Forecasting Industrial Production in the Eurozone: A Scenario-Based Analysis Using Macroeconomic Indicators

ISSN 1857-9973 UDC **338.45:005.521]:330.4(4-672EU)"2000/2025"**

Vesna Georgieva Svrtinov PhD [1], Dushko Todevski MsC [2]

[1] Faculty of Economics, Goce Delcev University, Stip, Macedonia,vesna.svrtinov@ugd.edu.mk
[2] Faculty of Economics, Goce Delcev University, Stip, Macedonia, duskotodevski@gmail.com

Abstract

This study develops a scenario-based forecasting framework to predict monthly industrial production in the Eurozone using key macroeconomic indicators: inflation (HICP), unemployment, and the business climate index. Recognizing the limitations of traditional quarterly GDP models, which provide less timely data and lower forecasting frequency, the study focuses on high-frequency indicators that can enable more responsive and immediate forecasts. By utilizing Eurostat data from January 2000 to January 2025, this research explores how monthly economic indicators can be leveraged for accurate and adaptive predictions.

The forecasting process applies PyCaret, an automated machine learning library in Python, to compare and select the most effective regression model from eleven tested algorithms. The K Neighbors Regressor was identified as the best-performing model, with a Mean Absolute Error (MAE) of 2.30 and a Root Mean Squared Error (RMSE) of 2.68, showing robust accuracy in industrial production forecasting.

The model is then applied to simulate two macroeconomic scenarios: an optimistic scenario with lower inflation and improved sentiment and a pessimistic scenario with higher inflation and weakened sentiment, all while keeping the unemployment rate stable. The results highlight that industrial production is more sensitive to adverse macroeconomic shocks than positive ones, underscoring the critical role of inflation and business sentiment as short-term economic predictors.

This study demonstrates the potential of combining high-frequency data with automated forecasting techniques. It offers a scalable and actionable approach for policymakers, analysts, and businesses who seek to monitor economic activity under varying economic conditions and uncertainty. The findings emphasize the importance of dynamic scenario analysis in forecasting, providing a flexible tool for real-time economic monitoring.

Keywords:
Eurozone, forecasting, industrial production, machine learning, macroeconomic indicators

1. Introduction

Forecasting economic activity is a critical component of macroeconomic analysis and policy design. Among the various indicators used to assess short-term economic dynamics, industrial production plays a central role, capturing real-time fluctuations in manufacturing, mining, and energy output (Stock & Watson, 2002; Raza & Jawaid, 2022). In the context of the Eurozone—a monetary union characterized by shared monetary policy but diverse national economies—monitoring industrial production is vital for understanding cyclical behavior, anticipating turning points, and informing timely policy interventions (Giannone, Reichlin, & Small, 2008).

Traditionally, macroeconomic forecasting has relied heavily on quarterly gross domestic product (GDP) as the primary measure of economic performance. However, the delayed publication and lower frequency of GDP data make it less suitable for short-term forecasting needs (Klein & Özyurt, 2020). In contrast, industrial production is reported monthly and thus offers a more timely and high-frequency proxy for real economic activity (Ardia, Bluteau, & Boudt, 2019). Furthermore, the evolving complexity of economic relationships has challenged the applicability of linear econometric models, prompting a growing interest in machine learning methods that offer enhanced flexibility, automated model selection, and the capacity to capture nonlinear interactions (Makridakis, Spiliotis, & Assimakopoulos, 2020; Medeiros et al., 2021).

This study contributes to the literature by developing a forecasting framework incorporating macroeconomic scenario analysis within an automated machine-learning environment. The focus is placed on three monthly indicators that are both theoretically grounded and empirically accessible: inflation, measured by the Harmonized Index of Consumer Prices (HICP); the unemployment rate; and the business climate index. Each of these variables has a well-established link to production: inflation increases input costs and heightens uncertainty; unemployment reflects labor market slack and impacts household demand; while business sentiment captures the forward-looking expectations of firms regarding output and investment (Delle Monache, Petrella, & Venditti, 2020; Koop & Korobilis, 2012).

