https://doi.org/10.26219/heal.aueb.10223

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Forecasting Macroeconomic Series Using Advanced Econometric Models

By

Evangelos {Christos}. Ntaflos

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfilment of the requirements for

the degree of Master in Quantitative Actuarial

and Financial Risk Management

Athens, Greece

March 2026

Chapter 1. Introduction

The investigation of the factors that affect macroeconomics has always been the objective of economic research, as well as the ways that they can be practically executed in economic policy. Some of the factors that are quite important are output growth, inflation, unemployment, and the financial conditions of each country, which are very important for policy institutions, financial markets, and private sector decision makers. For the central banks, the role of the central banks is mostly to guide monetary policy, while governments are responsible for fiscal interventions. Market participants rely on information regarding risk assessment and capital allocation. So many researchers focus on the ways to best model, predict, and generally translate the fluctuations of the macroeconomy.

For decades, and especially during the last years, empirical research of macroeconomics and the factors that affect it has changed substantially. This change happened mostly because of the widely available and high-frequency financial data, which, combined with the increased granular macroeconomic indicators, led to more information being available to researchers in comparison to previous years. This information gives the chance to financial markets to create a stream of data which represents risk perceptions, forward-looking assessments, but also expectations of economic conditions. Meanwhile, databases regarding macroeconomic systems and regulations have grown in breadth and depth due to the addition of indicators that represent labor markets, prices, production, sentiment, and even uncertainty. New developments, combined with increases in computational power, have led to the application of more effective and, at the same time, more sophisticated modeling techniques, which can handle high-dimensional, non-linear, and even mixed-frequency data bases.

These developments have offered a wider choice of modeling approaches. Traditional econometric methods, like autoregressive time series models and multivariate systems, are still viewed as important due to their high interpretability and the theoretical foundations that they were built upon. Although sometimes the assumptions of linearity, parameter stability, and even their relatively small information basis lead to many challenges, due to the complexity of the macrofinancial data that have been provided during the last years. However, machine learning methods seem to have gained more popularity, since they are more flexible and are able to capture non-linear relations between variables, but also are able to create more complex interactions among variables. At the same time, they are more durable to structural changes that classic methods find difficult to process. These methods, even if they were originally created for computer science and statistics purposes, have been used more and more in economic fields and offer important information regarding their predictive accuracy in certain fields and subjects.

Even if modeling techniques offer important benefits, the growing number of them leads to some critical and possible problems analytically. Many researchers believe that a single model cannot outperform when it comes to empirical applications. It is quite important to understand that each model can behave differently across different economic fields, the predictors that lead their outcomes, and the conditions that improve their accuracy. Also, it is very important to understand the purpose for which each one of the models is used. For example, a framework that focuses on predicting GDP growth would not be as appropriate for predicting inflation, explaining labor market turning points, or even forecasting financial asset prices. This shows that the effectiveness of each model varies not only based on statistical performance, but also on interpretability, robustness, and the alignment with the aims under which it is used, as well as the ways in which it is applied.

These considerations are particularly important when it comes to structural breaks, regime changes, and periods that are characterized by a high level of uncertainty, such as financial crises or when policy shifts change abruptly over time. Purely linear and static models present many limitations when used in such cases. However, the use of machine learning models, which are deemed as more flexible, leads to new challenges, especially regarding overfitting, the lack of transparency, and difficulties when it comes to the interpretation of the conclusions. Recent research seems to have mostly focused not only on evaluating the methods used, whether econometric or machine learning methods, but also on creating frameworks that aim to combine both the basic theoretical structure of economic models with the flexibility that machine learning approaches offer and their predictive abilities.

This background makes it clear that the objective of the present literature review is to investigate and offer a structured synthesis of the recent contributions to macroeconomic forecasting. More specifically, instead of focusing on a specific series, survey, or research methodology, the review highlights the important methodological developments and their implications that are directly connected to the modeling strategies presented in the research and proposed properly. Thus, the importance in the present research is mostly given to how different classes of models can handle information sets that are characterized by high dimensionality, non-linearity, and many times mixed frequencies, but also operate under real-time constraints. At the same time, the research focuses on the ways that these variables can be used as predictive variables in the field of macroeconomics.

To achieve the above, the present literature review contains three different thematic sections. Through the first section, it investigates econometric methods, both traditional time series models and factor-based models, in order to present their benefits but also their limitations in the modern forecasting of economic environments. Following this, the second section investigates mostly machine learning methods, especially giving special attention to linear models, tree-based algorithms, and deep learning architectures, highlighting their predictive abilities but also issues of interpretability. In the third and last section, hybrid macrofinancial models are investigated, which have the ability to integrate economic structure with machine learning flexibility. This section investigates the current methods that researchers use in macroeconomic forecasting. All three sections provide a foundation for the development of modeling frameworks in recent years.

Chapter 2. Econometric Models

2.1 VAR, ARIMA and traditional time-series structures

The traditional econometric models have been widely used in microeconomic analysis and forecasting. Through them, many empirical researchers have focused on the use of auto-aggressive structures, like the univariate models. Some examples are ARIMA and Vector Autoregressions (VAR), which is a multivariate system. Additionally, some extensions were also used, like the dynamic factor models, in order to incorporate some latent components. All of the above methods were used under assumptions that the relationships that were examined were linear and mostly stable over time. So, the values of the variables of the past were assumed to provide the needed information regarding the possible future results.

Generally, when certain indicators are taken into account, like output, inflation, employment, and the broad financial conditions, the classic model methods were usually applied in the field of macroeconomics in order to forecast. Their popularity is related to advantages that they have, such as their connection to economic theories and the fact that the parameters can easily be interpreted. Also, many researchers find that their statistical results can be easily understood and explained to others. Furthermore, autoregressive structures also present cyclical behavior, since the system evolves over time. However, they do not include any transmission mechanisms that are of huge importance when it comes to the macroeconomic field. According to the above, ARIMA and VAR methods and models have been widely used, both for institutions in the economic field and for academic and other types of research.

Although in recent years researchers have highlighted that the predictive power of classic time series models is really sensitive to economic environment changes, especially when it comes to the global financial situation, since policies change constantly and institutional frameworks do not present much stability. López-Estrada et al. (2025) focus on the effectiveness of VAR methods in order to predict electricity prices when a price-regulated setting was taken into consideration. The results with the VAR model were effective, as it seemed able to represent lagged interdependencies between the variables during specific periods of time with high accuracy. However, the accuracy seemed to be lower during structural breaks. This reveals some important limitations of traditional VAR methods. Even if the models are characterized by high resilience when it comes to linear parameters, the same ability makes them not appropriate for environments with constant regime shifts or evolving economic events. The findings lead to the conclusion that non-linear approaches and methods that are able to adapt to changing conditions should be explored further.

In contrast, Li et al. (2025) highlighted that the use of classical econometric frameworks in order to investigate mixed-frequency data and changes in regimes can lead to some forecasting benefits. Their study focused on the MS-MIDAS-LASSO model, which was used in order to predict the volatility of the S&P 500, using a set of macrofinancial predictors that included different frequencies. High-frequency financial information was used in the regression, and rolling simulation schemes were applied, as well as regime-alternating algorithms, showing that mixed-frequency models could outperform, almost always, simple AR and ARIMA methods and models. Also, through the results, it was confirmed that structural breaks are not captured by the classical models that were used, especially during periods of time that are characterized by financial stress. Additionally, the Markov-switching MIDAS specification was revealed to be quite effective when it comes to identifying regimes that are affected by crises and important shifts in volatility, which indicates that the traditional models used in the econometric field should be improved in order to include modern data and the challenges of the present.

Additionally, the general ARIMA model is usually expressed as:

$$\Phi(L)(1-L)^d y_t = \Theta(L)\varepsilon_t,$$

where L denotes the lag operator, $\Phi(L)$ and $\Theta(L)$ represent the autoregressive and moving-average polynomials, respectively, d is the order of differencing required to achieve stationarity, and ε_t is a white-noise error term. This method is used in order to capture serial correlation and stochastic trends in one time series. However, the univariate benefits also limit the ability when it comes to the interpretation of more than one macroeconomic variable.

In order to overcome this limitation, the multivariate VAR systems were used as autoregressive frameworks in order to model several endogenous variables jointly:

$$Y_t = c + A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + \varepsilon_t,$$

where Y_t is a vector of endogenous variables and the coefficient matrices A_i capture lagged interdependencies across the system. These models seem to be more effective when it comes to representing more complex dynamic interactions than ARIMA models, which are used mostly in order to analyze transmission mechanisms and policy responses in macroeconomic environments.

Besides the mentioned advantages, VAR models also present some constraints. First, when the number of variables in the model increases, the dimensionality grows to a high level and many times leads to an explosion of parameters. This increase is also connected to a higher risk of overfitting and multicollinearity, especially when the sample size is small or even limited. Thus, VAR models, even if they provide an easy and intuitive framework, are not effective in high-dimensional settings, which makes clear the growing need for techniques that reduce dimensionality or alternative models that can be equally effective in high-dimensional environments.

2.2 Dynamic Factor Models and Macroeconomic Predictability

Through recent empirical research, it was revealed that traditional techniques used in the econometric field face difficulties in managing large datasets. The higher the dimensionality of the data is, the more classical models seem to rely on their fixed parameter structures. Thus, the possibility of overfitting and multicollinearity increases, and their out-of-sample performance is also reduced. This limitation is mostly noticeable when the variables used in modern datasets are not appropriately shaped by dimension reduction mechanisms, leading to poor forecasting results.

Bachmair and Schmitz (2025) conducted a comprehensive study of macrofinancial forecasting in the United States and focused on the use of unregularized VAR models and conventional factor-based methods, which presented poor results when large datasets of variables were directly used in the models. Through these results, it was revealed that those classic models present a high level of sensitivity when it comes to the population of variables, and that the indiscriminate inclusion of financial variables tends to create noise rather than improve predictive results. According to these findings, it was also observed that the accuracy of the forecasts was at a lower level when dimensionality increased beyond a certain scale. According to the above, there is strong empirical support that traditional econometric frameworks are not properly equipped to effectively manage high-dimensional macrofinancial datasets, and that there is a need for dimensional reduction or regularization techniques in order for them to produce proper results.

Those models focus on the information that is given from many different time series and can extract a relatively small number of factors. These factors are the ones that represent mostly the economic forces that jointly drive the co-movement of the vast majority of the rest of the macroeconomic indicators. Thus, Dynamic Factors Models (DFMs) can reduce dimensionality and, at the same time be able to preserve the important information, limiting the problems regarding forecasting.

