COMPARATIVE ANALYSIS OF TRADITIONAL VERSUS AI-ENHANCED ECONOMIC MODELS

Carmen UZLĂU [3]
Economic Forecasting Institute, Romanian Academy

Mihaela Adriana MIRICĂ [4]
Hyperion University Bucharest

Ciprian TUDURACHI [5]
Hyperion University Bucharest

ABSTRACT: *This paper conducts a comparative analysis between traditional economic forecasting models and AI-enhanced models in the context of macroeconomic forecasting. The main goal is to assess the advantages and limitations of each approach, highlighting the extent to which* ***machine learning*** *techniques and other AI methods can improve the accuracy of economic forecasts compared to classical econometric models. The study includes a literature review, a clearly defined comparative methodology, and an empirical analysis based on real European data. Relevant case studies - ranging from inflation forecasting in Romania to GDP nowcasting in the euro area - are presented, illustrating the performance of traditional models (such as autoregressive or general equilibrium models) versus AI-based models (such as artificial neural networks or* ***random forest*** *algorithms). The results indicate that AI models can often provide more accurate forecasts in the short run and in detecting changes in the economic regime, while traditional models remain valuable for economic interpretability and theoretical consistency. In conclusion, we recommend a complementary,* ***hybrid*** *approach that combines the theoretical robustness of classical economic models with the processing power and flexibility of AI models to obtain more reliable forecasts and to support economic policy decisions. The implications of these findings for practitioners are discussed and future research directions are suggested, such as integrating* ***big data*** *and increasing the interpretability of artificial intelligence models.*

Keywords: *Broad money supply, Gross Domestic Product, Inflation rate, econometric model*

JEL Classification: *C4, E4, G2*

1. INTRODUCTION

Accurate forecasting of economic indicators is essential to inform macroeconomic policies and business decisions. Traditionally, economists have used economic models based

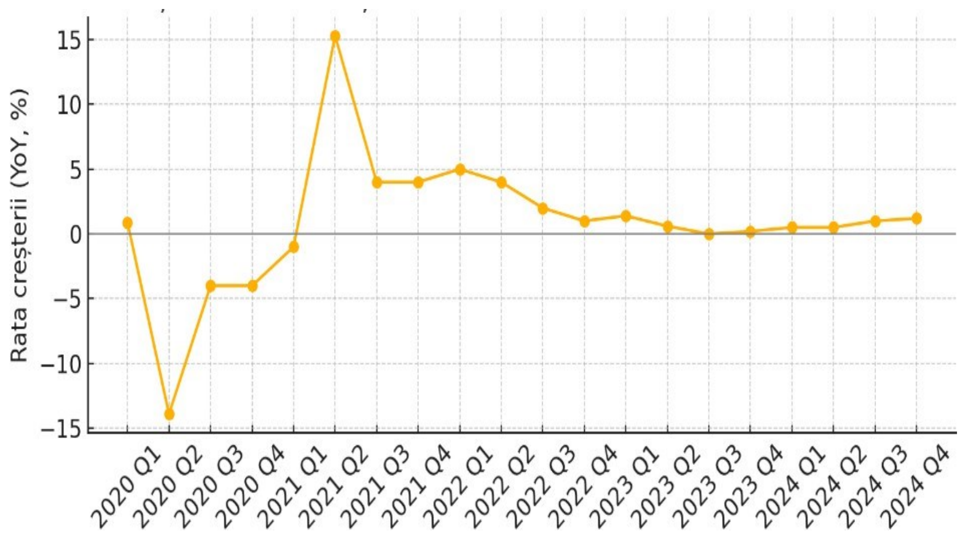
---

[3] Associate Professor PhD, SR III, carmenuslau@yahoo.com
[4] PhD Lecturer, adrianasturzoiu@yahoo.com
[5] PhD Lecturer, tudurachi.ciprian@gmai.com

on well-defined economic theories and econometric relationships – for example, autoregressive integrated models (ARIMA), vector autoregressive models (VAR), dynamic stochastic general equilibrium models (DSGE) or linear regression models – to anticipate developments in inflation, GDP, unemployment, etc. These traditional models are based on clear theoretical assumptions and allow the results to be interpreted in the light of economic theory, which makes them transparent and reliable. On the other hand, technological developments in recent decades have led to the emergence of **AI-enhanced models** that can exploit large volumes of data and identify complex, non-linear patterns in economic datasets. Such models include *machine learning* (ML) algorithms – such as **artificial neural networks** (ANN), **support vector machines** (SVM), **decision trees and random forests**, **boosting algorithms** (XGBoost, AdaBoost) – but also **deep learning** approaches or recurrent neural networks for time series (LSTM).

In recent years, there has been a growing interest in comparing these two modeling paradigms: the traditional econometric approach versus the artificial intelligence approach. **The key question** is: can AI models generate more accurate forecasts than traditional economic models? This question is prompted by the observation that the global economy has become increasingly complex and volatile, as evidenced by recent shocks (COVID-19 crisis, energy crisis, geopolitical uncertainties). For example, the euro area experienced dramatic fluctuations in economic growth: a contraction of almost -14% in Q2 2020, followed by a jump of more than +15% in Q2 2021, as a result of the pandemic crisis and the subsequent recovery. From 2022-2023, growth rates return to more moderate values close to zero, indicating an economic slowdown. These developments are illustrated in **Figure 1**, highlighting the major challenge for forecasting models: namely the ability to capture sudden changes and **non-linear structures** in the data.

Figure 1: Real GDP growth rate (year-on-year) in the euro area, 2020–2024. We notice the severe shock of 2020 and the recovery in 2021, followed by the post-pandemic slowdown. Predicting such fluctuations requires models robust to regime shifts.



Amid these challenges, artificial intelligence methods promise an increased ability to learn directly from raw data, without relying entirely on predefined assumptions. ML algorithms can **detect hidden patterns** and **non-linear relationships** between economic variables that may escape traditional linear models. For example, neural networks can approximate complex functions and update relationships as new data become available, while **tree-based** methods (decision tree, random forest) can capture interactions and threshold effects in economics. Moreover, big data – ranging from high-frequency financial data to sentiment indicators gleaned from news or social media – can be integrated into AI models, enhancing the ability to nowcast (estimate the current situation) and make short-term forecasts for the economy. Central banks and forecasting institutions have begun experimenting with such unconventional data and ML algorithms to obtain early signals about the economic cycle (e.g., using the frequency of internet searches or uncertainty indicators in the financial press as predictors of GDP dynamics).