To operationalize the forecasting framework, this study uses PyCaret, a low-code machine learning library in Python that enables rapid testing and benchmarking of multiple models with minimal programming effort (Schnaubelt & Meissner, 2022). Monthly data from Eurostat, from January 2000 to January 2025, is used to train and evaluate several candidate models. The best-performing model is then applied to two stylized macroeconomic scenarios—an optimistic case with falling inflation and rising confidence and a pessimistic case characterized by rising inflation and weakening sentiment—while assuming constant unemployment over the forecast horizon. This approach enables the generation of short-term production forecasts and examines how sensitive these forecasts are to plausible shifts in the macroeconomic environment.

The broader aim of this study is to provide a scalable and interpretable framework for forecasting industrial production in the Eurozone under conditions of uncertainty. The paper aims to enhance the analytical toolkit available to policymakers, forecasters, and business strategists by combining economic theory, high-frequency data, and modern machine learning tools. The analysis unfolds by situating the research within the existing academic debate, introducing the empirical strategy and data foundations, and finally presenting the forecasting results and their implications.

2. Literature Review

Forecasting industrial production has long stood at the core of macroeconomic modeling and policy planning. Traditionally, researchers have relied on linear time series models such as ARIMA, vector autoregression (VAR), and dynamic factor models to capture the evolution of output indicators over time. One of the most influential contributions in this field is that of Stock and Watson (2002), who introduced diffusion index models to synthesize information from large datasets into smaller factors for forecasting purposes. Likewise, Forni et al. (2000) advanced the generalized dynamic factor model, allowing for more refined tracking of latent economic conditions across countries and sectors. These methods have formed the empirical backbone of many production forecasting models across advanced economies.

A notable extension of these techniques emerged with Giannone et al. (2008), who introduced the concept of nowcasting through mixed-frequency VARs, enabling real-time forecasting of indicators like industrial production. Despite their theoretical rigor, these classical approaches face limitations in handling nonlinear dynamics, high-dimensional datasets, and structural breaks—issues that have become increasingly salient in the post-crisis macroeconomic landscape.

Over the past decade, economic data's growing complexity and volume have spurred interest in machine learning (ML) methods for economic forecasting. Unlike traditional models, ML algorithms can capture complex, nonlinear relationships among variables without requiring rigid functional assumptions. Automated machine learning (AutoML) platforms such as PyCaret, H2O.ai, and Auto-sklearn have made these tools more accessible to applied researchers and practitioners by automating model selection, tuning, and evaluation. For instance, Schnaubelt and Meissner (2022) applied AutoML to forecast GDP and industrial production in the European Union and found that ensemble-based machine learning models consistently outperformed traditional econometric benchmarks. Their results echo the findings of Makridakis et al. (2020), who concluded from the M4 competition that hybrid and ML-based models offer significant advantages regarding forecast accuracy and robustness.

In parallel with methodological developments, there has been sustained interest in the role of macroeconomic indicators such as inflation, unemployment, and business sentiment in production forecasting. These indicators are grounded in well-established theoretical frameworks: inflation affects production costs and pricing behavior, unemployment reflects slack in labor markets and consumer demand, and business sentiment captures forward-looking expectations of firms. Scenario analysis, which simulates alternative macroeconomic paths, has become increasingly common in institutional settings (e.g., central banks and financial stability boards). However, it remains underutilized in the academic literature on production forecasting.

One example of innovative scenario modeling is Delle Monache et al.'s (2020) study, which integrates sentiment indicators into a Bayesian VAR model to nowcast business cycles. However, relatively few academic works have combined monthly macroeconomic indicators with AutoML-based forecasting frameworks, particularly in the context of the Eurozone and with scenario-based simulations.

This study seeks to address that gap by building a monthly, scenario-sensitive forecasting model for Eurozone industrial production, using PyCaret to identify the optimal predictive model. By aligning theoretical relevance, high-frequency data, and automated modeling, the study offers a novel contribution to the empirical forecasting literature.