In a static factor representation, the observed data vector X_t is expressed as:

$$X_t = \Lambda f_t + u_t,$$

where f_t denotes a low-dimensional vector of unobserved common factors, Λ is the factor loading matrix, and u_t represents idiosyncratic components that are specific to individual series. According to the above, it is also possible for the models to exclude systematic movements that are driven by economic forces from series-specific noise. Additionally, the dynamic evolution of the factors is modeled by the use of autoregressive or VAR processes, which also enables the models to hold a structured time series representation and, at the same time, avoid an explosion in parameters that usually occurs in VAR models.

One of the main advantages of the models is their ability to manage parsimony and complexity when it comes to information sets, since they focus on a small number of common factors. Thus, DFMs can lower the risk of overfitting and, at the same time, incorporate important information from a large number of indicators. These characteristics make those models really useful in the macrofinancial field, since there is a small number of latent economic factors, like aggregate demand, monetary conditions, and financial stress, that affect a large number of other observed variables.

When it comes to practical use, the models have been applied to many environments, especially in equity indices, interest rates, exchange rates, commodity prices, and financial indicators that are related to the macroeconomic situation. The role of the included factors is also important, since their well-structured nature gives the opportunity to incorporate a large amount of data without having limitations in interpretability, as the factors that are extracted are associated with the most important concepts of economics. Additionally, DFMs can provide a framework for forecasting, nowcasting, and even robustness for scenario analysis, when there is uncertainty in the availability of data across frequencies and publication lags.

However, even if DFMs can improve greatly through the use of under-regularized multivariate models, they also present some limitations that should be taken into consideration. First of all, since there is a linear structure of the basic DFMs, their power to recognize and analyze non-linear relationships and regime-dependent dynamics that are present in macrofinancial data is limited. Additionally, research suggests that when there is a small number of common factors, relevant information may not be represented effectively, especially under structural changes and financial crises. Thus, researchers had to explore and find models that present more flexibility when it comes to factors, leading to approaches that combine machine learning and hybrid frameworks, which are also investigated in the following chapter.

Chapter 3. Machine Learning Models

3.1 Regularized Linear Models (Ridge, Lasso, Elastic Net)

Linear models that are regularized, for example Ridge, Lasso, and Elastic Net, are tools that are widely used in forecasting with macrofinancial variables, especially in settings with high-dimensional predictor spaces. Ordinary least squares regression minimizes only the residual sum of squares, while regularized regressions focus on explicit penalty terms that control the final size of the model coefficients. This regularization is really important when it comes to financial and macroeconomic applications, since the number of predictors many times exceeds the sample size that is available, which leads to unstable results and poor performance of the models, especially under standard estimation methods.

The main goal of regularization is to be able to control the complexity of the model and, at the same time, retain its predictive power. Thus, when coefficient estimates are shrunk, the mentioned methods lead to lower variance and mitigate overfitting, which is a problem commonly presented in datasets of macrofinancial fields that present a high level of multicollinearity and noise. Thus, regularized linear models act as a natural bridge between econometric regression and more advanced machine learning approaches that are more flexible, have high interpretability, and, at the same time, improve robustness in environments characterized by high-dimensionality.

Ridge regression introduces an ℓ_2 penalty on the magnitude of regression coefficients and can be expressed as

$$\min_{\beta} \sum_{t} (y_t - X_t^{\top}\beta)^2 + \lambda \sum_{j} \beta_j^2 \,,$$

where λ controls the strength of regularisation. This formulation shrinks all coefficients toward zero in a continuous manner, effectively stabilising estimation when predictors are highly correlated. Despite the fact that ridge regression does not force coefficients exactly to zero, it is nevertheless quite effective when there is a large number of predictors that carry small but non-negligible information. This characteristic is very important, especially when economic channels are dispersed across multiple indicators and are not concentrated in a small number of variables used in the models.

In contrast, the Least Absolute Shrinkage and Selection Operator (Lasso) imposes an ℓ_1 penalty:

$$\min_{\beta} \sum_{t} (y_t - X_t^{\top}\beta)^2 + \lambda \sum_{j} |\beta_j|.$$

The above technique is quite effective regarding sparsity and the shrinkage of some coefficients exactly to zero. Thus, the model is able to perform automatic variable selection. Lasso is more commonly used when researchers have to manage a large number of potential predictors and suspect that only some of them are truly relevant to their purpose. So, they select a parsimonious model like Lasso in order to ensure a high level of interpretability and, at the same time, reduce possible uncertainty, which makes it an attractive tool in forecasting but also in economic inference areas.

Although the Lasso model also presents some important limitations, especially when the predictors present a high level of correlation between them, which is quite common when it comes to macroeconomic and financial variables. When this happens, Lasso can select one variable from a group of correlated variables and remove the others, which can potentially lead to an unstable model. Elastic Net addresses this issue by combining ℓ_1 and ℓ_2 penalties, effectively blending the properties of Ridge and Lasso:

$$\min_{\beta} \left\{ \sum_{t=1}^{n} (y_t - X_t^{T}\beta)^2 + \lambda \left(\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^2 \right) \right\}$$

where λ controls the overall regularization strength, while $\alpha \in [0,1]$ determines the balance between the Lasso (L1) and Ridge (L2) penalties.

The hybrid model that uses Elastic Net combines variable selection and, at the same time, shrinks the coefficient estimates, especially when there is a high level of multicollinearity. Thus, Elastic Net is commonly used in macrofinancial datasets in which economic indicators often move together due to shared underlying causes.

Concluding the above, regularized linear models can be as effective as machine learning models, especially when it comes to economic forecasting. However, they have clear advantages in high-dimensional settings, while maintaining transparency and proper interpretability, which is often not present in complex machine learning algorithms. Meanwhile, the linear structure leads to limitations when it comes to the ability to capture non-linear relationships and regime-dependent dynamic patterns. Thus, researchers were forced to investigate more flexible approaches, like tree-based methods and deep learning architectures, which are presented in the following chapters.

3.2 Tree-Based Models (Random Forest, XGBoost)

Tree-based learning algorithms have been used more and more in the field of economic and financial forecasting, since they offer the ability to investigate non-linear relationships in the reaction function and help with complex decision boundaries, contrary to linear models. Tree-based methods do not rely on parametric assumptions on the functional form linking predictors to the outcome variables. Through these characteristics, the models can effectively handle changes, threshold effects, and asymmetric responses, which are quite frequent in the data sets of macroeconomic and financial environments. Additionally, their effectiveness and performance do not decrease as the number of predictors increases, so they are quite suitable for high-dimensional forecasting environments.

More specifically, Random Forests have emerged as a powerful learning technique that is created through the aggregation of many decision trees. Analytically, this is done through averaging predictions across a large number of trees grown on bootstrap samples and random subsets of predictors. These models lead to a lower level of bias and improved generalization results. Empirical findings show that Random Forests can manage effectively data with a high level of noise and multicollinearity, which are quite common in macrofinancial environments. Thus, these models often produce more accurate forecasting results in comparison to traditional linear models like ARIMA, especially when the data-generating process contains non-linear correlations and relationships.

Nti et al. (2019) provided an illustrative empirical application by applying Random Forest in order to predict stock market prices in the Ghana Stock Exchange. In the study, Random Forest methods were used with variables and importance measures in order to analyze macroeconomic indicators that can predict most strongly stock prices. From the results, it was made clear that Random Forest models can produce more effective predictions compared to ARIMA benchmarks, especially when it comes to forecasting accuracy. At the same time, they offer insights into the importance of each predictor separately. Thus, the variable importance structure differs from the traditional framework based on economic interpretation. This also led the researchers to connect predictive power to macroeconomic mechanisms, rather than using models just as black-box tools.

Besides Random Forest methods, gradient boosting algorithms like XGBoost have gained popularity, mostly due to their performance when it comes to predictive accuracy and their efficiency in computational applications. This specific building strategy is sequential, where each new tree focuses on correcting the errors of the previous ones. This strategy leads to the model having high complexity in order to represent patterns in the data. When it comes to macrofinancial forecasting, these models have been shown to perform quite well, especially in environments that present volatility clustering, regime changes, and many different predictor effects.

In the research of Neghab et al. (2025), an example of the application of tree-based methods in exchange rate forecasting was presented. Analytically, XGBoost and SHAP (Shapley Additive Explanations) values were used in order to investigate the contribution of individual predictors to the USD/CAD exchange rate movements. The results highlighted that the prices of oil and gold seemed to heavily affect the exchange rate dynamics, while also reflecting the commodity nature of the Canadian economy, which is linked to commodities. Additionally, indicators related to financial stress seemed to increase in importance, especially during periods of time with high levels of volatility. At the same time, the linear nature of macrofinancial variable relationships was not as important. The use of SHAP values allowed these findings to be presented as proper marginal contributions for each predictor individually. In this way, the transparency of the data and the results was enhanced.

According to the above results, it is clear that tree-based models can be more balanced when it comes to predictive accuracy and the economic insight they offer. Additionally, they are able to manage effectively non-linear relationships and high-dimensional predictor spaces, which makes them more effective in comparison to traditional econometric models. Furthermore, feature selection techniques like variable importance rankings and SHAP help with the interpretability of the results. According to the above, tree-based machine learning models seem to be a strong toolbox for macrofinancial forecasting, especially when the objective is robust prediction and aims to include the indicators that lead economic and financial environments to certain outcomes.

3.3 Deep Learning and LSTM Architectures

Recurrent neural networks, but also more advanced methods like long short-term memory networks, seem to gain more and more attention in economic and financial fields, especially in forecasting research, since they have the power to model sequential dependencies and capture temporal patterns. In comparison to traditional linear or time series models, RNN-based architectures are designed in order to manage ordered data and, at the same time, represent information from previous observations through their memory. This characteristic is essential, especially when it comes to financial and macroeconomic time series, since they are characterized by persistence, delayed effects, and many complex temporal relationships between variables.

Although standard RNNs present some specific limitations that are mostly connected to exploding gradients, which can lower their power to learn long-term dependencies, Long Short-Term Memory (LSTM) models focus on these issues by using gated memory cells that have the power to manage the flow of information through input, output, and forget gates. As an outcome, LSTMs are more able to represent short-term fluctuations but also longer-term dependencies between variables, like trends, cycles, or hidden patterns, which are quite common in the economic field.

Even though they are quite flexible, deep learning models can face certain challenges, especially in high-dimensional forecasting environments. More specifically, in finance, where open datasets are made of large numbers of technical indicators, sentiment measures, and macroeconomic variables, which are often correlated or even informative, without the right constraints neural networks may present overfitting on the data. This leads to poor generalization of the results that they produce.

In their research, Yang et al. (2022) investigated this specification by using a Lasso-regularized LSTM framework in order to predict possible stock price directions. Their method uses both technical indicators that are taken from previous price data and sentiment indicators that were extracted from specific financial texts. Thus, a version of Lasso-based variable selection is embedded in the deep learning architecture. Through this, they managed to reduce the level of dimensionality and, at the same time, the level of multicollinearity. Their results made it clear that the Lasso-LSTM model achieved a higher level of out-of-sample accuracy in comparison to the standard LSTM. This also highlights how important feature selection is, even when neural network models are quite flexible. Thus, this makes clear that insights from the literature also converge on the idea that different methods can benefit from structured regularization, especially when applied to economic and financial forecasting problems.

