QUANTITATIVE METHODS IN ECONOMICS
Volume XXVI, No. 2, 2025, pp. 60 – 79

FORECASTING IN TURBULENT TIMES: HOW ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING ARE RESHAPING MACROECONOMIC PREDICTION

Krystian Jaworski https://orcid.org/0000-0001-5186-9732
Collegium of World Economy, Department of Economics II
SGH Warsaw School of Economics
e-mail: kjawor@sgh.waw.pl
Nemanja Popović https://orcid.org/0009-0003-5686-9362
Faculty of Economics Podgorica
University of Montenegro
e-mail: nemanja.p@ucg.ac.me

Abstract: This paper examines how artificial intelligence and machine learning have reshaped macroeconomic forecasting in the volatile post-COVID era. Highlighting the use of ensemble methods, neural networks, and large language models, it illustrates their advantages in capturing nonlinear dynamics and processing complex data. Drawing on central bank case studies, the paper shows that AI enhances predictive power, though interpretability and robustness remain challenges. AI is best seen as a complement to, not a replacement for, traditional economic models and human judgment.

Keywords: macroeconomic forecasting, artificial intelligence; machine learning; large language models; inflation forecasting

JEL classification: C53, E37

INTRODUCTION

Macroeconomic forecasting has long been a cornerstone of economic policy planning. Before 2020, forecasters operated in a relatively stable environment by historical standards – often characterized by moderate fluctuations and well-understood business cycle dynamics. In the decades leading up to the COVID-19 pandemic, especially during the so-called "Great Moderation", the mid-1980s to mid-2000s, the volatility of key macroeconomic variables like output and inflation

https://doi.org/10.22630/MIBE.2025.26.2.6

was subdued. This stability enabled forecasters to rely on historical patterns and established models with some confidence. Standard forecasting tools included structural models grounded in economic theory (e.g. Dynamic Stochastic General Equilibrium models) and time series models (such as vector autoregressions and statistical trend projections), typically assuming that the future would resemble the recent past with only modest shocks. Indeed, large forecast errors were relatively infrequent in the absence of major disturbances, although events like the 2008–09 Global Financial Crisis had already highlighted that traditional models could falter in the face of unprecedented turmoil [Tchoketch-Kebir, Madouri 2024]. By the late 2010s, however, a prolonged period of low inflation and steady growth had perhaps lulled many forecasters into a false sense of security regarding the stability of economic relationships (a point underscored by the widespread failure to foresee the financial crisis or the sluggish recovery that followed). The period since 2020 has confronted forecasters with an unusually turbulent sequence of global shocks, each compounding the uncertainty of the previous one. First and foremost was the COVID-19 pandemic itself, which triggered the deepest global recession in living memory almost overnight (the "Great Lockdown"). The COVID-19 pandemic in 2020 marked a dramatic turning point for macroeconomic forecasting, exposing the fragility of standard forecasting approaches when faced with a truly novel disturbance. Given this context, the purpose of this paper is to examine how the COVID-19 shock and subsequent global upheavals have increased the complexity and uncertainty of macroeconomic forecasting, and how forecasting models and practices have been forced to adapt in response. The pandemic set off a chain reaction of successive global shocks – including supply chain breakdowns, energy price spikes, geopolitical conflicts, and an inflation surge – which together created a volatile and highly uncertain post-2020 environment for forecasters.

THE RISE OF AI IN FORECASTING

One of the most frequently noted innovations in macroeconomic forecasting in response to recent challenges has been the rapid integration of Artificial Intelligence and Machine Learning (AI/ML) techniques. Over the past few years, AI/ML tools have moved from the periphery of economic research toward the mainstream of forecasting practice, as traditional methods proved insufficient in the face of unusual shocks [Jouilil, Iaousse 2023]. The appeal of these methods lies in their ability to automatically detect complex patterns, utilize vast amounts of data (including novel data sources), and model nonlinear relationships that classical models might miss. In an environment where past equations broke down, machine learning offered a data-driven way to "let the data speak," potentially uncovering new predictive signals.