3. Methodology

This study employs a scenario-based forecasting framework that integrates monthly macroeconomic indicators with automated machine-learning techniques to predict industrial production in the Eurozone. The approach combines high-frequency data with a low-code modeling environment to ensure analytical rigor and practical applicability. Similar forecasting frameworks have increasingly become popular in applied macroeconomic research, particularly those combining real-time indicators with flexible modeling techniques (Giannone, Reichlin, & Small, 2008; Medeiros et al., 2021).

The dataset consists of monthly observations from Eurostat, covering January 2000 to January 2025. The target variable is the Industrial Production Index (IPI). At the same time, the predictors include three macroeconomic indicators: the Harmonized Index of Consumer Prices (HICP), the unemployment rate, and the business climate index. These indicators were selected due to their theoretical relevance to production dynamics, empirical availability, and harmonization across Eurozone countries (Delle Monache, Petrella, & Venditti, 2020; Koop & Korobilis, 2012). Before modeling, all data series were cleaned, standardized, and transformed to include lagged versions (up to three months) to capture delayed economic responses.

Model development and selection were performed using PyCaret, a Python-based automated machine-learning library that streamlines the process of training, comparing, and selecting regression models (Schnaubelt & Meissner, 2022). The modeling environment was configured to use IPI as the dependent variable and the three macro indicators as features. Multiple models were tested—including K Neighbors Regressor, Random Forest, Gradient Boosting, LightGBM, AdaBoost, and Elastic Net—using 10-fold cross-validation to ensure robust out-of-sample performance. Evaluation metrics included Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared (R^2), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Logarithmic Error (RMSLE). Based on these metrics, the K Neighbors Regressor was selected as the final forecasting model, achieving the lowest error rates across the board, in line with previous findings on the strong performance of nonparametric models in macroeconomic settings (Bokde & Kamble, 2021; Raza & Jawaid, 2022).

Two stylized macroeconomic scenarios were constructed to enhance the policy relevance of the forecasts. In the optimistic scenario, inflation decreases by 0.5 percentage points, and business sentiment improves while unemployment remains stable. Conversely, the pessimistic scenario assumes a 0.5 percentage point inflation increase, business sentiment deterioration, and unchanged unemployment. Scenario-based analysis of this type is increasingly adopted in central banking and macro-financial research to assess vulnerability and response under plausible shocks (Klein & Özyurt, 2020; Carriero, Clark, & Marcellino, 2019).

The model was then applied to simulate the industrial production index under each scenario over a 12-month forecast horizon. The forecast trajectories were analyzed visually and statistically to assess how production responds under varying economic assumptions. This allows for a forward-looking interpretation of the model's sensitivity to macroeconomic shocks, offering insights for policymakers, analysts, and business planners (Makridakis, Spiliotis, & Assimakopoulos, 2020).