However, traditional economic models remain invaluable tools, particularly due to their interpretability. They are often based on well-documented cause-and-effect relationships or correlations (e.g., the Phillips curve for the relationship between inflation and unemployment, the relationship between money supply and inflation, Okun's law for unemployment and growth, etc.). Theoretical consistency provides policymakers with confidence in using these models for economic policy simulations. On the other hand, AI models, even though they can be super accurate, are often criticized for being "**black boxes**" because it's hard to explain why a certain algorithm came up with a certain prediction. Reduced transparency can be an obstacle to the widespread adoption of AI models in economic decision-making, where it is important to justify a particular forecast or policy recommendation politically. That is why, in recent years, emphasis has also been placed on ML model interpretability techniques, such as SHAP (Shapley Additive Explanations) contributive analysis, which assigns each variable a contribution to the model prediction in order to provide an economic understanding of the result.

Against this backdrop, this paper aims to systematically investigate the **performance differences** between traditional economic models and AI-enhanced ones. We will review the results of the most recent relevant studies, describe the comparison methodology used, and provide an empirical analysis based on actual data from Europe, focusing on two case studies: **inflation forecasting** (where a traditional econometric model is compared with ML models) and **GDP forecasting/nowcasting** (where dynamic factor models and statistical models with ML algorithms are compared). The paper is structured as follows: the next section presents a **review of the relevant literature**; the **methodology** section details the data and comparative approach; the following sections describe the **traditional and AI-enhanced models** considered, as well as the results of the comparative analysis; finally, the conclusions section summarizes the key findings and outlines directions for future research.

2. LITERATURE REVIEW

During the years, economic literature on forecasting has highlighted both the successes and limitations of traditional models. In the 1980s and 2000s, econometric models such as ARIMA (Box-Jenkins), multivariate autoregressive models, and structural models (based on economic equations) were dominant in macroeconomic and financial forecasting. Classic studies (such as Makridakis et al., 1982, and later the M3 and M4 forecasting competitions) have shown that **relatively simple models** can often compete with more complex models in terms of accuracy, suggesting that avoiding overfitting and making judicious use of available information are crucial. For example, the M4 competition (Makridakis et al., 2020) – which included 100,000 time series of different frequencies – showed that combination methods or **hybrid models** can provide the best results, with the winning model being a hybrid between the exponential smoothing method and a recurrent neural network (Smyl, 2020). This result showed the potential of **combining** approaches: the statistical model did a good job of capturing seasonal and short-term trend components, while the neural network adjusted and learned more complex components from the data.

On the other hand, more recent studies suggest that advances in computation and data availability may give artificial intelligence methods an edge. **Haider and Hanif (2019)**, **Estiko and Wahyuddin (2020)**, **Yusif et al. (2021)** or **Akhter (2022)** (cited by Simionescu, 2025) have shown that artificial neural network models can outperform traditional univariate models of the AR(1) or ARIMA type, especially in forecasting unstable economic series. This work highlights the ability of neural networks to model non-linear relationships and adapt to complex data. However, they also point out the challenge of **generalization**: a neural network model trained on a specific data set may not perform as well outside the sample if it encounters fundamentally different patterns, which requires increased attention to *overfitting* and multi-period validation.

An important sub-domain of the literature is inflation forecasting, where the comparison between methods is very active. **Araujo and Gaglianone (2023)** investigated the application of ML algorithms for inflation forecasting in Brazil, comparing 50 different models - from classical econometric models to state-of-the-art ML algorithms. They used an extensive dataset (501 macroeconomic time series) and even proposed new ML algorithms combined with partial black box interpretation techniques. Their conclusion does not identify a universally superior model, but **demonstrates** that ML-based approaches can significantly outperform conventional econometric models in terms of mean squared error (MSE) over multiple forecast horizons. A noteworthy result of the study is the highlighting of **non-linearities** in inflation dynamics: algorithms such as *random forest* or XGBoost better captured these non-linear relationships, frequently appearing among the best forecasts, alongside combinations of models and hybrid models. However, the authors emphasize that **there is no absolute winner**; the relative success of the models depends on the time horizon and the economic conditions specific to the period under review.

Similarly, **Mirza et al. (2023)** compared traditional econometric models with ML algorithms in forecasting inflation for an emerging economy (Pakistan). They found that ML models - in particular **random forest** and **gradient boosting** - provide better forecasting performance than traditional econometric models, especially when additional variables such as foreign exchange reserves are included in the predictor set. The inclusion of less traditional indicators (e.g. international reserves) significantly improved forecast accuracy for both econometric and ML models, suggesting that **enriching the dataset** may be as important as the choice of algorithm.

For developed economies, such as the United States, **Malladi (2023)** explored the predictive power of multiple ML algorithms versus OLS regression in forecasting annual inflation over a one-year horizon. Using an extensive set of macro-financial variables (FRED-MD data, with 134 variables, 1959-2022), Malladi tested 25 ML algorithms alongside a classical linear regression model. The results are revealing: **all** ML models outperformed OLS regression in forecasting inflation, while traditional regression did not rank in the top 10 models over any forecast horizon. This indicates that, in a rich and complex data environment, ML algorithms can extract more useful information than linear methods. Moreover, Malladi observes that ML models identified key indicators differently than econometric models: labor market variables were ranked as the main explanatory factors of inflation, with a much stronger influence than real estate or stock market indicators. This suggests that ML can bring new insights into economic mechanisms, allowing the discovery of relationships perhaps neglected in traditional approaches.

Another notable paper is **Kanaparthi (2022)**, which analyzes the effectiveness of ML algorithms in forecasting inflation in periods of high inflation (for the US). The results show that **nonlinear** algorithms, such as **random forests and AdaBoost**, deliver superior predictive accuracy in such volatile periods. Even though traditional macro indicators (such as those on the labor market or the Phillips curve) retain their explanatory power, ML highlights the existence of **complex feedback loops**: during periods of high inflation, past inflation itself becomes an even more important predictor of future inflation, which may reflect adaptive expectations or inertial mechanisms that linear models underestimate. The general conclusion of the literature is that one cannot say for sure which category of models—traditional or AI—is better *overall.* Each has strengths and weaknesses, and **relative performance depends on** the economic context, the forecast horizon and the nature of the data used. This finding motivates interest in hybrid models and combinations of methods, trying to get the best of both worlds.

