysis of EMH provides understanding of how information is reflected in asset prices and the extent to which prices can be anticipated based on past or current information. In this context, the relationship between spot and futures prices in stock index becomes critical, as it shows how arbitrage opportunities may arise due to temporary misalignments in pricing. For instance, in this spot-futures prices relationship, it was verified thar transaction costs may create a "no-arbitrage" band, allowing prices to deviate, but without enabling profitable arbitrage opportunities (Modest & Sundaresan, 1983). Arbitrage strategies could also be constrained by maturities (Mackinlay & Ramaswamy, 1988), and eventually would be index futures leading spot prices, particularly during periods of high volatility, suggesting that futures markets may play a dominant role in price discovery (Stoll & Whaley, 1990).

This relationship between spot and futures prices not only reflects market dynamics but also informs strategies in asset allocation, particularly through ETF trading. Understanding the conditions under which arbitrage opportunities arise – or are constrained – reinforces the challenges investors face in capturing price inefficiencies. This is important for structuring portfolios balancing risk-return and leveraging the liquidity and diversification benefits. For example, the recent creation of Bitcoin ETFs, despite the regulatory barriers and the underlying exacerbated volatility, highlight the search for diversification benefits, even though the striking challenges in achieving consistent returns across different categories of ETFs (Diaw, 2024).

Different strategies through ETFs trading have been empirically examined, with emphasis on exploiting price inefficiencies to support portfolio adjustments (DiLellio & Stanley, 2011). Efficient asset allocation for individual investors reinforces how diversification across indices, sectors or regions may help mitigate risk (Emekter et al., 2022). The evidence suggests that this type of instrument reduces specific risks (Herzog et al., 2023), however there are still limitations in achieving positive alphas despite their exposure to raw and excess returns (Rompotis, 2024), illustrating challenges in surpassing benchmark performance. Examples include (Sheng & Ma, 2022), who applied machine learning models to high-frequency data to test pricing efficiency and construct ETF-based portfolios with minimal tracking errors, and Day & Lin (2019), who develop an AI framework by constructing a robot-advisor for ETF prediction and portfolio optimization.

In the pursuit of consistent returns through asset diversification, Timmermann (2006) noted that the problem of forming mean-variance efficient portfolios in finance is mathematically equivalent to that of combining forecasts. The approach has been explored to mitigate the unpredictability inherent in financial markets. This original literature evidenced a puzzle, demonstrating that combining forecasts of the same variable often yields better performance than relying on a single model (Bates & Granger, 1969; Granger, 1989, Clemen, 1989), and simple combinations often perform comparably to complex methods (Hsiao & Wan, 2014; Smith & Wallis, 2009). Claeskens et al. (2016) address this puzzle highlighting that estimating optimal weights often underperform due to the increased variance introduced during the estimation process, arguing that the equal-weight approach can outperform when the estimation error in weights is significant, providing a theoretical justification for simple combinations.

Considering this aspect, Wang et al. (2023) provide an extensive review of over fifty years of research on forecast combinations, as a robust alternative to relying on single model for both point and probabilistic forecasts. The ensemble methods underscore the practical value of combining forecasts across a range of asset classes, supporting diversification and risk management strategies. Empirical studies applying ensemble methods include Tsai et al. (2011, 2014) and Bravo & Ayuso (2021), Also, recent methods like the meta-learning strategy, based on multiple heterogeneous learning algorithms, may perform better compared to usual combination approaches (Bravo, 2023).

Independently of the approach, the recent literature suggests that even in high-dimensional contexts – somewhat similar to diversified ETF portfolios – combining methods prevents some model-specific risks. And for this type of financial instrument, where investment strategies seek exposure to different indices, regions, sectors, or asset classes, combining models can enhance decision-making by mitigating both systematic trends and unique idiosyncrasies. Consequently, forecast combination may help portfolio diversification, positioning traditional econometric models and new machine learning methods as essential tools for both researchers and investors.

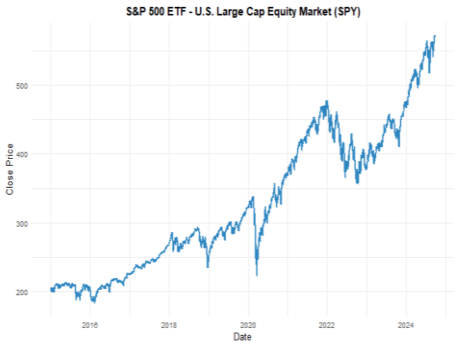
3 Data and methodology

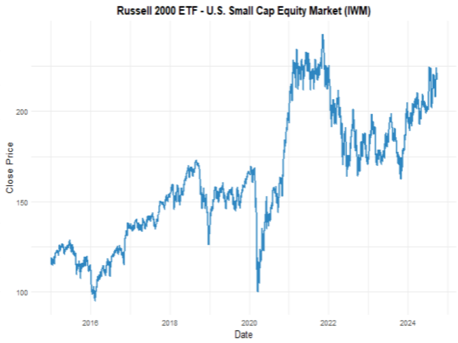
In this study, we use financial data accessed through the *getSymbols* function from the *quantmod* package in *R/RStudio*. The function allows us to collect daily prices for a range of financial assets, crucial for conducting up-to-date and comprehensive analysis. The selection covers a diverse spectrum of markets in terms of a portfolio diversification, including U.S. and international equities, bonds, cryptocurrency and commodity, to ensure a broad representation of financial asset classes.