In conclusion, LSTM-based models seem to be powerful tools when it comes to specific model designs that include feature selection and out-of-sample validation. These considerations should be mentioned, especially when deep learning architectures are used in a complementary way and not as substitutes for other machine learning and econometric methods.

3.4 Machine Learning in Macroeconomic Nowcasting

Besides traditional econometric applications, machine learning methods have been used quite frequently in macroeconomic nowcasting, where the main goal is to investigate economic conditions in real time by using asynchronously released data. Nowcasting is a challenging task, since macroeconomic indicators are often published with lags, are subject to revisions, and many times differ in frequency and timing in comparison to financial data. Thus, traditional econometric models are many times not able to accommodate real-time data effectively, especially in environments that present rapid economic changes.

Mamedli and Shibitov (2021) presented a detailed analysis regarding the use of machine learning methods in order to investigate real-time forecasting of the Russian Consumer Price Index. They showed that data vintages, publication lags, and seasonal adjustments allow a better management of real-time information, which also affects the forecasting results. Additionally, they highlighted that forecast accuracy can present non-linear behavior and can be quite sensitive to the timing of data releases, but also to the treatment of seasonality by modern models.

Supporting one of the most important contributions regarding the analysis is the comparison that took place between models estimated using revised data and those estimated under real-time conditions. Through the results, it was highlighted that when publication lags are ignored, as well as data revisions, there is a tendency to forecast over-optimistically, which offers an overly optimistic assessment regarding model performance. On the contrary, machine learning models that use real-time data seem to achieve more accurate results compared to traditional methods. This advantage is even more evident when there are episodes of high inflation volatility, which is especially important when information is used for policy decision-making in the economy.

These results highlight the importance of machine learning contribution to macroeconomic forecasting. The mentioned methods do not rely only on functional flexibility regarding the ability to adapt to irregular data, but also on real-time information and their flows. So, machine learning can produce valuable and important results, especially regarding forecasting applications and when they align with proper data availability and timing in the revision processes, the results are quite accurate.

3.5 Summary of the ML literature

Machine learning methods can be more effective and present better performance when large datasets, non-linear relationships, connections, and complicated associations are used. However, there are some limitations, like the lack of interpretability. These problems can be solved by using proper explainability approaches, like SHAP or other feature-selection-based methods.

Chapter 4. Hybrid and Macro-Finance Approaches

4.1 Merging Econometrics and Machine Learning

Hybrid models can usually combine econometric structures with the flexibility of machine learning and represent the current situation of both macroeconomic and macrofinancial forecasting. These models can lead to and give responses to the complementary strengths, but also the weaknesses, of more traditional econometric methods and modern machine learning techniques. Additionally, even if econometric models present interpretability, theoretical foundation, and well-defined statistical properties, they seem to struggle to represent non-linear variables, possible structural breaks, and high dimensionality of predictors. On the other hand, machine learning models have substantial gains, especially when it comes to predictive accuracy by investigating and incorporating complex patterns. However, they often present issues regarding transparency and economic interpretability. Thus, hybrid approaches aim to address these limitations in order to integrate both paradigms within a specific, but still unified, forecasting environment.

Stempien and Slepaczuk (2025) presented an example of this approach in their research. More specifically, they created a hybrid model that combined ARIMA and GARCH residual structures with machine learning algorithms, such as XGBoost, Support Vector Machines, and LSTM networks. Through these enhancements, the combined econometric structure was able to investigate linear dynamics, long memory, and persistence, while through machine learning the model was able to remain non-linear and flexible in capturing complex patterns. Additionally, they applied the approach to the S&P 500 and Bitcoin prices. The researchers used the hybrid model, which led to improvements in forecasting accuracy compared to pure econometric models or pure machine learning models when investigated separately. Thus, the final evaluation regarding forecast accuracy was improved by up to 5% in terms of mean squared error, which reveals how hybrid learning methods are more suitable in volatile financial environments.

López-Estrada et al. (2025) proposed a hybrid model using VAR and XGBoost frameworks in order to forecast financial results in datasets that are characterized by uncertainty regarding macroeconomic indicators. In their approach, they used the VAR structure in order to investigate lagged dependencies and dynamic correlations between variables, while XGBoost was used to predict non-linear patterns that the VAR framework could not manage or predict. The aim of this hybrid model was not to replace the econometric base, but to complement it by capturing possible learning deviations and non-linearities. Through their empirical results, they showed that the hybrid VAR and machine learning method outperformed both pure VAR models and machine learning models used alone in forecasting S&P 500 prices. This provided evidence that residual learning is a quite effective method in macrofinancial forecasting.

Formally, a general hybrid forecasting structure can be represented as

$$y_t = \hat{y}_t^{(ECON)} + g_\theta(Z_t) + \varepsilon_t,$$

where $\hat{y}_t^{(ECON)}$ denotes the prediction generated by a baseline econometric model, such as an ARIMA or VAR system, and $g_\theta(\cdot)$ is a non-linear machine learning function applied to a set of macro-financial predictors Z_t. The error term ε_t captures remaining unexplained variation. The above, reveal the nature of the hybrid models, that they include sections that are based on the theoretical interpretable foundation of econometry, while through the machine learning side mostly aims to investigate the complex relationships between the variables, higher-order dependencies, and non-linear effects.

One of the most important advantages of this hybrid structure is its robustness to possible structural changes and periods that face economic instability. Linear econometric models can perform adequately when relationships between variables are stable. However, when crises occur, policy regimes shift, and episodes of high uncertainty appear, the relationships may no longer be stable and can lead to forecasting errors and mistakes. Thus, hybrid models are more appropriate for such environments, since the machine learning parts can properly adapt to changing patterns without needing to be re-specified

Meanwhile, the core of the econometric field meets again some of the interpretability that is usually lost in machine learning methods. By making sure that predictions are well understood in a statistical modeling sense, hybrid approaches can offer a clear narrative and interpretable results, while still presenting a high level of flexibility in comparison to pure machine learning methods. Thus, hybrid methods offer a balance between interpretability and accuracy, which makes them quite useful for policy-relevant tasks and results, while also presenting a high level of predictive power when it comes to economic datasets.

4.2 Macro-Finance Linkage and Variable Importance

The main question in macrofinancial forecasting is related to whether financial variables can genuinely predict information about macroeconomic outcomes beyond what is already captured by traditional indicators used in the macroeconomic field. This problem significantly affects the design of the models that are used and the interpretation of policy implications. Financial variables, like equity indices, term spreads, and credit spreads, are forward-looking and reflect market expectations about the state of the economy. However, the extent to which this information can be effectively used remains an open question, especially in unstable economic environments.

Bachmair and Schmitz (2025) investigated this relationship by using the predictive content of a large range of variables, both financial and macroeconomic, for important macroeconomic outcomes. Through their analysis, many hypotheses were investigated, mostly regarding the importance of financial indicators, using both classical models and machine learning techniques that are more flexible. The results revealed that financial variables can improve the forecasts of macroeconomic variables, but the reliability of this improvement depends strongly on the different frameworks and methods that were used.

More specifically, their work highlighted that the use of VAR models does not seem to be able to present useful predictive results from financial variables when these are directly included in the model. Additionally, the linear structure and fixed-parameter nature that these systems present lead to limitations regarding their ability to investigate complex and non-linear correlations, which are quite frequent in financial markets and interact with the real economy. However, Random Forest and other machine learning algorithms usually appear better prepared to incorporate financial variables and contribute more to forecasting. This difference also highlights that the informational content of financial data, which is often latent and non-linear, increases the need for more flexible models even further.

4.3 Mixed-Frequency Hybrid Models

One more important challenge that is usually encountered in macrofinancial forecasting is related to the mixed-frequency nature of the data. Macroeconomic indicators are usually given at monthly or even quarterly frequency, while financial variables are available weekly, daily, or even multiple times within the same day. Thus, this mismatch can lead to a loss of important information or to misspecification when high-frequency data are incorporated directly into low-frequency variables.

Li et al. (2025) created MS-MIDAS-LASSO models that provide a structured framework in order to integrate mixed-frequency variables into forecasting. Through their approach, high-frequency variables were able to influence low-frequency outcomes with the help of flexible weighting schemes. At the same time, they applied a proper shrinkage technique in order to control model complexity. Through this combination, they were able to create an effective high-dimensional setting, with many predictors available at different frequencies and publication times. Variables like the VIX, as well as credit spreads and sentiment indicators, seem to hold predictive power for macroeconomic outcomes. This predictive power can be materialized only when the data are properly analyzed using appropriate econometric and machine learning techniques. Lasso and other shrinkage methods assist in isolating the most relevant predictors, while regime-switching specifications help the model adapt to possible structural changes during different stages of the economic cycle.

The above results make it clear that methodology choice, as well as data structure and model design, is very important, especially for hybrid forecasting models. These models show that they can provide a proper framework that offers informational advantages from both high-frequency and low-frequency data, so statistical robustness is not reduced. Overall, this represents a critical component of modern macrofinancial forecasting methods.

4.4 Macro-Technical-Sentiment Integration

Through recent studies, it was revealed that macrofinancial forecasting models have also included macroeconomic indicators, technical financial variables, and sentiment variables in unified models. Patsiarikas et al. (2025) investigated a machine learning environment with a combination of macroeconomic variables, technical indicators that were based on prices, and sentiment-based variables that could predict movements in the S&P 500. They showed a high level of predictive power, with a reported R-square value of 0.99 for some model specifications. Thus, it was highlighted that the potential effectiveness of this integration can vary between the different information sources that are included in forecasting models.

Meanwhile, the authors also acknowledged that the risk of overfitting was quite high when there was high dimensionality and when the frameworks were rich in information. More specifically, when multiple data sources were used and combined, especially when they were closely correlated, models could capture noise instead of underlying economic relationships. This concern also revealed that, although machine learning models can fit complex patterns at a high level, they may lead to poor out-of-sample results if there are not careful regularization and validation of the variables.

Concluding, the results of the research illustrate that the use of macroeconomic, technical, and sentiment integration is important. Hybrid models, especially when they use multiple data sources, can provide higher predictive accuracy and are practically more useful according to proper model validation. Lastly, they also present embedded feature selection, which leads to better interpretation of the outcomes.

Chapter 5. Summary of the Literature Review

The literature review gave many important insights that are related to the framework of the present thesis:

1. Financial variables present some mediocre predictive power, which increases when flexible ML or hybrid approaches are incorporated (Bachmair & Schmitz, 2025; Pirayesh Neghab et al., 2025).