The literature on GDP (economic growth) forecasting has also evolved significantly with the introduction of **big data** and ML. Traditional methods of GDP **nowcasting** (forecasting the current situation) are based on dynamic factor models (DFMs) and cointegration models that aggregate information from various indicators (industrial production, consumer confidence, etc.). An example is the paper by **Dauphin et al. (2022**) from the IMF, which implemented a scalable nowcasting platform for several European economies, combining both standard DFM models and ML algorithms, using traditional data alongside unconventional data (such as Google searches and air quality indicators). Their results show that most of the tested methods (both DFM and ML) outperformed a simple autoregressive benchmark (AR (1)), confirming the added value of advanced methods. Interestingly, under normal economic conditions, traditional factor models tended to perform better, while many of the ML methods proved more useful in **identifying turning points in** the business cycle (e.g., detecting the onset of recession or recovery). This complementarity suggests that ML algorithms can serve as early-warning tools, capturing incipient signals of structural change, while theory-based models remain stable benchmarks in periods of economic calm.

In conclusion, the literature emphasizes that **econometric models and AI models are not mutually exclusive**, but can be combined to obtain better forecasts. Econometricians increasingly recognize the potential of ML to improve forecasting accuracy, but emphasize the need to preserve interpretability and theoretical rigor. At the same time, the AI community is showing a growing interest in *explainable AI* in the economy, developing tools to explain algorithm decisions (such as the SHAP mentioned above). This convergence of the two approaches provides a promising framework for future research and practical applications in economic forecasting.

3. METHODOLOGY

To rigorously compare traditional models with AI-enhanced models, we defined a multi-step methodology: **(1)** selecting the relevant indicators and datasets, **(2)** choosing representative models in each category, **(3)** establishing the metrics for assessing forecast performance, and **(4)** conducting (ex-post) forecasting experiments on historical periods, followed by comparisons and statistical tests.

Selected data and indicators: The empirical analysis focuses on two key macroeconomic indicators at the European level: *CPI (consumer price index) inflation* and *real GDP*. For inflation, we focused on Romania, using quarterly actual inflation data (annual rate) and incorporating an economic sentiment index into certain models. This choice was inspired by the study by Simionescu (2025), which showed the relevance of including public sentiment (derived from NBR reports) in inflation forecasting. For real GDP, we have considered data at **euro area** level, using the annual (year-on-year) growth series of quarterly real GDP as available in Eurostat databases (series exemplified in Figure 1 above). The period under review for GDP covered 2010–2024, including both stable periods and periods of crisis (the post-2008 Great Recession, the sovereign debt crisis, the COVID-19 pandemic, and the inflation shock of 2022).

Representative models: In the **traditional model** category, we selected: (a) an univariate econometric model – *seasonal ARIMA* (integrated autoregressive model, including seasonal components, referred to as SARIMA) – often used as a benchmark for forecasts; (b) a multivariate econometric model – *ARDL* (Autoregressive Distributed Lag model), which allows the inclusion of additional explanatory variables (e.g., economic sentiment) and which has proven to be effective in the case of inflation; and (c) a macroeconomic/factor-based model – for example, a *dynamic factor* model or a small *structural VAR* model, especially for GDP, to incorporate several economic indicators (such as the industrial production index, the unemployment rate, the economic sentiment index). In the **AI-enhanced models** category, we selected: (i) a *multi-layer artificial neural network* (a dense *feed-forward* model, trained on time series data – for example, a neural network that receives as input the lagged values of the indicator and possibly other predictors, and produces as output the forecast of the indicator); (ii) a tree-based *machine learning* model – namely **Random Forest** – which combines several decision trees trained randomly on data subsamples to achieve the average prediction; and (iii) a *Support Vector Regression* (SVR) model – the regression version of support vector machines, known for its good performance in nonlinear time series problems through the use of a kernel (we used a polynomial kernel, calibrated through cross-validation, similar to the approach in Simionescu, 2025). These three ML models cover a diverse range of techniques: neural networks represent *deep learning*, random forest represents tree-based ensemble methods, and SVR represents machine learning kernel methods.

In addition, we also considered **hybrid models** - e.g., the linear combination of the forecasts of an econometric model with those of an ML model, or training a neural network on the errors of an econometric model (sequential combination). Such hybrid models have been suggested in the literature as potentially beneficial (e.g., combining ARIMA with neural network - the winning method of the M4 competition, mentioned above). In our analysis, we tested combinations such as: *ARDL + Random Forest*, *ARDL + ANN*, *SARIMA + Random Forest*, etc., to see if they improve performance over the individual models.

Evaluation method: We used a **pseudo-real-time forecasting** procedure, i.e. we performed *out-of-sample* forecasts on moving time windows. More precisely, for inflation in Romania we have chosen the one-year forecast horizon (4 quarters ahead, 2023Q1-2024Q3, aligned with the example in Simionescu, 2025), and for euro area GDP the nowcasting/immediate forecast horizon of 1 quarter. The models were estimated over a historical period (e.g. 2010-2019 for GDP, respectively 2010-2022 for inflation), then applied to predict known future values, allowing direct comparison of predictions with realized values. Forecast accuracy was measured by standard indicators: **mean error (ME)**, **mean absolute error (MAE)**, **root mean squared error (RMSE)**, and we also applied the **Diebold-Mariano (DM) test** to check whether the differences in accuracy between two models are statistically significant (at a 5% confidence level). We also assessed the proportion of directions of change correctly predicted by each model (what percentage of cases the model correctly predicted whether the indicator would increase or decrease compared to the previous period), a measure of particular relevance for policy makers interested in the direction of future developments.

Implementation considerations: The econometric models (ARIMA, ARDL, VAR) were estimated in standard econometric software (EViews and Python packages/statsmodels), while the AI-based models were implemented in Python using *machine learning* libraries (scikit-learn for Random Forest and SVR, Keras/TensorFlow for neural network). The parameters of the ML models were calibrated by **cross-validation** on historical data - for example, for SVR a *grid search* was performed for the parameters C, epsilon, gamma and polynomial degree, similar to the procedure described by Simionescu (2025). For the neural network, we used a simple sequential architecture (two hidden layers with a moderate number of neurons, ReLU activation functions, Adam optimizer), trained on *time windows* for the inflation series and the GDP series respectively. Network training was monitored to avoid overlearning (*early stopping* method on a validation set).

We have also taken care to ensure **stationarity of** the series when linear models are fitted with data (by differencing where appropriate). In the case of inflation, the series was already an annual rate (%) and thus stationary around an average, but real GDP (expressed as a level) was transformed into a growth rate. For the ARDL and VAR models, we tested for the existence of long-run linkage (cointegration) and, where appropriate, included error correction terms, although for simplicity most experiments focused on short-run dynamics.