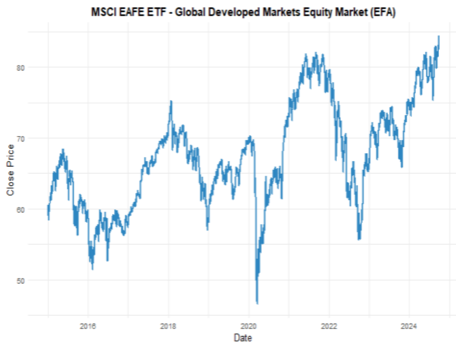
The data spans from January 1st, 2015, to September 30th, 2024, a rich time series for model training and evaluation. We analyze a representative group of 10 ETFs indices, holding a diversified portfolio that are traded on a stock exchange, allowing investors to buy and sell shares throughout the trading day (ticker symbol in parenthesis):

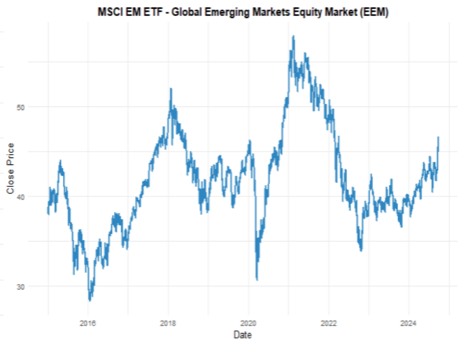
1) S&P 500 - U.S. Large Cap Equity Market (SPY)
2) Russell 2000 - U.S. Small Cap Equity Market (IWM)
3) MSCI EAFE - Global Developed Equity Market (EFA)
4) MSCI EM - Global Emerging Equity Market (EEM)
5) 7-10 Year Treasury Bond– U.S. Fixed Income (IEF)

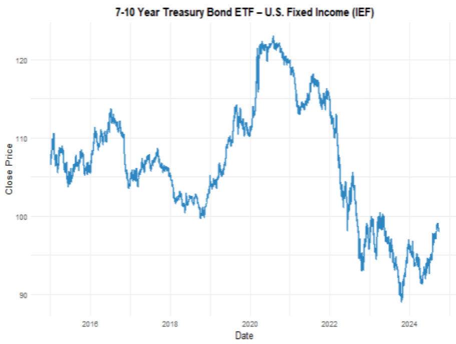
6) International Treasury Bond - Global Developed Markets Fixed Income (IGOV)
7) JP Morgan USD EM Bond - Global Emerging Markets Fixed Income (EMB)
8) iBoxx Investment Grade - U.S. Corporate Bonds (LQD)
9) SPDR Gold Shares ETF (GLD)
10) ProShares Bitcoin ETF (BITO)

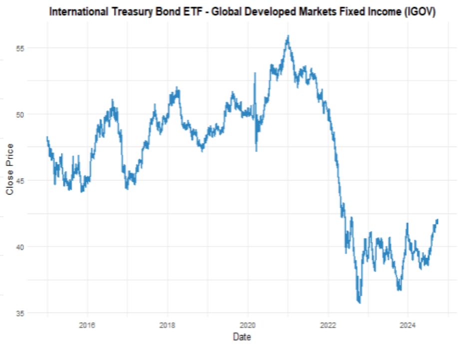












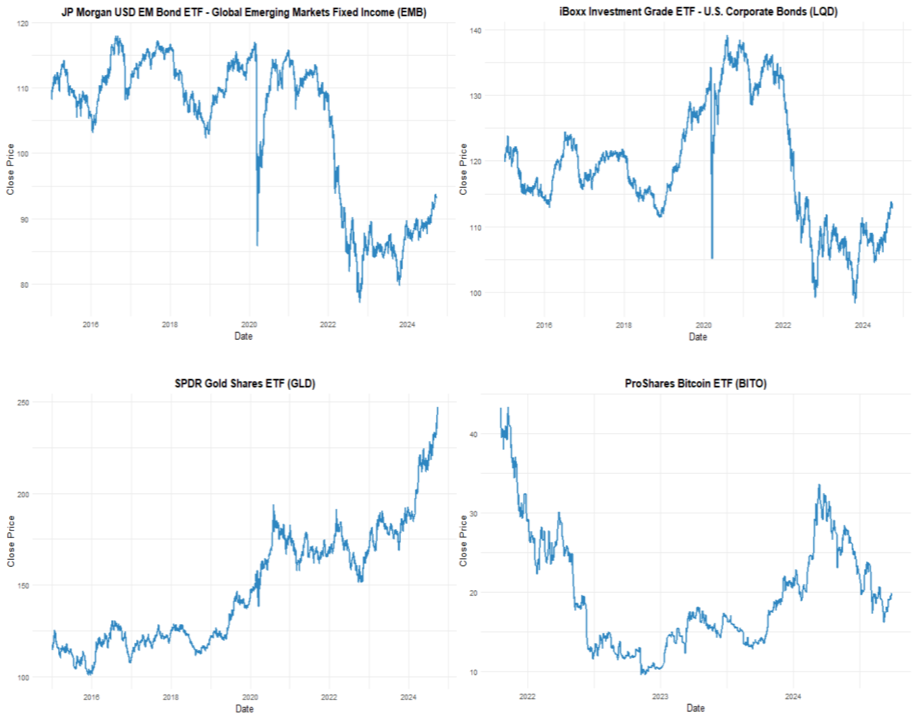


Fig. 1. Historical indices prices of ETFs: the 10 asset classes from 2015 to 2024.