2. Tree-based models like RF and XGBoost outperform with a great consistence the traditional linear models, when large number of predictors are available to them (Nti et al., 2019; López-Estrada et al., 2025).

3. The hybrid models present a high overall performance since they combine the econometric structure with the flexibility of the non-linear machine learning models (Stempien & Slepaczuk, 2025).
4. Feature selection is of high importance when the data sets present high-dimensionality (Yang et al., 2022; Li, 2025).
5. Mixed-frequency methods like MIDAS, are frequently used and quite important for data with higher-frequency data than the macroeconomic data (Li et al., 2025).
6. Real-time constraints such as publication lags affect the accuracy of forecasting (Mamedli & Shibitov, 2021).

These results support the modelling strategy of tracking financial indices (S&P500, NASDAQ, VIX), commodities (gold, oil), and macroeconomic indicators (unemployment, CPI, sentiment) to develop a predictive system based on econometric baselines and augmented ML models.

In general, the recent literature documents a clear transition in macroeconomic and macro-financial forecasting from traditional linear econometric models toward more flexible and data-driven approaches. Classical autoregressive frameworks remain valuable due to their interpretability, theoretical grounding, and statistical transparency. Nevertheless, empirical evidence consistently shows that their forecasting performance weakens in modern data environments characterized by non-linear dynamics, structural breaks, regime shifts, and high-dimensional predictor sets, limiting their effectiveness when economic systems become more complex and unstable.

Machine learning methods have gained prominence as a response to these limitations, offering the ability to model complex interactions and non-linear relationships across large information sets. Numerous studies report superior predictive accuracy compared to conventional econometric models, particularly when financial and macroeconomic variables are jointly considered. However, the literature also emphasizes that these improvements depend critically on careful model design. Without proper regularization, feature selection, and strict out-of-sample validation, ML models are prone to overfitting and often lack economic interpretability, reducing their usefulness for policy analysis and theoretical insight.

A dominant conclusion across studies is that hybrid forecasting frameworks provide an effective balance between structure and flexibility. By integrating econometric models that capture persistent linear relationships with machine learning components that model residual non-linearities and regime-dependent behavior, hybrid approaches consistently outperform both standalone econometric and pure ML models. Furthermore, research highlights the importance of aligning model choice with data characteristics, including mixed-frequency structures and real-time data constraints. These findings motivate the use of enriched macro-financial datasets and comparative forecasting frameworks, positioning hybrid models as a robust and state-of-the-art solution for forecasting in complex macro-financial environments.

Chapter 6. Empirical Design and Analysis

6.1 Objectives of the research and dependent variables

The main aim of the present study is to investigate the dynamics of financial market volatility according to the VIX index. So, the target variable is the algorithmic daily return of the index, which is measured by:

$$VIX\_ret\_log_t = \log(VIX_t) - \log(VIX_{t-1})$$

Log-returns makes sure that the variance is stabilized so the series are proper to use in linear econometric models, but also machine learning methods. So, the current forecasting exercise is focused only on one step ahead predictions.

6.2 Explanatory research variables

The explanatory variables used in the present study are major macro-financial indicators that usually are connected to the volatility transmission mechanisms. These include:

- Log-returns of the S&P 500 index (GSPC_ret)
- Log-returns of the NASDAQ index (IXIC_ret)
- Changes in U.S. Treasury yields (TNX_ret)
- Log-returns of oil prices (Oil_ret)
- Log-returns of gold prices (Gold_ret)
- Log-returns of the EUR/USD exchange rate (EUR_ret)
- Log-returns of the GBP/USD exchange rate (GBP_ret)

Additionally, all of the above are used after the proper transform into logarithmic return forms, to be compatible regarding scale but also to ensure stationarity.

6.3 Estimation and forecasting framework

To ensure the comparability among different modeling approaches, the same framework was used and applied to all these specifications. Analytically, for the model estimation and identification, a window of 1,200 observations was chosen. As for the performance of the forecasts, it is evaluated based on a one step ahead procedure, which has a fixed window. More specifically :

- At each step, the model is estimated using the most recent 1,200 observations.
- A one-period-ahead forecast is generated.
- The window then rolls forward by one observation.
- The procedure continues until the end of the sample.

Through this method, a clear and strict evaluation is ensured, as each specification is assessed under the same conditions. The same out of sample period is used, which also mitigates any bias. A naïve benchmark model was used as a reference, which is defined as:

$$\hat{y}_t = y_{t-1}$$

This benchmark model is the baseline and will be compared to all the models that will be assessed.

6.4 Modeling Approaches

The empirical analysis of the present thesis will compare multiple model classes to evaluate forecasting performance of linear, factor based, regularized, and nonlinear methods. The models are separated into five specific groups, which are the following:

1. **Linear time-series econometric models**
    - ARIMA
    - VAR
2. **Factor-based dimension reduction model**
    - Dynamic Factor Model
3. **Regularized linear regression models**
    - LASSO
    - Elastic Net

4. **Non-linear tree-based models**
    - CART
5. **Ensemble machine learning models**
    - Random Forest
    - XGBoost

The above modeling approaches allow a more structured comparison among the traditional econometric specifications and the modern machine learning methods that were applied.

6.5 Forecast evaluation methods

Regarding the forecast accuracy, it was evaluated based on two standard loss functions:

1. **Root Mean Squared Error (RMSE)**

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}$$

2. **Mean Absolute Error (MAE)**

$$MAE = \frac{1}{T}\sum_{t=1}^{T} | y_t - \hat{y}_t |$$

RMSE penalizes larger forecast errors more heavily, while MAE focuses on a robust measure of average absolute deviation. Using both of the functions contributes to the proper and comprehensive evaluation of predictive performance.

6.6 Comparative evaluation strategy

In order to quantify the relative performance of each one of the models, the percentage improvement over the naive benchmark was used, defined by:

$$Improvement(\%) = \frac{RMSE_{Naive} - RMSE_{Model}}{RMSE_{Naive}} \times 100$$

Through this measurement, an intuitive interpretation of forecasting gains across model classes was allowed. Also, since a consistent rolling estimation framework and uniform evaluation metrics were used, it was ensured that there was methodological comparability on all the specifications of the models.

Chapter 7. Empirical Results

ARIMA Model

In the ARIMA models, the dependent variable was the logarithmic daily return of the VIX index, while the focus was on the dynamics of financial market volatility. Alternative specifications were estimated through the use of 10,200 observations, while the selection was applied by the Akaike Information Criterion AIC and the Bayesian Information Criterion BIC. According to the results presented in Table 1, it seems that the ARIMA (1,0,1) model presents the lowest AIC value (AIC = -3176.882; BIC = -3156.522), which indicates that the fit relative to alternative lag structures is higher. In second place seems to be the ARIMA (2,0,1) model, which is characterized by a higher level of AIC (AIC = -3174.882). According to the above, it can be confirmed that the ARIMA (1,0,1) is characterized by a better balance when it comes to fit and parsimony.

Table 1. Candidate ARIMA Models and Information Criteria

p	d	q	AIC	BIC
1	0	1	-3176.9	-3156.5
2	0	1	-3174.9	-3149.4
1	0	2	-3174.9	-3149.4
1	0	3	-3172.9	-3142.4
3	0	1	-3172.9	-3142.4
2	0	2	-3172.9	-3142.4
3	0	3	-3172.8	-3132.1
3	0	2	-3170.9	-3135.3
2	0	3	-3170.9	-3135.3
0	0	3	-3168.8	-3143.3
0	0	2	-3168	-3147.7
3	0	0	-3166.8	-3141.4
2	0	0	-3166.7	-3146.3
0	0	1	-3166	-3150.7
1	0	0	-3165.4	-3150.1
1	1	2	-3165.1	-3144.8
0	0	0	-3162.6	-3152.4
2	1	2	-3162.4	-3136.9

2	1	3	-3161.3	-3130.7
3	1	2	-3161.2	-3130.7
3	1	3	-3159.3	-3123.6
0	1	3	-3157	-3136.7
3	1	1	-3155.8	-3130.3
2	1	1	-3155.7	-3135.3
0	1	2	-3155.1	-3139.8
1	1	1	-3154.5	-3139.3
1	1	3	-3153.2	-3127.7
0	1	1	-3151.8	-3141.7
3	1	0	-2828.7	-2808.4
2	1	0	-2759.9	-2744.6
1	1	0	-2610.6	-2600.4
0	1	0	-2256.2	-2251.1

In the following Table 2, the model with the minimum AIC value is presented, which also confirms that the model ARIMA (1,0,1) is the best specification for subsequent estimation and forecasting.

Table 2. Selected ARIMA Model Based on Minimum AIC

p	d	q	AIC	BIC
1	0	1	-3176.9	-3156.5

Through Table 3, the parameter estimates of the ARIMA(1,0,1) model are presented. It seems that the autoregressive coefficient is statistically significant and positive AR1 = 0.7867, SE = 0.0821, with a t statistic of 9.59, which indicates short term persistence in VIX log returns $p < .001$. Also, it seems that the moving average coefficient is statistically significant but negative MA1 = -0.8633, SE = 0.0675, t = -12.79, $p < .001$. This indicates that there is a corrective adjustment when it comes to the error dynamics. Lastly, the constant term is also statistically significant, even if small mean = -0.0026, SE = 0.0012, t = -2.17, $p < .05$, which seems to align with the unconditional mean being near zero, which is typical when it comes to daily financial return series.

Table 3. Parameter Estimates of the ARIMA Model

Parameter	Estimation	Std. Error
ar1	0.7867	0.0821
ma1	-0.8633	0.0675
mean	-0.0026	0.0012

The model adequacy is also evaluated by the use of the Ljung Box test, which focuses on the residual autocorrelation. According to the results in Table 4, the test statistic was Q (20) = 10.902 with p = .9487. So, it is revealed that the hypothesis related to no residual autocorrelation is not rejected. The result highlight that the ARIMA (1,0,1) specification sufficiently represent the linear dependence structure of the series and leaves no systematic serial correlation in the residuals.

Table 4. Ljung–Box Test for Residual Autocorrelation

Statistic	df	p value
10.902	20	0.9487

Regarding the predictive performance of the model, it is investigated with the utilization of rolling one-step-ahead forecasts with a specific window scheme. As it is presented in Table 5, the ARIMA (1,0,1) model presents RMSE = 0.0756 and MAE = 0.0529, when it comes to the comparison with the naive benchmark (RMSE = 0.1111; MAE = 0.0778). So, it means that there was a 31.9% reduction in RMSE compared to the benchmark, which makes it clear that there is a significant improvement in forecast accuracy.

Table 5. Rolling One-Step-Ahead Forecast Performance

Model	RMSE	MAE
Naive y[t-1]	0.1111	0.0778
ARIMA(1,0,1)	0.0755	0.0529

In Table 6 the last ten rolling forecasts are presented, in order to compare the results of the actual values, naive predictions, and ARIMA forecasts. The predictions of the

ARIMA model present a higher stability compared to the naive model. This indicates that the model smooths volatility and underreacts during abrupt regime shifts.