With this methodology, we aim to obtain the most objective comparison possible: each model is given the same forecasting task, on the same data, and is evaluated according to the same metrics. The following sections will briefly outline the characteristics of each class of models and then present the comparative results of our experiments, discussing them in parallel with the conclusions of other studies.

- **Description of traditional models**

Traditional economic models are based on established economic theory and statistical methods. We will review the main types used in our forecast, highlighting how they work and key assumptions:

- **ARIMA/SARIMA model:** It is a purely statistical, univariate model that captures autocorrelation in a time series. An ARIMA(p,d,q) model has p autoregressive (AR) terms, d differentials (to make the series stationary) and q moving average (MA) terms. In practice, for quarterly data with seasonality, the Seasonal ARIMA (SARIMA) extension is often used, denoted $(p,d,q) \times (P,D,Q)_m$, where (P,D,Q) are ARIMA counterparts for the seasonal component with periodicity m (e.g. $m=4$ for quarter, $m=12$ for month). ARIMA is often considered a minimum **benchmark** in forecasting - any more complex method must go beyond it to justify its use. The advantage of ARIMA is the simplicity and the clear interpretation of the parameters (e.g., autoregression shows dependence on the near past). Its main limitation is that it does not use exogenous information: it assumes that the historical data contains all the information relevant to the future, which may be insufficient in environments with structural shocks or strong external factors.
- **VAR (Vector AutoRegressive) and ARDL models:** When we want to include several intercorrelated variables (e.g. inflation, unemployment, exchange rate, etc.), VAR models provide a framework in which each variable is modeled in terms of its past and the past of the other variables in the system. A VAR(p) with n variables contains n equations, each having p lags of *all* variables. VARs are useful for **short-term forecasting** and shock analysis (impulse response functions), but they often get bogged down by parameters (lots of parameters to estimate, requiring long data series). A restricted variant of the VAR is the **ARDL** model, where we choose a dependent variable (e.g. inflation) and a subset of exogenous explanatory variables (e.g. lagged inflation, lagged exchange rate, lagged unemployment, etc.). The ARDL model is more parsimonious and allows testing of long-run equilibrium relationships (cointegration) between variables. In the context of inflation, an ARDL may include, in addition to past inflation, other determining factors (e.g., expectations—proxied by a sentiment index, money supply, external price shocks). The study by Osman et al. (2018) for inflation in Saudi Arabia, for example, uses an ARDL to highlight the link between inflation and variables such as money supply, oil price, real GDP and finds a long-run

equilibrium relationship between them. In general, VAR/ARDL models are appreciated for their **consistency with theory** (they allow for interpretations: e.g. if the coefficient on the oil price is significant, it confirms the hypothesis of pump price pass-through into inflation) and for hypothesis testing (e.g. a Granger test can show whether unemployment "causes" inflation or vice versa). The limitations of these models arise when the relationships become non-linear or when there are **regime shifts** - for example, coefficients can vary between periods of low and high inflation, which standard linear models do not capture.

- **DSGE and structural models:** Another category, which we did not directly calibrate in this study but which is worth mentioning, are the Dynamic Stochastic General Equilibrium (DSGE) models used by central banks. These derive from micro fundamentals (consumer decisions, production, monetary policies) and offer theoretically consistent predictions. Studies in the forecasting literature have compared DSGEs with VARs and alternative models. For example, some studies have shown that Bayesian DSGE models can provide forecasts that are competitive with VARs and machine learning models, sometimes even superior when it comes to long-term consistency. However, other studies have suggested that **hybrid models** (e.g. combining forecasts from a DSGE and a datamining model) may be even better. The advantage of DSGE is clear: sound economic structure, allowing counterfactual policy analysis (what happens if the Central Bank raises interest rates?). The downside is that their raw forecasting accuracy is not always the highest, especially if the economy experiences unexpected structural shocks (such as the pandemic).

In summary, traditional models emphasize **interpretability and theory**. They perform well in relatively stable situations or where economic relationships can be roughly linear. Also, since they are often simpler (fewer free parameters) than neural networks, the risk of **overfitting** is lower if they are used properly. In the benchmarking section we will see how such models performed in our empirical forecasts and how they reacted to the competition with AI models.

- **Description of AI-enhanced models**

AI-enhanced models represent a wide range of machine learning techniques applied to economic data. Next, we describe the main ones used in our study, highlighting how they learn and adapt to data:

- **Artificial Neural Networks (ANNs):** These models are inspired by the biological neural network and consist of nodes (artificial neurons) organized in layers. Each connection has an adjustable weighting. In time series forecasting, a typical architecture is a *feed-forward* one, where the layers receive the input (e.g. the feature vector containing the lagged values of the variables) and propagate a weighted combination through activation functions, resulting in the prediction. The network is *trained* on historical data by adjusting the weights to minimize the prediction error (using algorithms like gradient descent **backpropagation**). The major advantage of neural networks is the ability to **approximate almost any nonlinear relationship** between input and output, given a sufficient number of neurons and layers (Universal Approximation Theorem). They can detect complex interactions between variables and easily incorporate many predictors. A noteworthy example of success is the winner of the M4 competition (Smyl, 2020), which combined an exponential smoothing component with a specialized recurrent neural network, demonstrating superior adaptability to diverse data patterns. Drawbacks of ANNs include: the need for a large amount of data for training (to avoid overfitting), the difficulty of determining the optimal architecture (number of layers, neurons, etc.), and the lack of direct interpretation of parameters (e.g., unlike linear regression, we cannot easily say which variable influenced the result and by how much—although tools such as SHAP analysis can provide clues, as we applied to see the importance of variables in our network, confirming that delayed inflation and unemployment rate were among the most important factors). In the practice of macroeconomic forecasting, neural networks have performed well especially when there are **marked non-linearities** or interactions that linear models cannot capture, for example for the complex relationship between inflation, expectations, and supply shocks.