Traditional econometric and machine learning models are employed as the tools for time series forecasting, each one bringing unique strengths to address linear and non-linear patterns.

1) ARIMA (Auto-Regressive Integrated Moving Average): Usual statistical model combining autoregressive, differencing, and moving average elements, effective for linear dependencies in stationary data (Box & Jenkins, 1970).
2) ETS (Exponential Smoothing State Space Model): Uses exponential smoothing to handle level, trend, and seasonality in non-stationary time series, with automatic parameter selection for efficient forecasting (Hyndman et al., 2008).
3) SVM (Support Vector Machine): A machine learning method that separates data in high-dimensional space, effective for small datasets (Cortes & Vapnik, 1995).
4) RF (Random Forest): A versatile ensemble approach creating multiple decision trees on data subsets, combining their outputs for stable predictions (Breiman, 2001).
5) XGBoost (Extreme Gradient Boosting): A model with regularization that requires fine-tuning of parameters like learning rate and tree depth (Chen & Guestrin, 2016).
6) ARIMA-SVR (hybrid model combining ARIMA with Support Vector Regression): Integrates ARIMA for linear trends with SVR to model non-linearities, suitable for time series with both types of dependencies (Pai & Lin, 2005).

7) ARIMA-ANN (hybrid model combining ARIMA with Artificial Neural Networks): Combines ARIMA for linear trends with ANN for non-linear residuals, enhancing predictive accuracy in mixed-pattern data (Zhang, 2003).

Additional forecast combinations (ensemble techniques) are employed to mix predictions, improve robustness and accuracy, and give some insights about the advantages of combination:

8) Simple Average: A straightforward ensemble method that averages forecasts to balance individual biases, commonly used to explore the "Forecast Combination Puzzle" in empirical forecasting studies (Bates & Granger, 1969; Clemen, 1989).
9) Ridge Regression: Uses ridge regression with L2 regularization, which shrinks coefficients to balance the influence of correlated forecasts, reducing overfitting and enhancing stability in the forecast combination (Hoerl & Kennard, 1970).
10) Graphical Least Absolute Shrinkage and Selection Operator (G-LASSO): Based on the regularization with an L1 penalty in the regression, shrinking some coefficients to zero. This novel method estimates forecast error dependencies, revealing conditional relationships between forecasts and optimizing combination weights for improved accuracy in high-dimensional contexts where traditional methods may struggle (Lee & Seregina, 2022).

The ETFs daily prices are analyzed to allow us to forecast future performance over horizons of one month, one quarter, and one year. The forecasting models' performances are evaluated using the Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE), enabling us to assess the accuracy of each model's predictions across different timeframes.

4 Preliminary results

After estimating the models and forecast combinations algorithms, we present a summary of the preliminary findings. The results obtained are displayed in the following tables and graphically illustrated in the appendix, which already allow for some important highlights:

Overall superiority of Ridge Regression and Graphical LASSO models, suggesting that combinations of models are superior to each model individually: As can be seen in Table 2, the analysis of MAPE and RMSE shows much smaller prediction errors in the more complex models, specifically in the combinations that use regularization methods and optimization algorithms to combine forecasting models.

The simple average showed inferior performance to more complex combination algorithms, discarding the Forecasting Combination Puzzle hypothesis: The performance of the models in Table 2 also indicates that the simple average is inferior to more complex combinations and even to individual models, evidence contrary to the Forecasting Combination Puzzle hypothesis that is discussed in empirical literature.

Lower forecast error for fixed income asset classes (Treasury and corporate bonds) and higher forecast error for cryptocurrency and equity indices: As highlighted in Table 3, the forecasts for ETFs linked to fixed income assets generally present lower MAPE and RMSE among all the estimated models and in the three forecast horizons tested, which would be expected considering the lower volatility of these assets.

Lower forecast error for shorter horizons (as expected) but Ridge Regression and Graphical LASSO perform 1 year ahead almost as well as 1 month ahead forecasting: Naturally, in general, short-term projections have lower errors than long-term estimates. However, it is interesting to note that the best forecast combinations maintain better consistency than the rest of the models, not presenting such a severe loss of accuracy when comparing their forecasts 1 month ahead, 1 quarter ahead, and 1 year ahead.

Preliminary results show the forecasting model predicting better the ETF indices than priced in futures market, suggesting possible opportunities for spot-futures index arbitrage: The performance of the models allows us to infer about the predictability of assets in light of the EMH and arbitrage opportunities. For example, in Table 1, the comparison between the values predicted by the G-LASSO algorithm and the pricing of the underlying assets in the futures market suggests that this model would have been able to predict more accurately in the horizons analyzed, indicating the possibility of profitable long-short operations by investors.