Table 6. Last Ten Rolling Forecasts

Date	Actual	Naive	ARIMA
5/11/2025	-0.0535	0.1012	-0.0101
6/11/2025	0.0794	-0.0535	0.0022
7/11/2025	-0.0218	0.0794	-0.008
10/11/2025	-0.0807	-0.0218	1.6405
11/11/2025	-0.0183	-0.0807	0.0054
12/11/2025	0.0132	-0.0183	0.0006
13/11/2025	0.1329	0.0132	-0.0022
14/11/2025	-0.0085	0.1329	-0.0129
17/11/2025	0.1209	-0.0085	-0.0021
18/11/2025	0.0982	0.1209	-0.0116

Lastly, the model was also used this time with the full sample and an out of sample forecast was created. The results are presented in Tables 7 and 8, and the values gradually converge towards the unconditional mean. Also, it seems that the 80% and 95% confidence intervals get higher as the forecast horizon increases, which indicates that the forecast dispersion increases due to uncertainty. More specifically, the first out of sample forecast has a value of -0.0235 with a 95% confidence interval of [-0.1651, 0.1181], which highlights the inherent uncertainty that exists when it comes to short term volatility predictions.

Table 7. Out-of-Sample Forecasts with Confidence Intervals

Date	Estimation	LB 80%	UB 80%	LB 95%	UB 95%
19/11/2025	-0.0235	-0.1161	0.0691	-0.1651	0.1181
20/11/2025	-0.0198	-0.1126	0.0729	-0.1617	0.1220
21/11/2025	-0.0167	-0.1097	0.0761	-0.1588	0.1253
22/11/2025	-0.0142	-0.1072	0.0787	-0.1564	0.1280
23/11/2025	-0.0121	-0.1052	0.0809	-0.1544	0.1301
24/11/2025	-0.0104	-0.1035	0.0826	-0.1528	0.1319
25/11/2025	-0.009	-0.1021	0.0841	-0.1514	0.1334
26/11/2025	-0.0078	-0.1009	0.0853	-0.1502	0.1346
27/11/2025	-0.0068	-0.0999	0.0863	-0.1493	0.1356
28/11/2025	-0.0060	-0.0991	0.0871	-0.1485	0.1364

VAR Model

Continuing, a multivariate vector autoregression model was estimated, with the logarithmic returns of the VIX index being used as the dependent variable. Also, the exogenous variables were proper macrofinancial indicators. Regarding the lag length selection, it took place on 1,200 observations with the use of standard information criteria. In Table 8, the lag order selection results are presented. Both the Akaike Information Criterion AIC and the Final Prediction Error FPE indicate a lag order of ten. Additionally, the Hannan Quinn criterion HQ and the Schwarz Criterion SC or BIC suggest a lag order of one. Since there is a stronger penalty by the BIC and the goal is to maintain parsimony in the multivariate model, VAR(1) was chosen for the subsequent estimation.

Table 8. VAR Lag Order Selection Criteria (Identification Window)

Criterion	Lag
AIC(n)	10
HQ(n)	1
SC(n) (BIC)	1
FPE	10

In the following Table 9, the estimation results for the VAR (1) model are presented within the identification window. As it seems, the model presents LogLik = 28097.5, with AIC = -56051 and BIC = -55685. The results indicate a satisfactory fit, while the model also demonstrates a proper structure.

Table 9. Model Fit Criteria for VAR (1) in the Identification Window

Model	LogLik	AIC	BIC
VAR(1)	28097.5	-56051	-55685

Also, the stability of the model was investigated using the modulus of the eigenvalues of the companion matrix. The results in Table 10 make it clear that all the eigenvalue moduli are below one, while the largest root has a value of 0.3792. The results make it clear that the model process presents satisfactory stability and represents a stationary multivariate system.

Table 10. Stability Test of VAR (1) Based on Eigenvalue Moduli

Eigenvalue Moduli
0.3791
0.1143
0.1105
0.1105
0.0843
0.0843
0.0684
0.0499

Furthermore, there are residual diagnostic tests examined, while the Portmanteau test for residual correlation is presented in Table 11. According to the test statistics, a value of 1915.04 with 960 degrees of freedom and $p < 0.001$ is reported. This indicates that the null hypothesis of no serial correlation is rejected, which reveals the presence of residual autocorrelation in the multivariate system.

Table 11. Portmanteau Test for Residual Autocorrelation

Test	Statistic	df	p-value
Portmanteau	1915.04	960	0.000

Additionally, the multivariate ARCH LM test was used for heteroskedasticity, and the results are presented in Table 12. As it seems, the test statistic has a value of 20114.3 with 6480 degrees of freedom and $p < 0.001$, which indicates the rejection of the null hypothesis of no heteroskedasticity. The results are also consistent with the typical behavior regarding financial time series residuals.

Table 12. Multivariate ARCH LM Test for Residual Heteroskedasticity

Test	Statistic	df	p-value
Multivariate ARCH LM	20114.3	6480	0

Regarding the predictive results of the VAR(1) model, they were evaluated with a rolling one step ahead forecast and the use of a fixed window. As presented in Table 14, the VAR(1) model has a value of RMSE = 0.0759 and MAE = 0.0530, compared to the naïve benchmark RMSE = 0.1111 and MAE = 0.0778. This shows that the RMSE is reduced by almost 31.7% compared to the naïve benchmark, which suggests that the accuracy of the forecast was improved, even if there was residual autocorrelation and heteroskedasticity.

Table 14. Rolling One-Step-Ahead Forecast Evaluation for VIX_ret_log

Model	RMSE	MAE
Naive (y[t-1])	0.1111	0.0778

VAR(1)	0.0759	0.0531

In the following Table 15, the last 20 rolling forecasts of the chosen model are presented in order to compare the predicted values and the actual values. As it seems, the predictions of the model seem to be smoother compared to the naïve benchmark predictions, which means that there is a higher level of adaptability to volatility fluctuations, even if there are some forecast errors during periods of changes in the VIX returns.

Table 15. Last Twenty Rolling Forecasts of VAR(1)

Date	Actual	Naive	VAR
16/10/2025	0.2039	-0.0082	0.0019
21/10/2025	-0.0199	0.2039	-0.0004
22/10/2025	0.0400	-0.0199	0.0003
23/10/2025	-0.0725	0.0400	-0.004
28/10/2025	0.0391	-0.0725	-0.0021
29/10/2025	0.0300	0.0391	0.0012
30/10/2025	-0.0006	0.0300	-0.0057
31/10/2025	0.0308	-0.0006	-0.0129
3/11/2025	-0.0156	0.0308	-0.0001
4/11/2025	0.1012	-0.0156	-0.0022
5/11/2025	-0.0535	0.1012	-0.0065
6/11/2025	0.0794	-0.0535	-0.0047
7/11/2025	-0.0218	0.0794	-0.0065
10/11/2025	-0.0807	-0.0218	0.0016
11/11/2025	-0.0183	-0.0807	0.0076
12/11/2025	0.0132	-0.0183	0.0013
13/11/2025	0.1329	0.0132	-0.0002
14/11/2025	-0.0085	0.1329	-0.0137
17/11/2025	0.1209	-0.0085	-0.0032
18/11/2025	0.0982	0.1209	-0.0065

Continuing, the full sample was used in order to re estimate the model, with the results being presented in Table 16. It is evident that the last VAR(1) model achieved

LogLik = 95387.5, AIC = -190631, and BIC = -190178. This suggests that there is an improvement in the fit when all the observations are used.

Table 16. Model Fit Criteria of Final VAR(1) (Full Sample)

Μοντέλο	LogLik	AIC	BIC
VAR(1)	95387.5	-190631	-190178

Additionally, an out of sample forecast was generated for VIX_ret_log. The results in Table 17 present the forecasted values with 80% and 95% confidence intervals. This means that the forecasts gradually converge and approach zero, with the confidence intervals remaining substantial, making it clear that forecast uncertainty is inherent in volatility modeling. Additionally, the first out of sample forecast has a value of -0.0068 with a 95% confidence interval of [-0.1488, 0.1352].

Table 17. Out-of-Sample Forecasts of VAR(1) with 95% Confidence Intervals

Date	Estimation	LB 95%	UB 95%
19/11/2025	-0.0068	-0.1488	0.1351
20/11/2025	-0.0022	-0.1448	0.1404
21/11/2025	-0.0020	-0.1446	0.1406
22/11/2025	-0.0021	-0.1447	0.1405
23/11/2025	-0.0020	-0.1447	0.1405
24/11/2025	-0.0020	-0.1447	0.1405
25/11/2025	-0.0020	-0.1447	0.1405
26/11/2025	-0.0020	-0.1447	0.1405
27/11/2025	-0.0020	-0.1447	0.1405
28/11/2025	-0.0020	-0.1447	0.1405

Dynamic Factor Model

Regarding the dynamic factor model framework, principal component analysis was applied to the financial variables in order to reduce dimensionality and to extract common dynamic factors. The main aim was to summarize the joint variation of the macrofinancial indicators that were chosen into fewer components that would be able to represent the dominant sources of comovement.

The results of the principal component analysis are presented in Table 18. It is evident that the first factor explains 33.28% of the total variance, while the second factor explains 22.18%, and the third and fourth explain 14.83% and 13.28%. Overall, the first four factors explain a total of 83.57% of the variance, which exceeds the 80% threshold, so these four factors were used for the subsequent modeling.

Table 18. Explained Variance by Principal Components (Dynamic Factor Model)

Factor	Explained Variance	Cumulative Variance
Factor 1	0.3328	0.3328
Factor 2	0.2218	0.5546
Factor 3	0.1483	0.703
Factor 4	0.1328	0.8357
Factor 5	0.0891	0.9249
Factor 6	0.0684	0.9933
Factor 7	0.0067	1

The retained factors were subsequently used as regressors in a linear model with VIX logarithmic returns as the dependent variable. The regression results are reported in Table 2. The intercept is negative and statistically significant (Estimate = -0.002, SE = 0.0007, t = -2.69, p = .007).

Factor 1 exhibits a strong negative and highly significant effect (Estimate = -0.0320, SE = 0.0005, t = -64.24, p < .001). Factor 2 is also negative and statistically significant (Estimate = -0.0130, SE = 0.0006, t = -21.38, p < .001). Factor 3 remains significant with smaller magnitude (Estimate = -0.00350, SE = 0.0007, t = -4.68, p < .001). Factor 4 displays a positive and statistically significant association with VIX log-returns (Estimate = 0.0173, SE = 0.0007, t = 21.87, p < .001). The statistical significance of all retained factors confirms the presence of strong common macro-financial dynamics influencing volatility.