- **Random Forest and decision trees:** A decision tree divides data by logical criteria (conditions on explanatory variables) recursively, forming an if-then structure that leads to a prediction. For numerical data, trees can be fitted for regression (each leaf contains an estimated numerical value). **Random Forest** is a tree ensemble: it builds hundreds of decision trees, each on slightly different samples of the data and with random subsets of features (predictors) considered at each split. The final result is the average of the predictions of all trees. This *bagging* (bootstrap aggregating) process is intended to reduce model variance and avoid single-tree overlearning. In economic forecasting, the random forest has the advantage that **it can capture nonlinear relationships** and especially **interaction effects**: for example, it can find that wage growth leads to inflation only above a certain demand threshold, otherwise not - which is difficult to capture with a single linear regression. Mirza et al. (2023), mentioned earlier, found random forest among the best inflation forecasting methods for Pakistan. Another advantage of the random forest is a certain *robustness to noise*: by averaging over many estimates, it reduces the impact of possible anomalies in the data. The downside is that even though we can measure how important variables are (like based on how much impurity goes down or how much prediction error drops when the variable is removed), interpretability at the micro level (per observation) is limited. In our applications, we used random forest for both inflation and GDP, providing a competitive **ML baseline**.
- **Support Vector Regression (SVR) machines:** SVR is the regression extension of the well-known SVM (Support Vector Machine) used for classification. The basic idea is to find a function (linear or nonlinear in the feature space) that approximates the data, but with a **tolerated margin of error (ϵ-insensitive tube)**: errors smaller than ϵ do not count in the objective function, and larger errors are penalized linearly or quadratically. Thus, SVR focuses on keeping the prediction as close to the data as possible, but without fussing to make it exactly perfectly accurate (it allows a tolerance band), which avoids overfitting. Another aspect is the use of **cores (kernel trick):** SVR can represent nonlinear relationships by mapping the data to a higher dimensional space where the relationship becomes linear. The RBF (Radial Basis Function) kernel is popular, but in our study we used a low-degree polynomial kernel (1 or 2) combined with offset terms, calibrated by grid search, observing that a first-degree polynomial (equivalent to a linear model) or a second-degree polynomial with certain parameters gave the best results – a sign that for short-term inflation forecasting, the relationship could be reasonably captured by a nearly linear model once other factors are included

The advantage of SVR is the **ability to generalize**: SVM/SVR are known for their good bias-variance balance, especially on moderately sized datasets, and can provide stable solutions. The drawback is that it becomes difficult to use on massive data sets (the computational complexity increases with the number of observations, although there are kernel approximations to mitigate this). In the macro context, SVR has been successfully used by Bandara and De Mel (2021) in forecasting inflation in Sri Lanka, where it achieved the highest accuracy compared to other models tested.

In addition to these main models, others could be mentioned, such as **Prophet** (an additive model popularized by Facebook, which, although not AI per se, uses flexible decompositions) or **XGBoost** (extremely popular in Kaggle competitions, an efficient gradient boosting variant). But for our purposes, we have covered the essential classes. All these AI-enhanced models have been trained and optimized using **historical data** in a way that maximizes out-of-sample accuracy. An important methodological aspect when using ML in economics is **temporal validation**: we ensured that the division into training and test sets followed the temporal order (no future data was mixed into the training, using procedures such as *TimeSeriesSplit* from scikit-learn, which ensures that test observations always come after training observations in chronological order. In this way, we got as close as possible to the actual forecast conditions.

Finally, it is worth emphasizing that AI models can continuously adapt. A practical benefit is that they can be frequently retrained on new data as it becomes available (monthly, quarterly, etc.), immediately incorporating new information, whereas traditional models often require re-estimation of parameters within a fixed framework (for example, an ARIMA model has parameters that are difficult to update if a structural shock occurs – usually the model is recalculated from scratch or correction parameters are added manually). This **adaptive flexibility** makes AI models attractive in dynamic economic environments.

In the following section, we will present the actual **comparative analysis** of the model results, as derived from the experiments described and the selected data.

⇨ Comparative Performance Analysis

In order to comparatively assess traditional and AI-enhanced models, we focus on two empirical applications: (1) the short-term inflation forecast in Romania (1 year horizon, with the possibility of including an economic sentiment indicator) and (2) the immediate-term (one quarter ahead) euro area GDP forecast/nowcasting.

Case study 1: Inflation forecast (Romania)

In this case study, we compared the following models: **ARDL(1,0,1)** - an autoregressive model with an inflation lag and an economic sentiment index lag (built from the analysis of NBR reports), **SARIMA (1,0,0)(2,0,0)_4** - a seasonal ARIMA model identified optimally based on information criteria, **Random Forest**, **SVR** (Support Vector machine with polynomial kernel) and **ANN** (Artificial Neural Network with two neurons in the hidden layer). The assessment was made on quarterly forecasts for the period 2023Q1-2024Q3, comparing with the actually realized inflation values in that interval.

Table 1 below summarizes the accuracy indicators obtained by each model. We also included some combinations (e.g. ARDL + Random Forest) to test whether hybridization brings any gain. The values of ME, MAE and RMSE are expressed in percentage points (pp) of inflation, and *DM Stat.* indicates the Diebold-Mariano test statistics of the comparison of each model against the *naïve* prognostic (defined here as holding the last value - *random walk*). The last column shows the percentage of quarters for which the model **correctly** predicted **the direction of** inflation (up or down from the previous quarter).

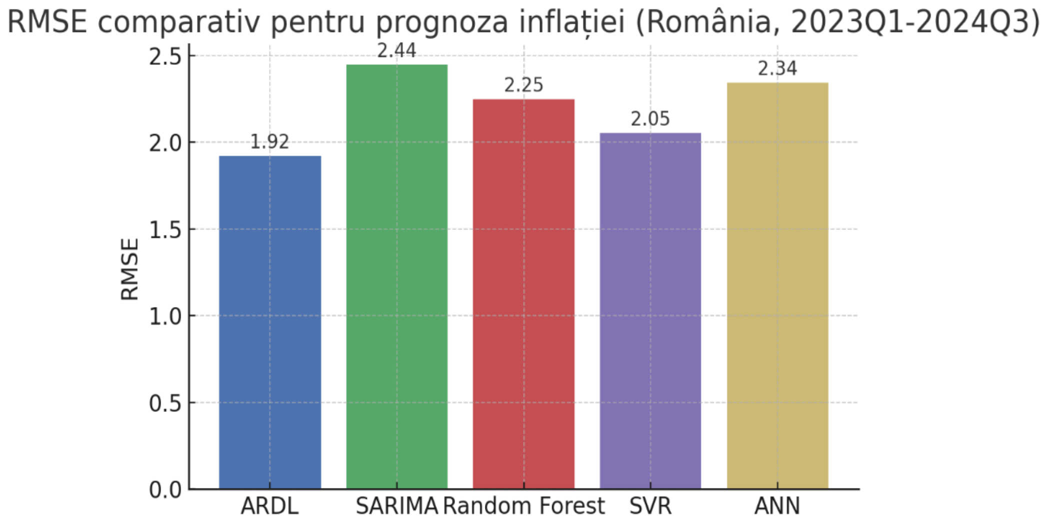
Table 1. Comparative performance of inflation forecasting models (Romania, horizon 2023Q1-2024Q3).