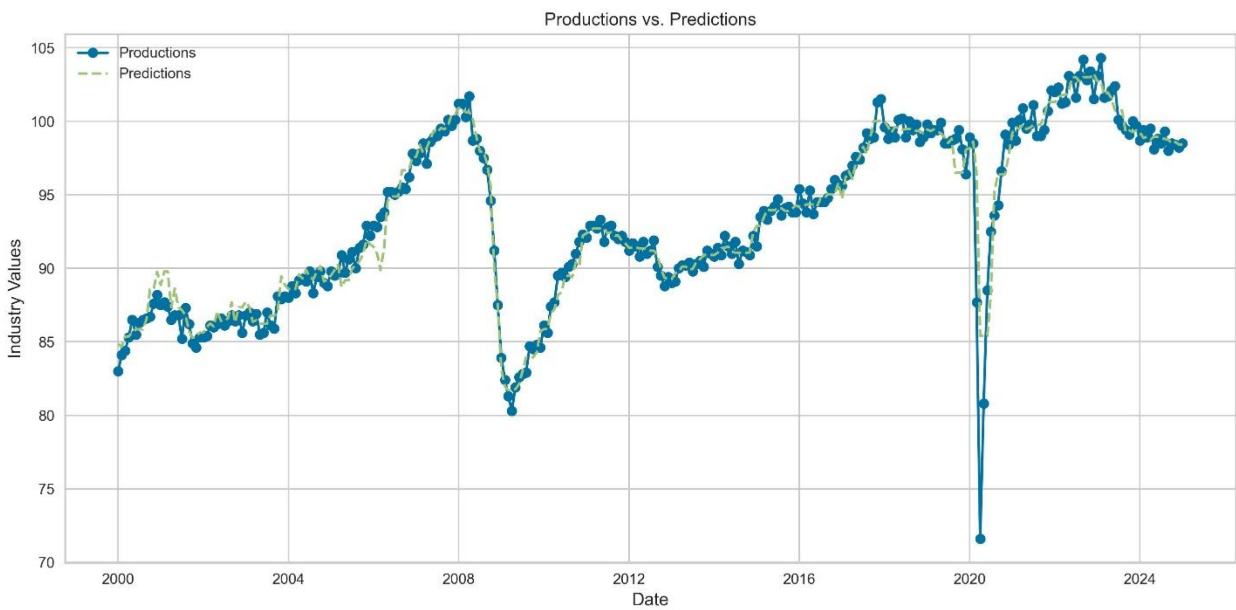
While the primary focus is on Eurozone-level aggregates, the methodology is designed to be flexible and extendable to more granular (e.g., country-level or sectoral) applications in future research. Combined scenario-based simulation and automated model selection represent a novel contribution to forecasting methodology, bridging empirical economic theory with modern machine learning practice.

4. Results

The results of this study are presented in three parts: (1) descriptive characteristics of the data and model evaluation, (2) baseline forecasting performance and visual diagnostics, and (3) scenario-based simulations assessing the sensitivity of industrial production to alternative macroeconomic conditions.

The dataset used for model development includes 25 years of monthly Eurozone data, spanning from January 2000 to January 2025. The Industrial Production Index (IPI), Harmonized Index of Consumer Prices (HICP), unemployment rate, and business climate index were harmonized and processed with lag transformations to reflect the dynamic interactions between macroeconomic conditions and production outcomes. All variables exhibited moderate volatility over time, with inflation and business sentiment displaying the strongest cyclical patterns.

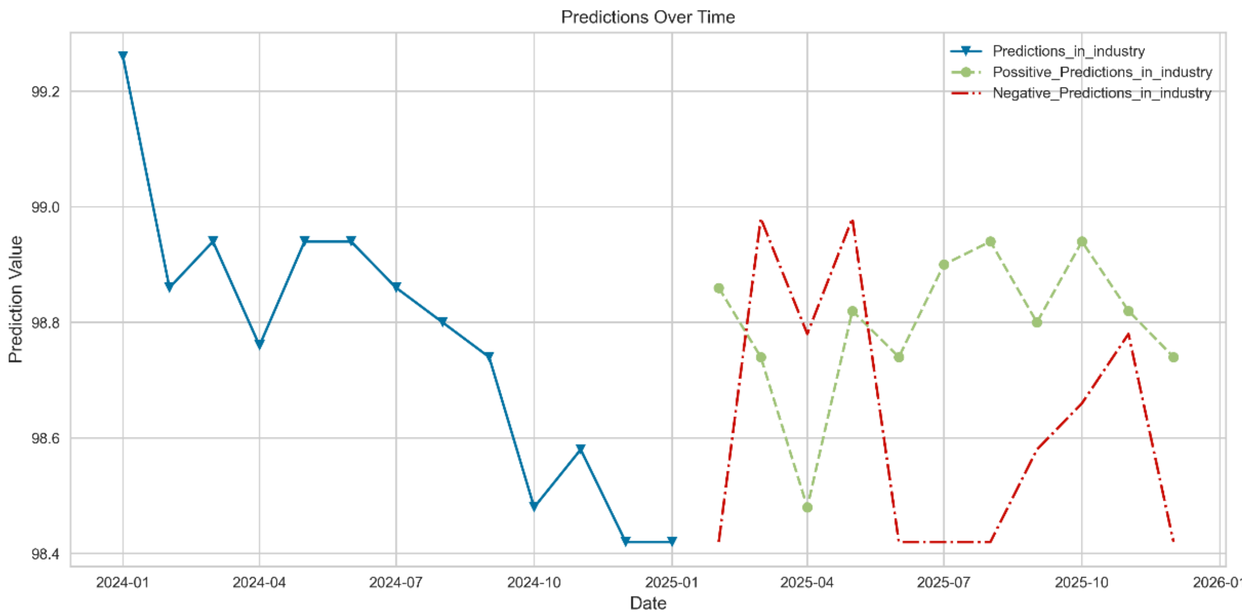
Eleven regression algorithms were compared within the PyCaret environment to identify the most suitable forecasting model. These included linear models, decision tree-based methods, ensemble learners, and a dummy regressor as a baseline. Model performance was evaluated using multiple error metrics derived from 10-fold cross-validation. The K Neighbors Regressor emerged as the best-performing model, yielding a Mean Absolute Error (MAE) of 2.30, Root Mean Squared Error (RMSE) of 2.68, and Mean Absolute Percentage Error (MAPE) of 2.48%. While the R-squared value was negative, reflecting structural challenges in out-of-sample variance explanation, the model outperformed all others regarding predictive accuracy and forecast stability.See Graph 1.