Table 19. Regression of VIX Log-Returns on Dynamic Factors

Predictor	Estimate	Std. Error	t value	p value
Intercept	-0.0021	0.0007	-2.69	0.007
Factor 1	-0.0321	0.0005	-64.24	<0.001
Factor 2	-0.0131	0.0006	-21.38	<0.001
Factor 3	-0.0035	0.0007	-4.68	<0.001
Factor 4	0.0173	0.0007	21.87	<0.001

To investigate the predictive performance of the dynamic factor model, a rolling one step ahead forecast was used. The results are presented in Table 20, and it seems that the naïve benchmark presents values RMSE = 0.1111 and MAE = 0.0778, while the DFM model with the use of the four factors has values RMSE = 0.0521 and MAE = 0.0350. This shows that the RMSE was reduced by 53.1% compared to the naïve benchmark, which is an important improvement when it comes to the accuracy of the forecasts.

Table 20. Rolling One-Step-Ahead Forecast Evaluation (DFM vs Naive)

Model	RMSE	MAE
Naive (y[t-1])	0.1111	0.0778
DFM (4 factors)	0.0521	0.035

Table 21 presents the last 20 rolling one step ahead DFM forecasts in order to compare them with the actual observations and the naïve benchmark predictions. As it seems, the DFM predictions have a stronger alignment with the actual volatility movements compared to the naïve benchmark predictions, especially during periods of abrupt fluctuations.

Table 21. Rolling One-Step-Ahead DFM Forecasts (Last 20 Observations)

Date	Actual	Naive	DFM
16/10/2025	0.2040	-0.0082	0.0210
21/10/2025	-0.0199	0.2040	0.0091
22/10/2025	0.0400	-0.0199	0.0317
23/10/2025	-0.0725	0.0400	-0.0276
28/10/2025	0.0391	-0.0725	-0.0199
29/10/2025	0.0300	0.0391	-0.0123
30/10/2025	-0.0006	0.0300	0.0493
31/10/2025	0.0309	-0.0006	-0.0173
3/11/2025	-0.0156	0.0309	-0.0134
4/11/2025	0.1013	-0.0156	0.0661
5/11/2025	-0.0535	0.1013	-0.0249
6/11/2025	0.0795	-0.0535	0.0627
7/11/2025	-0.0218	0.0795	0.0046
10/11/2025	-0.0807	-0.0218	-0.0776
11/11/2025	-0.0183	-0.0807	0.0031
12/11/2025	0.0132	-0.0183	-0.0006
13/11/2025	0.1330	0.0132	0.0802
14/11/2025	-0.0085	0.1330	0.0038
17/11/2025	0.1210	-0.0085	0.0366
18/11/2025	0.0982	0.1210	0.0427

CART Model

Additionally, the classification and regression tree model was used as a nonlinear benchmark specification. The model is an intermediate approach between traditional econometric models and more advanced machine learning methods and techniques. The aim was to investigate the potential of nonlinear relationships that may occur between the macrofinancial variables and the VIX logarithmic returns, while having a proper level of interpretability.

To analyze the predictive performance of the CART model, a rolling one step ahead forecast method was used with a fixed window. The results are presented in Table

22, with the naïve benchmark having values of RMSE = 0.1111 and MAE = 0.0778 and the CART model having values of RMSE = 0.0508 and MAE = 0.0355. So, it seems that the RMSE was reduced approximately 54.3% compared to the naïve benchmark, which indicates an improvement in the accuracy of the forecast. Also, there is a visible reduction in MAE, which confirms the precision of the model.

Table 22. Rolling One-Step-Ahead Forecast Evaluation for the CART Model

Model	RMSE	MAE
Naive (y[t-1])	0.1111	0.0778
CART	0.0508	0.0355

In Table 23, the rolling forecasts of the CART model are presented in order to compare them with the actual observations and the naïve predictions. As it seems, the CART predictions have a higher alignment with the VIX returns compared to the naïve benchmark, especially during periods of high volatility. More specifically, on 2025-11-13, the actual return has a value of 0.1330, with the naïve forecast having a value of 0.0132 and the CART forecast having a value of 0.1040, which shows an improvement in capturing volatility. Similarly, on 2025-11-10, the actual return was -0.0807, with the forecast of the CART model having a value of -0.0659 and the naïve forecast having a value of -0.0218, which also was closer to the actual value.

Table 23. Rolling CART Forecasts for the Last 20 Observations

Date	Actual	Naive	CART
16/10/2025	0.2040	-0.0082	0.0239
21/10/2025	-0.0199	0.2040	-0.0258
22/10/2025	0.0400	-0.0199	0.0080
23/10/2025	-0.0725	0.0400	-0.0273
28/10/2025	0.0391	-0.0725	-0.0149
29/10/2025	0.0300	0.0391	0.0030
30/10/2025	-0.0006	0.0300	0.0755
31/10/2025	0.0309	-0.0006	-0.026
3/11/2025	-0.0156	0.0309	-0.0249
4/11/2025	0.1013	-0.0156	0.0740
5/11/2025	-0.0535	0.1013	-0.0275
6/11/2025	0.0795	-0.0535	0.0746
7/11/2025	-0.0218	0.0795	-0.0146
10/11/2025	-0.0807	-0.0218	-0.0659
11/11/2025	-0.0183	-0.0807	-0.0150
12/11/2025	0.0132	-0.0183	0.0041
13/11/2025	0.1330	0.0132	0.1040
14/11/2025	-0.0085	0.1330	-0.0150
17/11/2025	0.1210	-0.0085	0.0770
18/11/2025	0.0982	0.1210	0.0305

The variable importance analysis also presents more insights about the nonlinear structure that the CART model presents. According to the results in Table 24, the most important predictor is the returns of the US equity index, more specifically GSPC_ret S&P 500 returns, which has the highest score of importance with a value of 13.10. In second place is IXIC_ret NASDAQ returns with a score of 9.03, while the remaining variables have lower importance values, including TNX_ret 0.63, Oil_ret 0.56, EUR_ret 0.33, GBP_ret 0.16, and Gold_ret 0.13. This reveals that equity market movements constitute the main driver of VIX fluctuations, with interest rates, commodities, and exchange rates having a secondary role, while still being important.

Table 24. Relative Variable Importance in the Final CART Model

Variable	Importance
GSPC_ret	13.1
IXIC_ret	9.03
TNX_ret	0.63
Oil_ret	0.56
EUR_ret	0.33
GBP_ret	0.16
Gold_ret	0.13

LASSO Model

The LASSO model was used as a regularized regression framework in order to simultaneously predict and perform variable selection in high dimensional settings. For that reason, rolling one step ahead forecasts were used with a specific window scheme in order to be able to compare the results with the previous models that were used, both econometric and nonlinear.

The forecast performance of the LASSO model is presented in Table 25. As it seems, the naïve benchmark presents values of RMSE = 0.1111 and MAE = 0.0778, while the LASSO model has values of RMSE = 0.0518 and MAE = 0.0349. This indicates that there was a reduction in RMSE reaching 53.4% compared to the naïve benchmark results, which suggests that there is higher predictive accuracy of the model. Also, the reduction in MAE makes clear that the regularization model performs properly when it comes to short term volatility forecasting.

Table 25. Rolling One-Step-Ahead Forecast Evaluation for the LASSO Model

Model	RMSE	MAE
Naive (y[t-1])	0.1111	0.0778
LASSO	0.0518	0.0349

In the following Table 26, the last 20 rolling forecasts of the LASSO model are presented in order to compare the results of the actual observation, the naïve prediction, and the LASSO predictions. It is evident that the LASSO model predictions have a better alignment with the VIX returns in comparison to the naïve benchmark. More specifically, on 2025-11-13, the return was 0.1330, with the naïve forecast having a value of 0.0132 and the LASSO forecast having a value of 0.0792, which gives better tracking when it comes to volatility spikes. Similar results are presented on 2025-11-10, with the actual return being -0.0807, the LASSO forecast being -0.0661, and the naïve forecast being -0.0218, with the LASSO remaining closer to the actual return.

Table 26. Rolling LASSO Forecasts for the Last 20 Observations

Date	Actual	Naive	LASSO
16/10/2025	0.2040	-0.0082	0.0319
21/10/2025	-0.0199	0.2040	-0.0206
22/10/2025	0.0400	-0.0199	0.0262
23/10/2025	-0.0725	0.0400	-0.0208
28/10/2025	0.0391	-0.0725	-0.0222
29/10/2025	0.0300	0.0391	-0.0057
30/10/2025	-0.0006	0.0300	0.0520
31/10/2025	0.0309	-0.0006	-0.0189
3/11/2025	-0.0156	0.0309	-0.0124
4/11/2025	0.1013	-0.0156	0.0583
5/11/2025	-0.0535	0.1013	-0.0148
6/11/2025	0.0795	-0.0535	0.0595
7/11/2025	-0.0218	0.0795	0.0033
10/11/2025	-0.0807	-0.0218	-0.0661
11/11/2025	-0.0183	-0.0807	0
12/11/2025	0.0132	-0.0183	0.0137
13/11/2025	0.1330	0.0132	0.0792
14/11/2025	-0.0085	0.133	-0.0067
17/11/2025	0.1210	-0.0085	0.0339
18/11/2025	0.0982	0.1210	0.0385

Through the regularization procedure, strong sparsity in the final model was estimated when using the total sample. According to the results in Table 27, only 2 of the variables had a non-zero coefficient, which are GSPC_ret S&P 500 returns and IXIC_ret NASDAQ returns, with coefficients of -2.1618 and -1.1929, respectively. So, it seems to address a negative association with the VIX returns, which is also strong.

Additionally, the elimination of the other macrofinancial variables makes it clear that when the information contained in major US equity indices is accounted for, it is evident that the other predictors have limited contribution to the marginal explanatory power. This supports that the US stock market primarily is a channel of information and uncertainty to the VIX, with the other macrofinancial indicators having weaker predictive power.

Table 27. Variables Selected by the LASSO Model in the Full Sample

Variable	Coefficient
GSPC_ret	-2.1618
IXIC_ret	-1.1929

Elastic Net Model

The elastic net model was used as a regularized machine learning framework which presents both the properties of LASSO and ridge regression and focuses on enhancing predictive accuracy and at the same time offering a stable selection of variables in an environment with high correlations. In order to achieve that, the L1 and L2 penalizations were blended to avoid excessive sparsity and at the same time present proper robustness even if there is multicollinearity.

In Table 28, the performance of the elastic net model forecasting was investigated using rolling one step ahead forecasts with a properly fixed window scheme. As it seems, the naïve benchmark presents values of RMSE = 0.1111 and MAE = 0.0778, and the elastic net model has values of RMSE = 0.0515 and MAE = 0.0348. According to the results, the RMSE is reduced by 53.6% compared to the naïve benchmark, which seems that the model presents a higher level of predictive accuracy. Additionally, the MAE indicator is lower, which confirms that the model offers consistency and stability regarding each estimation approach. Also, the performance of the forecast is compared to that of the LASSO and CART models, which places the elastic net model in the top performing linear models category.