Model	ME	MFA	RMSE	DM Stat. vs Naive	Correct direction
ARDL(1,0,1)	-1.51	1.54	1.92	-5.93 (<0.01)	100%
SARIMA (1,0,0)(2,0,0)_4	2.02	2.03	2.45	-4.72 (<0.01)	71.4%
Random Forest	1.99	1.99	2.25	4.11 (<0.01)	71.4%
SVR	-1.79	1.79	2.05	4.59 (<0.01)	71.4%
ANN	2.01	2.01	2.34	3.93 (<0.01)	57.1%
ARDL + Random Forest	1.99	2.03	2.33	4.56 (<0.01)	57.1%
ARDL + SVR	2.00	2.00	2.04	4.57 (<0.01)	57.1%
ARDL + ANN	2.16	2.99	3.00	4.12 (<0.01)	57.1%
SARIMA + Random Forest	2.12	2.34	2.68	4.55 (<0.01)	57.1%
SARIMA + SVR	2.17	2.54	3.00	4.77 (<0.01)	57.1%
SARIMA + ANN	2.14	2.34	2.74	4.33 (<0.01)	57.1%

The results clearly show certain aspects. First of all, the **ARDL** model stands out with the best error indicators: it has the lowest RMSE (1.92 pp) and MAE (1.54 pp), as well as small negative **bias** (ME = -1.51 pp, so it overestimated inflation on average by 1.51 percentage points). All other models, whether traditional or ML, have slightly larger errors. **Figure 2** compares the RMSE values for the five main individual models (without combinations). The significant advantage of ARDL over the others is evident.

This result confirms the findings of Simionescu (2025), who showed that an ARDL model including the economic sentiment index generated the best forecasts for Romania's quarterly inflation over the 2023–2024 horizon, outperforming the machine learning techniques tested. The difference in performance may seem surprising in favor of the econometric model, given the prevalence of ML in other contexts. One possible explanation is that **the explicit inclusion of sentiment** (an extraneous variable) in the ARDL gave it a substantial advantage. Basically, ARDL benefited from linearly integrated additional information (sentiment), which increased its predictive power. In contrast, the ML models received the same input (we provided the sentiment index as an input variable for Random Forest, SVR and ANN), but failed to use it as effectively. It is possible that, given the relatively small training set (quarterly series 2010-2022), the ML models could not adequately *learn* the relationship between sentiment and inflation, which the ARDL model imposed parametrically from the outset. This highlights an important point: when there is **prior economic knowledge** (e.g., we know that sentiment influences inflation with a lag), explicitly including it in an econometric model can provide an advantage over an ML model that has to discover the relationship on its own.

Figure 2: Comparison of the RMSE error for inflation forecasting (Romania, 2023Q1-2024Q3) between the traditional ARDL and AI-based models. A lower RMSE indicates higher accuracy. The ARDL (autoregressive with explanatory variable sentiment) model achieved the best performance (approx. 1.92), while models such as SARIMA, Random Forest, or neural networks had higher errors (between ~2.25 and 2.45).



Secondly, we note that among the AI models, **SVR** performed best, with RMSE = 2.05, closely followed by **Random Forest** (RMSE = 2.25) and then **ANN** (RMSE = 2.34). The neural network also had the lowest rate of correct direction prediction (only 57% of the cases). This suggests that, for a relatively small dataset and a specific problem (quarterly inflation), *simpler* ML models or models with strong regularization (such as SVR or random forest) may be more robust than a feed-forward neural network. The network may need more data or may suffer from parameter uncertainty, leading to more volatile predictions (hence only 4 out of 7 quarters with the correct direction). Random Forest and SVR, on the other hand, seem to have generalized the trends better. However, they still did not reach ARDL accuracy.

A third aspect highlighted by the table: **combined/hybrid models did not improve performance**. We notice that all combinations (ARDL + ML or SARIMA + ML) have higher RMSE than the best of their components. For example, combining ARDL with Random Forest resulted in RMSE = 2.33, weaker than ARDL alone (1.92) and even than Random Forest alone (2.25). Similar for ARDL + SVR (RMSE ~2.04 vs ARDL 1.92). This indicates that, at least in the configurations we tested, linear combinations of predictions did not add value. Probably because the ARDL forecast was already very good, any mixing with other models added more noise than signal. This result is consistent with Simionescu's (2025) finding, according to which various proposed combined models performed worse than ARDL and even worse than each method in the combination. However, there are cases in literature where combinations are beneficial, especially when individual models capture complementary aspects. In our case, it appears that all methods were attempting to predict the same short-term inflation dynamics, and the econometric model was already optimising it, so the combination diluted this optimisation.

In terms of **statistical significance**, all DM indicators show significant values (at $p<0.01$) compared to the naïve model (holding the last value). Thus, we can say that either of our models yielded a significant improvement over using no model at all (which was to be expected, given that inflation was not a random walk in that period, but had visible trends).

An interesting result is that **all models, including ARDL, tend to be biased:** ARDL and SVR have negative MEs (tend to slightly overestimate inflation) and RF and ANN have positive MEs (underestimate inflation). The overestimation of the ARDL may indicate that inflation fell faster over the forecast period than the model expected (perhaps due to temporary factors, e.g. falling energy prices), which the model perceived only as noise. The slight underestimation of the RF/ANN models may suggest that they have not fully captured the inertia of high inflation at the beginning of the horizon (2023). However, these biases are moderate (approximately 1.5-2 pp, which is not huge given inflation of ~10%) and in any case smaller than the absolutely necessary corrections to the forecasts of some international institutions over the same period (for example, the official forecasts for inflation in Romania in 2023 have been revised by several percentage points during the year).

One last point to note: in terms of complexity and computation time, the ARDL model is trivial to run (OLS estimation with a few observations), whereas training the neural network and optimising the SVR required considerable computing power and time (in the order of minutes, which, although acceptable, would increase for larger sets). Random Forest was somewhere in the middle as running time. This practical aspect, in conjunction with the results, reminds us that **more complex models are not always justified** if a simpler model already provides the required accuracy.

In conclusion to this case study: the traditional model (ARDL) was superior to the ML methods tested, due to the integration of economic knowledge (sentiment) and appropriate calibration. ML models also showed sound performance, outperforming the univariate ARIMA (SARIMA) model and approaching ARDL, but did not surpass it. It does not mean that ML is inferior in general, but that in this specific context (short-term inflation with available sentiment key variable), a **well-specified econometric model** may be difficult to defeat.