Table 1. ETF index predicted for 30.09.2024 with G-LASSO vs. Futures market for each forecast horizon.

		30.09.2023 **1 year**		
	Sport	*Futures**	*G-LASSO*	
S&P 500 (SPY)	**427.5**	**431.2**	**574.9**	
Russell 2000 (IWM)	**176.7**	**178.1**	**220.2**	
Gold (GLD)	**171.5**	**171.4**	**237.7**	
		30.06.2024 **1 quarter**		
	Spot	*Futures*	*G-LASSO*	
S&P 500 (SPY)	**544.2**	**550.3**	**560.7**	
Russell 2000 (IWM)	**202.9**	**204.6**	**218.6**	
Gold (GLD)	**215.0**	**215.1**	**241.0**	
		30.08.2024 **1 month**		**30.09.2024** **Effective**
	Spot	*Futures*	*G-LASSO*	**Index**
S&P 500 (SPY)	**563.7**	**564.9**	**569.9**	**573.8**
Russell 2000 (IWM)	**220.1**	**220.6**	**220.1**	**220.9**
Gold (GLD)	**231.3**	**230.4**	**245.0**	**243.1**

** Calculated as the percentage of variation futures contracts for the underlying asset on each date for 30.09.2024*

5 Concluding remarks and further research

This study aimed to provide preliminary empirical evidence on the application of forecasting models to a significant group of financial assets, assessing the accuracy of econometric, machine learning, and alternative forecast combination methods. In addition to evaluating their predictive power, we compared the estimated values from the most accurate models with futures market prices to gain insights into the predictability of financial assets from the perspective of market efficiency within the EMH framework and the potential existence of arbitrage opportunities in portfolio management investment strategies.

Firstly, the findings, though preliminary, support the use of ensemble methods in modelling financial time series. The methodologies applied in this study confirm the advantages of forecast combination for reducing prediction errors, consistent with empirical literature spanning over fifty years. The evidence suggests that machine learning methods generally outperform traditional econometric models in forecasting asset classes, and that more complex algorithms based on regularization and optimized model weighting offer the highest accuracy across various horizons tested, even surpassing simple combinations. This also contributes to the discussion of the forecast combination puzzle in literature.

Additionally, we sought to interpret the results in light of theories on market (in)efficiency, particularly regarding the predictability of financial asset prices. The preliminary findings of this study align with the notion of arbitrage opportunities between spot and futures market prices, as the most effective forecasting models appear to predict ETF index values more accurately than market pricing in derivatives. This generally supports prior studies arguing for market inefficiencies due to various factors, such as information asymmetry, transaction costs, institutional market power, among others.

Finally, the results suggest potential areas for further development. For instance, additional robustness tests and structural break analyses could be applied, particularly to some traditional econometric models. Moreover, machine learning algorithms could benefit from refined parameterization processes to assess sensitivity to hyperparameter tuning, particularly in Ridge Regression and G-LASSO. Testing more flexible moving window forecast horizons, beyond the three evaluated here, could also enhance the analysis.

References

1. Bates, J. M., & Granger, C. W. J. (1969). The Combination of Forecasts. Journal of the Operational Research Society, 20(4), 451–468.
2. Box, G. E. P. ., & Jenkins, G. M. . (1970). Time series analysis; forecasting and control. Holden-Day. https://books.google.com/books/about/Time_Series_Analysis.html?hl=pt-BR&id=5BVfnXaq03oC
3. Box, T., Davis, R., Evans, R., & Lynch, A. (2021). Intraday arbitrage between ETFs and their underlying portfolios. Journal of Financial Economics, 141(3). https://doi.org/10.1016/j.jfineco.2021.04.023