Table 28. Rolling One-Step-Ahead Forecast Evaluation for the Elastic Net Model

Model	RMSE	MAE
Naive (y[t-1])	0.1111	0.0778
Elastic Net	0.0515	0.0348

In the following Table 29, the last 20 rolling forecasts of the elastic net model are presented in order to compare them with the actual values and the naïve predictions. It is evident that the predictions of the elastic net model have better alignment with the realized VIX returns compared to the naïve benchmark. More specifically, on 2025-11-13, the return was 0.1330, with the naïve forecast being 0.0132 and the elastic net forecast being 0.0785, which is an improvement regarding the investigation of the volatility spikes. Also, on 2025-11-10, the return was -0.0807, with the elastic net forecast being -0.0655 and the naïve forecast being -0.0218, which also reveals that the elastic net forecast is more accurate.

 https://doi.org/10.26219/heal.aueb.10223

Table 29. Rolling Elastic Net Forecasts for the Last 20 Observations

Date	Actual	Naive	Elastic Net
16/10/2025	0.2040	-0.0082	0.0316
21/10/2025	-0.0199	0.2040	-0.0171
22/10/2025	0.0400	-0.0199	0.0262
23/10/2025	-0.0725	0.0400	-0.0202
28/10/2025	0.0391	-0.0725	-0.0222
29/10/2025	0.0300	0.0391	-0.0057
30/10/2025	-0.0006	0.0300	0.0512
31/10/2025	0.0309	-0.0006	-0.0181
3/11/2025	-0.0156	0.0309	-0.0123
4/11/2025	0.1013	-0.0156	0.0586
5/11/2025	-0.0535	0.1013	-0.0139
6/11/2025	0.0795	-0.0535	0.0589
7/11/2025	-0.0218	0.0795	0.0034
10/11/2025	-0.0807	-0.0218	-0.0655
11/11/2025	-0.0183	-0.0807	-0.0001
12/11/2025	0.0132	-0.0183	0.0124
13/11/2025	0.1330	0.0132	0.0785
14/11/2025	-0.0085	0.1330	-0.0063
17/11/2025	0.1210	-0.0085	0.0330
18/11/2025	0.0982	0.1210	0.0388

Additionally, the model estimated on the full sample seems to be able to retain the returns of the major US equity indices as the dominant predictors of the VIX values. According to the results in Table 30, GSPC_ret presents a coefficient of -1.883, with IXIC_ret having a coefficient of -1.457, indicating a negative relation between the equity market returns and the VIX values, which aligns with the feedback regarding the volatility effect in the financial sector. Also, the rest of the macrofinancial variables are excluded, which means that once proper equity market information is used, the other predictors do not contribute importantly to the explanatory power. This reinforces that the US stock market primarily acts as a source of uncertainty toward the VIX.

Table 30. Variables Selected by the Elastic Net Model in the Full Sample

Variable	Coefficient
GSPC_ret	-1.883
IXIC_ret	-1.457

Random Forest Model

Regarding the random forest model, it presents a strong nonlinear benchmark when it comes to the transition from the traditional econometric approaches to the machine learning techniques, generally creating a large number of decision trees using proper bootstrap sampling and random feature selection. The model can capture properly the complexity of the interaction effects and the nonlinear relationships when it comes to financial time series.

The predictive performance of the random forest model was investigated using once again a one step ahead forecast with a fixed window scheme. The results that are presented in Table 31 reveal that the benchmark has values of RMSE = 0.1111 and MAE = 0.0778, with the random forest model having values of RMSE = 0.0485 and MAE = 0.0337. So, the RMSE value was reduced 56.4% compared to the naïve benchmark, which means that there is a higher improvement than the previous models that were estimated so far. Also, the MAE value is lower, which confirms that the forecasting of the present model is superior when it comes to its accuracy.

Table 31. Rolling One-Step-Ahead Forecast Evaluation for the Random Forest Model

Model	RMSE	MAE
Naive (y[t-1])	0.1111	0.0778
Random Forest	0.0485	0.0337

Through Table 32, the last 20 rolling forecasts of the random forest model are presented in order to compare them to the actual observations and the naïve predictions. More specifically, it seems that the random forest predictions have a higher level of responsiveness when it comes to sudden volatility adjustments and also have better tracking of the VIX returns compared to the naïve benchmark. More specifically, on

2025-11-13, the actual return was 0.1330, with the naïve forecast being 0.0132 and the random forest forecast having a value of 0.1154, which reveals that the random forest model presents a closer alignment with the realized volatility spike. Similar are the results on 2025-11-10, with the actual return being -0.0807, with the random forest forecast being -0.0713 and being closer to the real value compared to the naïve forecast with the value of -0.0218.

Table 32. Rolling Random Forest Forecasts for the Last 20 Observations

Date	Actual	Naive	Random Forest
16/10/2025	0.2040	-0.0082	0.0412
21/10/2025	-0.0199	0.2040	-0.0397
22/10/2025	0.0400	-0.0199	0.0309
23/10/2025	-0.0725	0.0400	-0.0388
28/10/2025	0.0391	-0.0725	-0.0279
29/10/2025	0.0300	0.0391	-0.0114
30/10/2025	-0.0006	0.0300	0.0685
31/10/2025	0.0309	-0.0006	-0.0153
3/11/2025	-0.0156	0.0309	-0.0255
4/11/2025	0.1013	-0.0156	0.0800
5/11/2025	-0.0535	0.1013	-0.0245
6/11/2025	0.0795	-0.0535	0.0751
7/11/2025	-0.0218	0.0795	-0.0131
10/11/2025	-0.0807	-0.0218	-0.0713
11/11/2025	-0.0183	-0.0807	-0.0098
12/11/2025	0.0132	-0.0183	-0.0185
13/11/2025	0.1330	0.0132	0.1154
14/11/2025	-0.0085	0.1330	-0.0104
17/11/2025	0.1210	-0.0085	0.0748
18/11/2025	0.0982	0.1210	0.0640

Continuing, the permutation importance is presented and analyzed in Table 33. According to the results, GSPC_ret is characterized by the highest importance score with a value of 0.00316, followed by IXIC_ret with a value of 0.00142. This pattern indicates that U.S. equity market returns constitute the dominant predictors of VIX values within the model. The remaining macro-financial variables appear to have comparatively lower, yet non-zero, contributions. Specifically, EUR_ret records an importance of 0.00013, Oil_ret 0.00007, GBP_ret 0.00004, and Gold_ret 0.00001. In contrast, TNX_ret presents a slightly negative importance score of -0.00004, suggesting a negligible and potentially destabilizing contribution to predictive performance.

Table 33. Permutation Importance in the Final Random Forest Model

Variable	Permutation Importance
GSPC_ret	0.00316
IXIC_ret	0.00142
EUR_ret	0.00013
Oil_ret	0.00007
GBP_ret	0.00004
Gold_ret	0.00001
TNX_ret	-0.00004

XGBoost Model

The Extreme Gradient Boosting model, XGBoost, is employed as the primary boosting based machine learning approach using a rolling one step ahead forecasting scheme for VIX_ret_log. A fixed training window is implemented in order to ensure comparability with the previously estimated econometric and ensemble models, while the naive benchmark model is also included as a reference point.

According to the results presented in Table 34, the naive benchmark records RMSE = 0.1111 and MAE = 0.0778. In contrast, the XGBoost model achieves RMSE = 0.0486 and MAE = 0.0334. This implies an approximate 56.3% reduction in RMSE relative to the naive benchmark, indicating a substantial improvement in forecasting accuracy. A similarly strong reduction is observed in MAE, further confirming the superior predictive performance of the boosting model compared to all previously examined specifications. The findings suggest that XGBoost is capable of effectively capturing complex nonlinear interactions and dynamics inherent in volatility behavior.

Table 34. Rolling One-Step-Ahead Forecast Evaluation for the XGBoost Model

Model	RMSE	MAE
Naive (y[t-1])	0.1111	0.0778
XGBoost	0.0486	0.0334

Continuing, Table 35 presents the last two rolling forecasts of the XGBoost model alongside the corresponding actual values and the naive benchmark forecasts in order to facilitate direct comparison. The evidence indicates that the XGBoost predictions exhibit a high degree of responsiveness to abrupt volatility movements and demonstrate superior tracking ability of realized VIX returns relative to the naive benchmark. More specifically, on 2025-11-13 the actual return reached 0.1329, while the naive forecast was only 0.0132. In contrast, the XGBoost forecast amounted to 0.0988, showing a substantially closer alignment with the observed volatility spike. A similar pattern is observed on 2025-11-10, where the actual return was -0.0807. The XGBoost forecast of -0.0629 approximated the realized value much more closely compared to the naive forecast of -0.0217. These results further reinforce the ability of the boosting model to adjust more effectively to sudden shifts in volatility dynamics.

Table 35. Rolling XGBoost Forecasts for the Last 20 Observations

Date	Actual	Naive	XGBoost
16/10/2025	0.2039	-0.0082	0.0384
21/10/2025	-0.0199	0.2039	-0.0106
22/10/2025	0.0400	-0.0199	0.0237
23/10/2025	-0.0725	0.0400	-0.0273
28/10/2025	0.0391	-0.0725	-0.0240
29/10/2025	0.0300	0.0391	-0.0159
30/10/2025	-0.0006	0.0300	0.0570
31/10/2025	0.0308	-0.0006	-0.0200
3/11/2025	-0.0156	0.0308	-0.0175
4/11/2025	0.1012	-0.0156	0.0713
5/11/2025	-0.0535	0.1012	-0.0229
6/11/2025	0.0794	-0.0535	0.0714
7/11/2025	-0.0218	0.0794	-0.0142
10/11/2025	-0.0807	-0.0218	-0.063
11/11/2025	-0.0183	-0.0807	-0.0136
12/11/2025	0.0132	-0.0183	-0.0099
13/11/2025	0.1329	0.0132	0.0988
14/11/2025	-0.0085	0.1329	-0.0108
17/11/2025	0.1209	-0.0085	0.0582
18/11/2025	0.0982	0.1209	0.0538

Additionally, feature importance is further examined through gain and cover frequency metrics, which are presented in Table 36. The results indicate that the highest contribution belongs to GSPC_ret, with Gain = 0.6397, followed by IXIC_ret with Gain = 0.1914. These values suggest that U.S. equity market returns account for the largest share of the model's explanatory power. The remaining macro-financial variables exhibit smaller but non-negligible contributions in terms of gain and cover frequency. Overall, the findings reinforce the dominant role of U.S. equity market returns in explaining VIX dynamics, while the rest of the macro-financial variables operate in a complementary manner without demonstrating equivalent predictive strength within the boosting framework.