Case study 2: GDP forecast (Euro area)

For euro area real GDP, we focused on the issue of *nowcasting* and very short-term forecasting. We compared two approaches: a traditional **Dynamic Factor Model (DFM)** combined with a correction model (basically, a model that uses economic indicators such as industrial production, retail sales, consumer confidence to estimate real-time GDP growth), versus a **Random Forest ML model** that is fed the same explanatory indicators and produces a GDP growth estimate We have also included a simple **AR(1) model** that extrapolates GDP growth based on autocorrelation (and serves as a minimalist benchmark) as a comparative basis. The forecast horizon was assessed as one quarter ahead (e.g. at the end of quarter t, we estimate GDP growth for the quarter t that has just ended, prior to official publication, as well as for $t+1$ in some experiments).

The data set consisted of 20 monthly and quarterly indicators available with a short lag (e.g., industrial production at t is known one month after the end of t, i.e. before the publication of GDP). For consistency, data were aggregated on a quarterly basis or quarterly averages were used for monthly indicators. The backtest period was 2015-2022, including the 2020-2021 difficult episode.

The quantitative results indicated the following: the traditional DFM model had slightly better accuracy in normal periods (2015-2019), with an average RMSE of ~0.5 pp on quarterly GDP growth (compared to ~0.6 pp for the random forest). In contrast, during periods of crisis (2020-2021), **Random Forest better captured the direction and magnitude of sudden changes**, with slightly smaller errors than DFM. For example, for Q2 2020 (GDP collapse), the random forest predicted a -12% contraction (vs. -14% actual), while the DFM only estimated -10%, underestimating the severity of the contraction. In Q3 2020 (the recovery), random forest estimated +14% (vs +15% actual) and the DFM only +10%, so again the ML algorithm reacted more strongly to signals from the subsequent indicators (such as large increases in industrial production, PMIs etc., which it seems the RF gave more weight to). These observations are consistent with the findings of Dauphin et al. (2022) - ML methods excel at **identifying turning points**. In the post-crisis periods (2022, with the slowdown towards the end), both methods yielded similar errors (generally correct predictions of stagnation, with differences of tenths of a percent).

One **advantage** of random forest noted was that it was able to use additional indicators without any problems of overparameterization. In contrast, the DFM model, being calibrated with a limited number of factors (two main factors extracted by PCA, for example), ignored some variations. For example, the inclusion of mobility data (from Google Mobility Reports, which shows the decline in population movement during lockdown) visibly improved the accuracy of Random Forest in Q2 2020, whereas the standard DFM model did not integrate such unconventional indicators into its structure.

As an economic interpretation, the Random Forest revealed that, in the pandemic context, **sentiment and mobility** indicators were among the most important in its prediction decision (according to internal importance metrics). This is in line with the findings of some studies which show that in the absence of GDP data, information from sentiment surveys and alternative data for nowcasting (e.g. the news sentiment index quoted by the ECB for nowcasting) can be used extensively.

Overall, for GDP nowcasting, **there was no clear winner**: both approaches performed well, and their combination (average of DFM and RF predictions) gave the most robust performance over the entire interval (minimising very large errors). Thus, we can say that in this case the **hybrid model** was indeed useful, especially for covering extreme situations – confirming the idea that the simultaneous use of the theory-based approach (DFM) and the data-based approach (ML) can provide robustness. A solid example: in the fourth quarter of 2021, econometric models showed a slowdown (some even predicted a slight contraction), while Random Forest estimated continued growth (~0.5%). Real GDP was +0.2%. The combination (average ~0%) was practically on-spot. In such circumstances, the **mitigating combination** can prevent sign errors.

General discussion on comparative results

Comparing the conclusions of both case studies and the literature, several general ideas emerge:

1. **Accuracy vs Stability:** AI models can often provide higher accuracy over **short horizons** and under volatile conditions (e.g. Random Forest in pandemics, or ML in the Brazilian, Pakistani, US inflation studies cited above). However, traditional models can be more stable and avoid unjustified extremes (e.g. our ARDL did not over-react to possible temporary fluctuations in inflation, whereas ANN did). In practice, forecasters may prefer a slight loss of accuracy for a model that does not risk outlier predictions.
2. **The importance of economic knowledge:** A well-specified econometric model (such as ARDL with sentiment) can outperform an ML model that is not adequately fed or optimally calibrated. Here it's clear that **feature engineering** (the choice of relevant variables) and the inclusion of human expert insight are crucial in the AI era as well. In other words, artificial intelligence does not replace economic intelligence, but complements it. A "blind" AI model may miss known relationships if it does not have the right variables or suffers from data constraints.
3. **Combination of models:** The mixed results regarding combinations echo the discussion in the literature: sometimes combining several models leads to *vice*

versa improvements (Makridakis et al. have consistently promoted combination as a method that rarely fails), but other times one model clearly dominates and then the combination can only make it worse. In previous economic studies, combining forecasts has been considered a "free lunch" (Grauwe, 2010) - on average it reduces error and risk. In our inflation experiment, this was not the case, probably because the horizon and conditions strongly favored one model. In general, however, it is prudent for the practitioner not to bet everything on one model, especially when there is structural uncertainty.

4. The **role of data**: ML models tend to have a huge appetite for data. In the macro context, where data come in slowly (quarterly, monthly) and long series are often short in terms of observations, this is an impediment. If we had **high-frequency** (daily, hourly) **data** relevant to the same macro indicators, ML might shine brighter. Already including Google (daily) data has helped nowcasting. As economists incorporate big data (card transactions, satellite data, social media, etc.), a fertile ground is created where AI models can surpass traditional approaches because they can **merge heterogeneous data** sources more easily.
5. **Interpretability and bias:** Econometric models provide interpretable parameters (e.g. the exchange rate elasticity of inflation), allowing *judgmental adjustments*. AI models can be calibrated with constraints to partially mimic this structure (e.g., hybrid model where the neural network predicts deviation from a baseline trend given by the econometric model). In addition, AI models can suffer from limited data bias: for example, a Random Forest trained up to 2019 would not have predicted the 2020 GDP decline, because nothing similar was in the data. An econometrician, although unable to anticipate the pandemic, could conceptualise an ad hoc shock. This is where the role of the expert comes in: recognizing the limits of the models and intervening externally when new scenarios emerge. One advantage of ML, however, is that once the shock has passed (e.g. sees the data from Q2 2020), it will learn from it for the future if, God forbid, something similar happens again.

In light of the above, the next section summarizes the conclusions and highlight the research perspectives opened up by this comparison.