4. Bravo, J. M. (2023, September). Ensemble methods for Stock Market Prediction. MIDAS 8th Workshop on MIning DAta for Financial ApplicationS.
5. Bravo, J. M., & Ayuso, M. (2021). Forecasting the Retirement Age: A Bayesian Model Ensemble Approach. Advances in Intelligent Systems and Computing, 1365 AIST, 123–135. https://doi.org/10.1007/978-3-030-72657-7_12
6. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS
7. Brown, D. C., Davies, S. W., & Ringgenberg, M. C. (2021). ETF Arbitrage, Non-Fundamental Demand, and Return Predictability. Review of Finance, 25(4). https://doi.org/10.1093/rof/rfaa027
8. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016, 785–794. https://doi.org/10.1145/2939672.2939785/SUPPL_FILE/KDD2016_CHEN_BOOSTING_SYSTEM_01-ACM.MP4
9. Claeskens, G., Magnus, J. R., Vasnev, A. L., & Wang, W. (2016). The forecast combination puzzle: A simple theoretical explanation. International Journal of Forecasting, 32(3), 754–762. https://doi.org/10.1016/j.ijforecast.2015.12.005
10. Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, 5(4), 559–583. https://doi.org/10.1016/0169-2070(89)90012-5
11. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning 1995 20:3, 20(3), 273–297. https://doi.org/10.1007/BF00994018
12. Day, M. Y., & Lin, J. T. (2019). Artificial intelligence for ETF market prediction and portfolio optimization. Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2019, 1026–1033. https://doi.org/10.1145/3341161.3344822
13. Diaw, A. (2024). The Long Way to Cryptocurrencies Commoditization: Learning from Bitcoin ETF Prices? The Journal of Beta Investment Strategies, jbis.2024.1.061. https://doi.org/10.3905/jbis.2024.1.061
14. DiLellio, J., & Stanley, D. (2011). ETF trading strategies to enhance client wealth maximization. Financial Services Review, 20(2), 145–163.
15. Emekter, R., Jirasakuldech, B., & Beaves, R. (2022). Efficient Asset Allocation for Individual Investors in the ETF World. Financial Planning Research Journal, 8(1), 79–98. https://doi.org/10.2478/fprj-2022-0004
16. Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance, 25(2), 383–417.
17. Fama, E. F. (1991). Efficient Capital Markets: II. The Journal of Finance, 46(5), 1575–1617. https://doi.org/10.1111/J.1540-6261.1991.TB04636.X
18. Gowani, R., & Kanjiani, Z. (2024). Advanced LSTM Neural Networks for Predicting Directional Changes in Sector-Specific ETFs Using Machine Learning Techniques.
19. Granger, C. W. J. (1989). Invited Review Combining Forecasts-Twenty Years Later. Journal of Forecasting, 8(3), 167–173.
20. Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. American Economic Review, 70(3), 393–408.
21. Herzog, B., Jones, J., & Safaee, S. (2023). Remember to Diversify Your Active Risk: Evidence from US Equity ETFs. Journal of Beta Investment Strategies, 14(2), 6–16. https://doi.org/10.3905/jbis.2023.1.033

22. Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12(1), 55. https://doi.org/10.2307/1267351
23. Hsiao, C., & Wan, S. K. (2014). Is there an optimal forecast combination? Journal of Econometrics, 178(PART 2), 294–309. https://doi.org/10.1016/j.jeconom.2013.11.003
24. Hyndman, R., Koehler, A., Ord, K., & Snyder, R. (2008). Forecasting with exponential smoothing: the state space approach. In Springer Science & Business Media. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71918-2
25. Lee, T.-H., & Seregina, E. (2022). Combining Forecasts under Structural Breaks Using Graphical LASSO. https://arxiv.org/abs/2209.01697v2
26. Lo, A. W., & MacKinlay, A. C. (1988). Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test. The Review of Financial Studies, 1(1), 41–66. https://doi.org/10.1093/RFS/1.1.41
27. Mackinlay, A. C., & Ramaswamy, K. (1988). Index-Futures Arbitrage and the Behavior of Stock Index Futures Prices. Review of Financial Studies, 1(2), 137–158. https://www.jstor.org/stable/2962006
28. Madhavan, A. N. (2016). Exchange-Traded Funds and the New Dynamics of Investing. In Exchange-Traded Funds and the New Dynamics of Investing. https://doi.org/10.1093/acprof:oso/9780190279394.001.0001
29. Marshall, B. R., Nguyen, N. H., & Visaltanachoti, N. (2013). ETF arbitrage: Intraday evidence. Journal of Banking and Finance, 37(9). https://doi.org/10.1016/j.jbankfin.2013.05.014
30. Modest, D. M., & Sundaresan, M. (1983). The Relationship Between Spot and Futures Prices in Stock Index Futures Markets: Some Preliminary Evidence. Journal of Futures Markets, 3.
31. Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. Omega, 33(6), 497–505. https://doi.org/10.1016/J.OMEGA.2004.07.024
32. Piovezan, R. P. B., de Andrade Junior, P. P., & Ávila, S. L. (2024). Machine Learning Method for Return Direction Forecast of Exchange Traded Funds (ETFs) Using Classification and Regression Models. Computational Economics, 63(5). https://doi.org/10.1007/s10614-023-10385-4
33. Rompotis, G. (2024). The Performance of Commodity ETFs in the United States. Journal of Beta Investment Strategies, 15(2).
34. Saha, K., Madhavan, V., & Chandrashekhar, G. R. (2021). Relative efficiency of equity ETFs: an adaptive market hypothesis perspective. Applied Economics Letters, 28(14). https://doi.org/10.1080/13504851.2020.1804045
35. Samuelson, P. A. (1965). Proof That Properly Anticipated Prices Fluctuate Randomly. Industrial Management Review, 6(2), 41–49.
36. Sheng, Y., & Ma, D. (2022). Stock Index Spot–Futures Arbitrage Prediction Using Machine Learning Models. Entropy 2022, Vol. 24, Page 1462, 24(10), 1462. https://doi.org/10.3390/E24101462
37. Smith, J., & Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. Oxford Bulletin of Economics and Statistics, 71(3), 331–355. https://doi.org/10.1111/j.1468-0084.2008.00541.x
38. Stoll, H. R., & Whaley, R. E. (1990). The Dynamics of Stock Index and Stock Index Futures Returns. Journal of Financial and Quantitative Analysis, 25(4), 441–468.
39. Timmermann, A. (2006). Chapter 4 Forecast Combinations. Handbook of Economic Forecasting, 1, 135–196. https://doi.org/10.1016/S1574-0706(05)01004-9