Table 36. Feature Importance in the Final XGBoost Model (Gain, Cover, Frequency)

Variable	Gain	Cover	Frequency
GSPC_ret	0.6397	0.3343	0.2112
IXIC_ret	0.1914	0.20393	0.1458
EUR_ret	0.0364	0.0856	0.1228
Oil_ret	0.0360	0.0739	0.1298
GBP_ret	0.0337	0.0969	0.1305
TNX_ret	0.0323	0.1292	0.1322
Gold_ret	0.0302	0.0758	0.1273

Comparative Evaluation of Forecasting Models

7.1 Summary Comparison of Model Performance

In Table 37 the forecasting performance of all the models that were used based on rolling one step ahead predictions are presented. It is also summarized through the forecasting measures RMSE and MAE. Additionally, the naive model is once again used as a benchmark, with all the other models being evaluated regarding their improvements. Evidence shows that the relative improvement through the reduction of RMSE compared to the naive model reveals an important difference between the RMSE of the naive model and the RMSE of all the other models. It seems to be systematic when it comes to forecasting performance regarding the model complexity and flexibility, which appears to increase compared to the naive model. However, the nonlinear machine learning models seem to present lower forecast error.

Table 37. Summary Comparison of Forecasting Performance Across Models

Model	Category	RMSE	MAE	RMSE Improvement vs Naive
Naive	Benchmark	0.1111	0.0778	0.00%
ARIMA(1,0,1)	Linear econometric	0.0756	0.0529	31.90%
VAR(1)	Linear econometric	0.0759	0.0531	31.70%
Dynamic Factor Model	Linear factor-based	0.0521	0.035	53.10%
LASSO	Regularized linear regression	0.0518	0.0349	53.40%
Elastic Net	Regularized linear regression	0.0515	0.0348	53.60%
CART	Non-linear decision tree	0.0508	0.0355	54.30%
Random Forest	Non-linear ensemble	0.0485	0.0337	56.40%
XGBoost	Boosting non-linear	0.0486	0.0335	56.30%

7.2 Comparison of Linear Econometric Models

The models ARIMA and VAR presented an important improvement compared to the naïve benchmark, since the RMSE was reduced by 32%. So, the VIX incorporates predictable linear dependence both on past values but also on the financial variables. Although the two models presented almost identical performances, which means that the use of the macrofinancial variables in the linear framework did not provide any increase in the predictivity of the models. So, when the linear autoregressive dynamics are captured, other variables cannot contribute in a statistically significant way, especially under linear specifications.

In contrast, the dynamic factor model seems to have the best performance, reaching an RMSE reduction of 53%. This extremely good performance reflects the ability of the model to summarize the information of many financial variables into a smaller number of factors. So, the noise is reduced while the signal to noise ratio in the forecasting is enhanced. This suggests that the volatility of the market is mostly affected by a number of dynamic forces and can be extracted in an effective way when the dimensions are reduced with proper techniques.

7.3 Comparison of Regularized Linear Models

Regarding the LASSO and the elastic net models, it was revealed that the forecasting performance was improved compared to the econometric models. More specifically, the performance was comparable to those of the dynamic factor model, since the RMSE value was reduced by almost 53%. This improvement is mostly driven by the regularization mechanism, which conducts proper selection of the variables in order to shrink their relevant coefficients close to zero, so overfitting and multicollinearity are also reduced. Additionally, the variables of S&P 500 and NASDAQ returns were consistently chosen as the main and most important predictors, which also confirms that the US equity market primarily acts as a transmission mechanism of uncertainty toward the VIX value. As for the exclusion of the other macrofinancial variables, which were less relevant, it is clear that this enhances both the stability and the predictive accuracy of the models.

7.4 Comparison of Non-Linear Machine Learning Models

As for the nonlinear machine learning models, which were CART, random forest, and XGBoost, they presented the highest level of forecasting performance among all the models that were examined. Most specifically, random forest and XGBoost presented the largest RMSE reductions, which were higher than 56% compared to the naïve benchmark. It seems that the nonlinear models' efficiency is connected to the ability they have to capture complex interactions, asymmetries, and nonlinear connections that cannot be represented efficiently with linear frameworks. Additionally, the financial volatility dynamics are connected with threshold effects and directional structures that can be modeled efficiently with the help of tree-based ensemble techniques.

7.5 Overall Interpretation and Comparative Conclusions

Through the comparison between the models, it was revealed that the forecasting performance differs a lot. The simple linear econometrics models offer some improvements compared to the naïve benchmark, but they are limited compared to other models that offer greater flexibility and accuracy.

Through the analysis, it is revealed that part of the VIX dynamics can be fully investigated through linear relationships, however an important part of them is connected and more impactful in nonlinear structures and interaction effects. Additionally, the important role of US equity index returns that appears to be confirmed in all the models highlights the importance of stock market dynamics when it comes to volatility. The performance of ensemble based methods that was better than the rest makes clear that the use of nonlinear machine learning techniques can offer strong advantages regarding forecasting financial market volatility. Overall, the results reveal that the machine learning models, and especially random forest and XGBoost, are the most effective when it comes to predicting VIX fluctuations, since they can capture the complex and nonlinear nature of the financial market more successfully than the rest of the models.

Chapter 8. Conclusions

The present analysis focuses on investigating the forecasting performance of various econometric and machine learning models in order to predict the daily logarithmic returns of the VIX index using a unified rolling one step ahead framework with a fixed identification window created in a total of 1,200 observations. It was made clear that the analysis ensures strict out of sample evaluation, while it was fully compatible across the models in order to properly compare them.

Through the empirical results, it was made clear that there is a specific hierarchy regarding the predictive power that the models present. The classical linear econometric models like ARIMA and VAR improved the naïve benchmark results by reducing the forecast error by 32%. So, these results confirm that short term linear dependence in VIX dynamics exists, although the limited incremental gains by adding other macro-financial variables in the linear framework highlight that linear models are not sufficient in capturing the complexity of the volatility movements.

Continuing, the Dynamic Factor Model and the LASSO and Elastic Net models, which represent the regularized regression approaches, presented a substantial gain when it came to their predictive accuracy, which led to a lower RMSE over 53%. According to the results, the importance of dimensionality reduction and variable selection techniques was revealed in order to reduce the noise and also isolate the most important sources of macrofinancial comovement. The models consistently identified S&P 500 and NASDAQ returns as the major variables, scoring the central role of US equity markets when it comes to the transmission of information and uncertainty to the VIX values.

As for the strongest forecasting performance, it seems to belong to the nonlinear ensemble models and more specifically the random forest and XGBoost, which reduced the RMSE value by more than 56% relative compared to the results of the naïve benchmark. This indicates that the model's ability to capture the complexity of the interaction effects, asymmetries, and also the nonlinear correlations that are present in financial volatility dynamics is highly important. Also, it is indicated that an important portion of the information is connected to nonlinear structures that cannot be effectively investigated using linear specifications.

Between all the models that were applied, US equity index returns seem to be the most important driver of the VIX fluctuations, with interest rates, commodities, and exchange rates contributing but in a negligible dimension, since the two major equity indices are the most supportive channel through which uncertainty is reflected when it comes to volatility. The study revealed that modern machine learning methods, especially the ensemble-based techniques, present important advantages regarding the forecasting of financial market volatility. However, the linear econometric models provide information and are interpretable. The total empirical evidence highlights that the nonlinear machine learning methods present superior predictive accuracy, especially when it comes to environments that face structural complexity and important interaction effects.

The present findings present certain practical implications regarding risk management, derivative pricing, and volatility trading methods. Also, the use of machine learning models in volatility forecasting frameworks can enhance predictive performance in short term periods of time, but also lead to better decision making under periods with high uncertainty. Future research could focus on the use of alternative volatility measures and the exploration of multi step ahead forecasts. It could also compare formal forecast tests like the Diebold Mariano procedure and examine the stability of the models across different market regimes.

Bibliography

Bachmair, K., & Schmitz, N. (2025). *Forecasting macro with finance*. Cambridge Working Papers in Economics (No. 2574). University of Cambridge.

Goulet-Coulombe, P. (2024). *The macroeconomy as a random forest*. Journal of Applied Econometrics, 39(3), 401–421. https://doi.org/10.1002/jae.3030

Goel, H., & Sharma, M. (2025). *Hybrid machine learning models for global macroeconomic forecasting: A case study of G20 economies*. International Journal of Progressive Research in Engineering Management and Science, 5(11), 400–407. https://doi.org/10.58257/IJPREMS44624

Li, X., Liang, C., & Ma, F. (2025). *Forecasting stock market volatility with MS-MIDAS-LASSO*. Annals of Operations Research, 352(3), 613–652. https://doi.org/10.1007/s10479-022-04716-1

Lopez-Estrada, D., Joaqui-Barandica, A., & Orozco-Ceron, J. (2025). *Hybrid VAR–XGBoost modelling for electricity tariffs*. Technologies, 13(11), 495. https://doi.org/10.3390/technologies13110495

Mamedli, M., & Shibitov, D. (2021). *Machine learning real-time CPI forecasting*. In IFC–Bank of Italy Workshop "Machine Learning in Central Banking" (pp. 1–18). Bank for International Settlements.

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2019). *Random forest-based feature selection of macroeconomic variables for stock market prediction: A case study of the Ghana Stock Exchange*. American Journal of Applied Sciences, 16(7), 200–212. https://doi.org/10.3844/ajassp.2019.200.212

Patsiarikas, M., Papageorgiou, G., & Tjortjis, C. (2025). *Using machine learning on macroeconomic, technical, and sentiment indicators for stock market forecasting*. Information, 16(7), 584. https://doi.org/10.3390/info16070584

Pirayesh Neghab, D., Cevik, M., Wahab, M. I. M., & Basar, A. (2025). *Explaining exchange-rate forecasts with macroeconomic fundamentals using interpretive machine learning*. Computational Economics, 65(4), 1857–1899. https://doi.org/10.1007/s10614-024-10617-1

Sheng, Z., Liu, Y., Hu, H., & Liu, L. (2025). *A two-stage forecasting framework with LASSO-based feature selection and a non-stationary Autoformer*. Electronics, 14(10), 2059. https://doi.org/10.3390/electronics14102059

Stempien, D., & Slepaczuk, R. (2025). *Hybrid models for financial forecasting: Combining econometric, machine learning, and deep learning models*. arXiv. https://doi.org/10.48550/arXiv.2505.19617

Yang, J., Wang, Y., & Li, X. (2022). *Prediction of stock price direction using the LASSO-LSTM model combines technical indicators and financial sentiment analysis*. PeerJ Computer Science, 8, e1148. https://doi.org/10.7717/peerj-cs.1148