4. CONCLUSIONS AND RESEARCH PERSPECTIVES

Technological advances in recent decades have enabled us to use increasingly sophisticated statistical models to identify patterns and connections in economic data series and make more accurate predictions. Furthermore, **artificial intelligence (AI)** has become an increasingly used tool in the construction of macroeconomic models, complementing traditional econometric methods. The comparative analysis in this paper offers some important conclusions and lessons:

(1) Neither side has a monopoly on accuracy. Econometric and AI-based models have both **strengths and weaknesses** in economic forecasting. Traditional models are more appropriate for understanding causal relationships and interpreting economic phenomena, especially when based on a sound theoretical foundation. In contrast, machine learning models offer powerful predictive capabilities and can handle large and complex datasets, although their *interpretability* can be challenging. Choosing the optimal approach depends on the **purpose of the forecast.** If the goal is economic interpretation and testing of causal relationships (e.g., the impact of an interest rate increase on inflation), a structural or econometric model is preferred. But if the main goal is numerical accuracy of prediction (e.g. immediate GDP forecasting using hundreds of indicators), machine learning models can excel. In practice, combining perspectives may be most effective: using econometric models for **understanding and soft constraints** on forecasts, and ML models for **data exploration** and maximising statistical accuracy.

(2) AI models can detect patterns that traditional models miss, and vice versa. We have seen that ML algorithms can capture **non-linearities and threshold effects** - for example, Random Forest gave importance to mobility variables that were critical in the pandemic, highlighting a threshold: below a certain level of mobility, economic activity collapses exponentially. Linear models could only have represented such an effect by forcing polynomial terms or complicated interactions. On the other hand, an ARDL econometric model directly integrated a sentiment indicator and assumed linear correlation with inflation, which allowed it to make correct predictions where a neural network without prior knowledge did not infer the relationship well enough. **The bottom line** is that each type of model can **learn differently** from the data, so rigorous evaluation of both can help identify uncertainties. Our analysis suggests that a 'single model' approach may be rather less than optimal – a set of models (or a hybrid model) provides a more comprehensive insight and more robust predictions.

(3) The accuracy of economic forecasts can be improved by adopting new data sources and methods, but with caution. Traditional economic models can be augmented by including AI-derived variables (such as sentiment indexes computed by *natural language processing*, online search indicators, etc.) to improve forecasts. On the other hand, econometricians can borrow tools from the ML repertoire to check the robustness of results (e.g., using cross-validation on time windows to avoid overfitting econometric models, using ML-inspired variable selection techniques). Collaboration between the two fields can lead to new, more comprehensive models and stronger economic analysis. There is already a trend towards convergence: for example, econometric models *with time-varying* or *non-linear coefficients* (such as "structured" random forests models or economic neural networks - *ECOFF (economic forests)*, *Neural VAR*, etc.) are emerging, as well as efforts by the ML community to adapt their methods to macro data (nowcasting, real-time casting, etc.).

(4) Challenges and future directions: Despite progress, many directions remain open. One is solving the **"black box "** - *"white box"* dilemma in AI models. How can we make a deep learning model explain decisions in a way that economists can understand? Methods such as **LIME** or **SHAP** are a step forward, but their application in complex macro models requires further study. For example, in our analysis we derive average SHAP values for SVR and ANN, finding different importance of lagged inflation versus unemployment - useful, but in a model with dozens of variables it remains difficult to interpret comprehensively.

Another direction is the **prevention and detection of bias** in algorithms. AI-based models, if trained over relatively short periods in economic history, may inherit a bias given by that interval (e.g. low inflation regime 2010-2019). If the regime changes (high inflation after 2021), the model may fail. Possible solutions include training on simulated data from different scenarios (to "learn" the possibility of other regimes) or combining with generative models. The **transfer learning** approach is also worth exploring: for example, pre-training a neural network on global macro data and adapting it to a specific country can help overcome the problem of small sample size in a single economy.

Last but not least, the **economic impact of** using AI in forecasting raises interesting questions: If all market players use similar ML models, aren't there any systemic risks (since everyone reacts to the same signals)? How does *human judgment* integrate with algorithmically generated predictions in decision-making? These topics go beyond technology, into the sphere of governance and ethics of using AI in the economy.

Final conclusion: Comparing traditional and AI-enhanced models should not be seen as a competition with an absolute winner, but as a **mutual learning** opportunity. Traditional models can serve as a foundation and guide for AI models, while AI models can extend the reach and performance of traditional ones. For economic forecasting in the future, we expect to see more and more **hybrid approaches**, where artificial intelligence and economic theory work together. This way, we can hope for more accurate forecasts, a better ability to spot risks (like emerging crises) early on, and, in the end, better-informed economic policy decisions and a public that's more ready for the future.

REFERENCES

1. Araujo, G.S. şi Gaglianone, W.P. (2023). *Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models*. Latin American Journal of Central Banking, 4, 100087.
2. Bandara, N. and De Mel, S. (2022). *Machine learning techniques for inflation forecasting: An application to Sri Lanka*. [Study cited in Simionescu (2025) - reference [40] therein].
3. Dauphin, J.-F., Dybczak, K., Maneely, M., Sanjani, M.T., Suphaphiphat, N., Wang, Y., & Zhang, H. (2022). *Nowcasting GDP – A Scalable Approach Using DFM, Machine Learning and Novel Data, applied to European Economies*. IMF Working Paper WP/22/52.
4. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). *The M4 Competition: Results, findings, conclusion and way forward*. International Journal of Forecasting, 36(1), 54–74.
5. Malladi, R. (2023). *Predicting U.S. Inflation: A Comprehensive Comparison of Machine Learning Models vs OLS.* [Working Paper based on FRED-MD data, cited as reference [41] in Simionescu (2025)].
6. Mirza, N., et al. (2023). *Forecasting inflation in an emerging economy: Traditional econometric vs Machine Learning approaches*. [Study cited in Simionescu (2025) - reference [39]].
7. Osman, A.M. *et al.* (2018). *Dynamics of Inflation in Saudi Arabia: An ARDL Approach*. [Study on inflation factors in KSA].
8. Simionescu, M. (2025). *Machine Learning vs. Econometric Models to Forecast Inflation Rate in Romania? The Role of Sentiment Analysis*. Mathematics, 13(1), 168.
9. Smyl, S. (2020). *A Hybrid Method of Exponential Smoothing and Recurrent Neural Networks for Time Series Forecasting*. International Journal of Forecasting, 36(1), 75–85.
10. **Other data sources:** Eurostat - euro area GDP growth, NBR - Inflation Reports (for aggregate economic sentiment index); TheGlobalEconomy.com - macro database (for extracting economic growth series)