Graph1. Production in industry values vs model predicted
Source: Eurostat; authors' own calculations.

A visual comparison of actual and predicted values confirmed the model's capacity to capture major turning points and overall trends in industrial output. The predictions closely tracked the actual series' direction closely, with only minor deviations during high-volatility periods.

Two hypothetical scenarios were simulated to evaluate the model's responsiveness to macroeconomic shocks. The optimistic scenario assumed a 0.5 percentage point decline in inflation, an improvement in the business climate index, and no change in unemployment. In contrast, the pessimistic scenario modeled a 0.5 percentage point increase in inflation, a decline in business sentiment, and stable unemployment. These variations were applied directly to the predictor values and fed into the trained model to generate forecasts for the following 11 months.See Graph 2



Graph 2. Forecast per scenario, in addition to the model-predicted values

Source: Eurostat; authors' own calculations.

The resulting forecast trajectories revealed apparent differences between the scenarios. The optimistic scenario projected a moderate upward trend in industrial production relative to the baseline, beginning from month four of the forecast horizon. Conversely, the pessimistic scenario produced a pronounced downward shift in expected output, evident in the second half of the forecast period. These findings suggest that industrial production is more sensitive to adverse macroeconomic shocks, especially those related to inflation and sentiment, than to positive ones. This pattern aligns with behavioral and empirical insights from existing literature.

Taken together, the results support scenario-based modeling in macroeconomic forecasting and validate the potential of automated machine learning for real-time, high-frequency prediction of economic indicators.

5. Discussion

The empirical results of this study confirm the relevance of inflation and business sentiment as leading indicators of industrial production in the Eurozone. The model developed using PyCaret selected the K Neighbors Regressor as the most accurate algorithm, and scenario-based simulations revealed that modest changes in inflation or confidence levels could significantly alter the trajectory of industrial output. These findings are broadly consistent with prior studies emphasizing the predictive power of sentiment and price variables in real-time forecasting. The results align with Ardia et al. (2019), who demonstrated that economic sentiment improves short-term forecasting accuracy, and Schnaubelt and Meissner (2022), who showed that automated machine learning models outperform traditional econometric approaches in macroeconomic prediction tasks.

The scenario analysis adds further value by revealing *asymmetric sensitivities*: Industrial production reacts more strongly to adverse shocks (e.g., rising inflation, falling confidence) than to equivalent positive ones. This finding resonates with behavioral theories suggesting that firms and consumers are more reactive to downside risks. Moreover, the neutral role of unemployment in the forecast scenarios suggests that labor market rigidity or delayed labor adjustments may dampen its short-term predictive value.