40. Tsai, C. F., Hsu, Y. F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. Applied Soft Computing Journal, 24, 977–984. https://doi.org/10.1016/j.asoc.2014.08.047
41. Tsai, C. F., Lin, Y. C., Yen, D. C., & Chen, Y. M. (2011). Predicting stock returns by classifier ensembles. Applied Soft Computing Journal, 11(2), 2452–2459. https://doi.org/10.1016/j.asoc.2010.10.001
42. Wang, X., Hyndman, R. J., Li, F., & Kang, Y. (2023). Forecast combinations: An over 50-year review. In International Journal of Forecasting (Vol. 39, Issue 4, pp. 1518–1547). Elsevier B.V. https://doi.org/10.1016/j.ijforecast.2022.11.005
43. Zhang, P. G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

Table 2. MAPE and RMSE performance evaluation for each forecast horizon
- ranked by forecasting model (red represents higher errors and blue represents lower errors)

Forecasting Horizon	Model	S&P 500 (SPY)		Russell 2000 (IWM)		MSCI Developed (EFA)		MSCI Emerging (EEM)		US Treasury Bonds (IEF)		Developed Bonds (IGOV)		Emerging Bonds (EMB)		US Corporate Bonds (LQD)		Gold (GLD)		Bitcoin (BITO)	
		MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
1 month	ARIMA	1.280	8.928	1.903	5.419	1.437	1.459	2.443	1.359	0.792	0.877	0.588	0.299	0.778	0.888	0.849	1.064	1.971	6.056	10.579	2.133
	ETS	1.280	8.926	1.909	5.463	1.374	1.393	2.446	1.360	0.779	0.860	0.585	0.298	0.778	0.888	0.870	1.093	2.101	6.587	10.415	2.090
	SVM	2.988	17.913	2.332	6.367	1.499	1.460	3.662	1.704	2.980	2.974	1.752	0.756	0.494	0.574	1.919	2.257	4.029	10.646	7.856	1.932
	RF	1.359	9.107	1.948	5.611	1.223	1.241	2.449	1.360	0.768	0.847	0.576	0.293	0.767	0.863	0.866	1.087	2.143	7.059	11.251	2.226
	XGBoost	1.366	9.085	1.896	5.353	1.209	1.218	2.386	1.357	0.875	0.971	0.614	0.314	0.802	0.934	0.938	1.194	2.191	7.286	10.792	2.151
	ARIMA-SVR	1.331	9.314	1.931	5.549	1.567	1.567	2.519	1.367	0.710	0.781	0.576	0.292	0.748	0.816	0.796	0.991	1.957	5.886	10.009	2.043
	ARIMA-ANN	1.280	8.929	1.905	5.438	1.442	1.464	2.444	1.359	0.799	0.886	0.589	0.300	0.781	0.894	0.851	1.067	1.970	6.033	10.567	2.131
	Simple Average	1.327	9.348	1.909	5.541	1.365	1.380	2.539	1.359	1.040	1.137	0.671	0.333	0.732	0.833	0.953	1.217	2.063	6.862	8.634	1.831
	Ridge Regression	0.872	6.224	1.606	4.209	1.049	1.048	1.862	0.943	0.400	0.485	0.264	0.136	0.331	0.364	0.337	0.470	0.762	2.169	4.167	0.865
	G-LASSO	0.866	6.210	1.574	4.193	1.032	1.038	1.853	0.946	0.432	0.537	0.266	0.131	0.334	0.363	0.339	0.471	0.768	2.175	3.975	0.842
1 quarter	ARIMA	1.755	11.793	5.875	14.761	2.555	2.476	2.030	1.227	2.284	2.670	3.945	1.894	1.921	2.252	2.195	3.042	5.243	14.276	25.670	5.920
	ETS	1.780	11.922	5.858	14.730	2.629	2.542	2.028	1.226	2.304	2.694	3.943	1.894	1.921	2.252	2.223	3.079	5.850	15.865	17.279	3.764
	SVM	3.755	24.380	3.497	8.476	4.422	3.875	3.530	1.757	2.752	3.125	1.664	0.811	1.525	1.586	1.283	1.632	4.929	12.765	18.703	3.945
	RF	2.165	14.401	5.838	14.692	2.765	2.667	1.977	1.194	2.355	2.750	3.936	1.890	1.931	2.262	2.227	3.084	5.526	15.198	16.716	3.658
	XGBoost	2.282	15.054	5.711	14.449	2.641	2.554	1.998	1.207	2.434	2.839	3.939	1.892	2.029	2.355	2.273	3.140	5.544	15.236	17.548	3.815
	ARIMA-SVR	1.705	11.603	5.771	14.561	2.472	2.399	1.996	1.206	2.220	2.590	3.933	1.890	1.860	2.188	2.139	2.966	5.082	13.942	26.015	5.995
	ARIMA-ANN	1.746	11.757	5.870	14.751	2.548	2.471	2.027	1.225	2.288	2.675	3.952	1.897	1.943	2.273	2.198	3.046	5.243	14.275	25.676	5.921
	Simple Average	1.652	11.370	4.763	12.400	2.115	2.062	1.905	1.129	2.363	2.759	3.398	1.652	1.609	1.890	2.038	2.826	5.341	14.493	6.659	1.657
	Ridge Regression	1.601	11.229	2.687	6.691	1.526	1.622	1.939	1.124	0.535	0.623	2.224	1.011	0.474	0.536	0.462	0.609	1.072	3.115	5.954	1.429
	G-LASSO	1.601	11.229	2.682	6.685	1.466	1.518	1.914	1.094	0.498	0.594	0.472	0.237	0.463	0.520	0.455	0.599	1.071	3.105	5.943	1.430
1 year	ARIMA	12.721	75.131	11.783	27.041	10.602	9.268	7.471	3.677	3.207	3.653	6.301	2.819	6.673	6.510	5.385	6.455	14.415	36.423	64.983	16.499
	ETS	15.006	89.167	11.761	26.992	10.505	9.190	7.427	3.660	3.207	3.653	6.280	2.811	6.525	6.366	5.332	6.393	14.415	36.422	37.091	10.106
	SVM	2.735	17.606	8.622	19.310	4.782	4.266	15.226	6.724	2.774	3.081	9.572	4.277	6.327	6.230	3.803	4.599	8.372	24.179	22.773	6.344
	RF	14.634	87.300	11.660	26.757	10.045	8.817	7.139	3.544	2.893	3.345	5.744	2.591	6.087	5.933	4.935	5.911	13.805	35.261	36.840	10.058
	XGBoost	14.786	88.070	11.619	26.661	10.050	8.822	6.980	3.478	2.925	3.378	5.810	2.619	6.157	6.003	5.008	5.999	13.892	35.427	36.855	10.061
	ARIMA-SVR	12.674	74.844	11.947	27.427	10.653	9.314	7.481	3.681	3.262	3.706	6.373	2.848	6.672	6.507	5.439	6.520	14.468	36.547	64.833	16.466
	ARIMA-ANN	12.718	75.113	11.779	27.031	10.601	9.267	7.474	3.679	3.213	3.659	6.309	2.823	6.683	6.519	5.391	6.463	14.402	36.398	64.987	16.500
	Simple Average	11.594	68.613	9.312	21.163	8.480	7.395	4.485	2.272	2.548	2.962	4.087	1.911	4.862	4.669	4.227	5.076	13.341	34.284	45.758	11.961
	Ridge Regression	1.953	12.181	3.345	7.759	2.284	2.186	1.655	0.893	1.666	1.812	2.444	1.261	1.475	1.581	2.000	2.554	2.204	5.666	16.173	4.330
	G-LASSO	1.899	11.950	3.330	7.714	1.682	1.597	1.656	0.882	1.666	1.811	2.298	1.061	1.409	1.486	2.001	2.560	2.070	5.276	15.012	4.113