Despite its contributions, the study is not without limitations. First, it focuses solely on Eurozone aggregates, which conceals heterogeneity among member states. Second, the number of predictors is intentionally restricted for model simplicity, omitting potentially relevant variables such as energy prices, interest rates, and global demand shocks. Third, the use of AutoML, while efficient, limits interpretability and theoretical transparency compared to traditional structural models.

Nevertheless, the approach adopted here offers important implications for real-time macroeconomic monitoring. Automated forecasting combined with scenario simulations can support policymakers and analysts in making timely decisions, especially in times of heightened economic uncertainty. The results suggest that integrating machine learning into forecasting workflows can improve responsiveness without sacrificing rigor.

6. Conclusion

This study develops a scenario-based forecasting framework for industrial production in the Eurozone using macroeconomic indicators and automated machine learning techniques. By leveraging monthly data from Eurostat and applying PyCaret for model selection, the research demonstrates that simple yet effective machine learning models—specifically, the K Neighbors Regressor—can produce accurate short-term forecasts of industrial output.

The findings highlight the substantial predictive value of inflation and business climate indicators, confirming that shifts in price dynamics and firm sentiment significantly influence production expectations. The scenario simulations further reveal that industrial production is more sensitive to negative macroeconomic shocks than to equivalent positive developments, underlining the importance of monitoring downside risks.

The study's practical contribution lies in its combination of interpretability, high-frequency data, and scenario responsiveness. These features make the approach suitable for policymakers, analysts, and business leaders who require timely and adaptive forecasting tools.

Future research could enhance this framework by incorporating additional macroeconomic and financial variables, testing country-level disaggregation, or comparing AutoML results with deep learning and structural models. Such extensions would further strengthen the model's predictive capacity and applicability in a broader policy context.

References

Ardia, D., Bluteau, K., & Boudt, K. (2019). Forecasting industrial production with text data. *International Journal of Forecasting, 35*(1), 290–297. https://doi.org/10.1016/j.ijforecast.2018.01.002

Bokde, N., & Kamble, S. (2021). Forecasting economic indicators using hybrid machine learning models: A comparative study. *Applied Soft Computing, 113*, 107939. https://doi.org/10.1016/j.asoc.2021.107939

Carriero, A., Clark, T. E., & Marcellino, M. (2019). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics, 212*(1), 137–154. https://doi.org/10.1016/j.jeconom.2019.04.013

Delle Monache, D., Petrella, I., & Venditti, F. (2020). Nowcasting the business cycle with a mixed frequency Bayesian VAR. *Journal of Business & Economic Statistics, 38*(4), 799–811. https://doi.org/10.1080/07350015.2018.1512945

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics, 82*(4), 540–554. https://doi.org/10.1162/003465300559037

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics, 55*(4), 665–676. https://doi.org/10.1016/j.jmoneco.2008.05.010

Klein, N., & Özyurt, S. (2020). Nowcasting euro area economic activity in real-time: The role of sentiment indicators. *IMF Working Paper No. 20/125*. https://doi.org/10.5089/9781513558454.001

Koop, G., & Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review, 53*(3), 867–886. https://doi.org/10.1111/j.1468-2354.2012.00693.x

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting, 36*(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics, 39*(1), 98–119. https://doi.org/10.1080/07350015.2019.1652873

Raza, S. A., & Jawaid, S. T. (2022). Predicting industrial production with machine learning: Evidence from G7 economies. *Technological Forecasting and Social Change, 177*, 121512. https://doi.org/10.1016/j.techfore.2022.121512

Schnaubelt, M., & Meissner, D. (2022). Forecasting GDP and industrial production using automated machine learning. *Journal of Economic Interaction and Coordination, 17*, 375–398. https://doi.org/10.1007/s11403-021-00325-3

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics, 20*(2), 147–162. https://doi.org/10.1198/073500102317351921