Table 3. MAPE and RMSE performance evaluation for each forecast horizon
- ranked by asset class (red represents higher errors and blue represents lower errors)

Forecasting Horizon	Model	S&P 500 (SPY) MAPE	S&P 500 (SPY) RMSE	Russell 2000 (IWM) MAPE	Russell 2000 (IWM) RMSE	MSCI Developed (EFA) MAPE	MSCI Developed (EFA) RMSE	MSCI Emerging (EEM) MAPE	MSCI Emerging (EEM) RMSE	US Treasury Bonds (IEF) MAPE	US Treasury Bonds (IEF) RMSE	Developed Bonds (IGOV) MAPE	Developed Bonds (IGOV) RMSE	Emerging Bonds (EMB) MAPE	Emerging Bonds (EMB) RMSE	US Corporate Bonds (LQD) MAPE	US Corporate Bonds (LQD) RMSE	Gold (GLD) MAPE	Gold (GLD) RMSE	Bitcoin (BITO) MAPE	Bitcoin (BITO) RMSE
1 month	ARIMA	1.280	8.928	1.903	5.419	1.437	1.459	2.443	1.359	0.792	0.877	0.588	0.299	0.778	0.888	0.849	1.064	1.971	6.056	10.579	2.133
	ETS	1.280	8.926	1.909	5.463	1.374	1.393	2.446	1.360	0.779	0.860	0.585	0.298	0.778	0.888	0.870	1.093	2.101	6.587	10.415	2.090
	SVM	2.988	17.913	2.332	6.367	1.499	1.460	3.662	1.704	2.980	2.974	1.752	0.756	0.494	0.574	1.919	2.257	4.029	10.646	7.856	1.932
	RF	1.359	9.107	1.948	5.611	1.223	1.241	2.449	1.360	0.768	0.847	0.576	0.293	0.767	0.863	0.866	1.087	2.143	7.059	11.251	2.226
	XGBoost	1.366	9.085	1.896	5.353	1.209	1.218	2.386	1.357	0.875	0.971	0.614	0.314	0.802	0.934	0.938	1.194	2.191	7.286	10.792	2.151
	ARIMA-SVR	1.331	9.314	1.931	5.549	1.567	1.567	2.519	1.367	0.710	0.781	0.576	0.292	0.748	0.816	0.796	0.991	1.957	5.886	10.009	2.043
	ARIMA-ANN	1.280	8.929	1.905	5.438	1.442	1.464	2.444	1.359	0.799	0.886	0.589	0.300	0.781	0.894	0.851	1.067	1.970	6.033	10.567	2.131
	Simple Average	1.327	9.348	1.909	5.541	1.365	1.380	2.539	1.359	1.040	1.137	0.671	0.333	0.732	0.833	0.953	1.217	2.063	6.862	8.634	1.831
	Ridge Regression	0.872	6.224	1.606	4.209	1.049	1.048	1.862	0.943	0.400	0.485	0.264	0.136	0.331	0.364	0.337	0.470	0.762	2.169	4.167	0.865
	G-LASSO	0.866	6.210	1.574	4.193	1.032	1.038	1.853	0.946	0.432	0.537	0.266	0.131	0.334	0.363	0.339	0.471	0.768	2.175	3.975	0.842
1 quarter	ARIMA	1.755	11.793	5.875	14.761	2.555	2.476	2.030	1.227	2.284	2.670	3.945	1.894	1.921	2.252	2.195	3.042	5.243	14.276	25.670	5.920
	ETS	1.780	11.922	5.858	14.730	2.629	2.542	2.028	1.226	2.304	2.694	3.943	1.894	1.921	2.252	2.223	3.079	5.850	15.865	17.279	3.764
	SVM	3.755	24.380	3.497	8.476	4.422	3.875	3.530	1.757	2.752	3.125	1.664	0.811	1.525	1.586	1.283	1.632	4.929	12.765	18.703	3.945
	RF	2.165	14.401	5.838	14.692	2.765	2.667	1.977	1.194	2.355	2.750	3.936	1.890	1.931	2.262	2.227	3.084	5.526	15.198	16.716	3.658
	XGBoost	2.282	15.054	5.711	14.449	2.641	2.554	1.998	1.207	2.434	2.839	3.939	1.892	2.029	2.355	2.273	3.140	5.544	15.236	17.548	3.815
	ARIMA-SVR	1.705	11.603	5.771	14.561	2.472	2.399	1.996	1.206	2.220	2.590	3.933	1.890	1.860	2.188	2.139	2.966	5.082	13.942	26.015	5.995
	ARIMA-ANN	1.746	11.757	5.870	14.751	2.548	2.471	2.027	1.225	2.288	2.675	3.952	1.897	1.943	2.273	2.198	3.046	5.243	14.275	25.676	5.921
	Simple Average	1.652	11.370	4.763	12.400	2.115	2.062	1.905	1.129	2.363	2.759	3.398	1.652	1.609	1.890	2.038	2.826	5.341	14.493	6.659	1.657
	Ridge Regression	1.601	11.229	2.687	6.691	1.526	1.622	1.939	1.124	0.535	0.623	2.224	1.011	0.474	0.536	0.462	0.609	1.072	3.115	5.954	1.429
	G-LASSO	1.601	11.229	2.682	6.685	1.466	1.518	1.914	1.094	0.498	0.594	0.472	0.237	0.463	0.520	0.455	0.599	1.071	3.105	5.943	1.430
1 year	ARIMA	12.721	75.131	11.783	27.041	10.602	9.268	7.471	3.677	3.207	3.653	6.301	2.819	6.673	6.510	5.385	6.455	14.415	36.423	64.983	16.499
	ETS	15.006	89.167	11.761	26.992	10.505	9.190	7.427	3.660	3.207	3.653	6.280	2.811	6.525	6.366	5.332	6.393	14.415	36.422	37.091	10.106
	SVM	2.735	17.606	8.622	19.310	4.782	4.266	15.226	6.724	2.774	3.081	9.572	4.277	6.327	6.230	3.803	4.599	8.372	24.179	22.773	6.344
	RF	14.634	87.300	11.660	26.757	10.045	8.817	7.139	3.544	2.893	3.345	5.744	2.591	6.087	5.933	4.935	5.911	13.805	35.261	36.840	10.058
	XGBoost	14.786	88.070	11.619	26.661	10.050	8.822	6.980	3.478	2.925	3.378	5.810	2.619	6.157	6.003	5.008	5.999	13.892	35.427	36.855	10.061
	ARIMA-SVR	12.674	74.844	11.947	27.427	10.653	9.314	7.481	3.681	3.262	3.706	6.373	2.848	6.672	6.507	5.439	6.520	14.468	36.547	64.833	16.466
	ARIMA-ANN	12.718	75.113	11.779	27.031	10.601	9.267	7.474	3.679	3.213	3.659	6.309	2.823	6.683	6.519	5.391	6.463	14.402	36.398	64.987	16.500
	Simple Average	11.594	68.613	9.312	21.163	8.480	7.395	4.485	2.272	2.548	2.962	4.087	1.911	4.862	4.669	4.227	5.076	13.341	34.284	45.758	11.961
	Ridge Regression	1.953	12.181	3.345	7.759	2.284	2.186	1.655	0.893	1.666	1.812	2.444	1.261	1.475	1.581	2.000	2.554	2.204	5.666	16.173	4.330
	G-LASSO	1.899	11.950	3.330	7.714	1.682	1.597	1.656	0.882	1.666	1.811	2.298	1.061	1.409	1.486	2.001	2.560	2.070	5.276	15.012	4.113