One family of AI techniques making inroads is machine learning ensemble methods, such as random forests and gradient boosting machines (GBMs). These are powerful nonlinear regression algorithms that combine many simple prediction rules to capture complex interactions in the data. A random forest, for example, builds numerous decision trees on random subsets of data and averages them, which tends to improve accuracy and robustness compared to a single tree. Boosting methods like XGBoost and LightGBM iteratively build an ensemble of trees, correcting errors at each step, and often achieve very high predictive accuracy. Central bank researchers have found these methods particularly useful for forecasting in volatile conditions. For instance, the European Central Bank (ECB) has begun using tree-based ensemble models as supplementary forecasting tools alongside its core structural models [Lenza et al. 2023a]. These models can ingest a large number of input variables – commodity prices, financial indicators, dozens of sectoral confidence indices, etc. – and algorithmically select the most predictive combinations. Machine learning can find patterns without an a priori model specification. Moreover, ensemble methods naturally accommodate nonlinear relationships. This was especially important during the pandemic and subsequent shocks, as relationships like "oil prices → inflation" or "financial stress → credit availability" likely have threshold effects and interactions that linear models don't capture. Indeed, a commentary from the Czech National Bank [CNB 2025] notes that the main advantage of methods like random forests and gradient-boosted trees is their ability to capture nonlinear relationships and complex inflation dynamics, which proved "especially important during sudden shocks" such as COVID-19 and the inflation spike that followed. In practical terms, tree-based models were better able to fit the unusual post-pandemic inflation process (where, for example, inflation might respond nonlinearly once supply chain delays exceeded a certain length). Empirical tests have borne this out: researchers Lenza et al. [2023a] at the ECB find that an advanced ensemble method, *quantile random forests*, can not only forecast the path of inflation but also quantify the uncertainty by predicting the entire distribution (quantiles) of future inflation. This allows policymakers to see risks to the forecast (e.g. a certain probability of very high inflation) and plan accordingly, a critical capability in uncertain times.

Another set of AI tools making waves are neural networks, particularly advanced architectures like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, as well as the more recent Transformer models. Neural networks are nonlinear function approximators inspired by the brain's neurons; when given sufficient data, they can, in principle, approximate very complex mappings from inputs to outputs. Deep learning networks with many layers became famous for image and speech recognition in the 2010s, but they have also been applied to time series forecasting. LSTM networks, a type of RNN, are explicitly designed to handle sequence data and remember long-term dependencies. This makes them well-suited to economic time series that may have multi-quarter trends or delayed effects. For example, an LSTM can, in theory, learn that "a shock now might affect output more strongly three quarters ahead" if such patterns exist in the data, without the modeler needing to pre-specify lag lengths. During the pandemic, some institutions experimented with LSTMs to forecast indicators like

GDP and inflation [Li 2024]. A study at Banco de la República (Colombia's central bank) used LSTM models to forecast Colombian inflation one year ahead [Cárdenas-Cárdenas et al. 2023]. They tried two variants: one with only past inflation as input and another incorporating additional variables (like exchange rates, oil prices, etc.). They found that the LSTM with a richer input set significantly outperformed traditional ARIMA models, especially at longer horizons, with the gains becoming most pronounced beyond 6 months ahead. This suggests that the LSTM was able to capture nonlinear interactions between inflation and other indicators (and perhaps global factors) that a linear model could not, thus yielding more accurate medium-term forecasts.

Similarly, other researchers have applied deep learning to GDP nowcasting, sometimes feeding in an array of high-frequency indicators as a multivariate input to an LSTM or a Transformer model. The Transformer architecture, which underlies modern Large Language Models (LLMs), has also been adapted for time series forecasting in experimental studies. Transformers excel at capturing long-range dependencies in sequences using attention mechanisms. Early results indicate that large models with well-designed learning can sometimes provide more accurate predictions than traditional parametric models in diverse scenarios. In other words, an LLM that has learned to predict patterns (even patterns in text) can be surprisingly powerful when that prediction ability is transferred to economic data – though this is still a cutting-edge area of research.

A notable example of LLM use in forecasting comes from experiments in inflation forecasting using GPT-style models. Analysts at the St. Louis Fed [Faria-e-Castro, Leibovici 2024] tested an LLM (Google's PaLM 2, a large language model) on the task of projecting U.S. inflation and compared its performance to professional forecasters. They found that the LLM, after appropriate training/prompting, could estimate inflation trends more accurately than human forecasters in many cases over the 2019–2023 period. The PaLM-based model was *more accurate in most years and across nearly all forecast horizons* than the median professional forecast. This is a striking result, suggesting that LLMs (which incorporate vast textual knowledge, potentially including real-time news and narratives) picked up on the brewing inflation pressures earlier or interpreted the pandemic/economic data in a way that gave them an edge. However, the study also highlights a key issue: the LLM is essentially a black box, and it is "not entirely clear how [the] AI arrives at its predictions". Complementary evidence comes from Lopez-Lira and Tang (2024), who show that LLMs can forecast indirectly through sentiment analysis of news headlines, with LLM-based sentiment scores outperforming traditional methods and accuracy improving with model size. We will revisit this black-box problem in the risk section, but it is noteworthy that AI models have begun to rival expert judgment in forecasting tasks. The Czech National Bank [2025] undertook a similar exercise with two proprietary AI models (nicknamed "OpenAI o1" and "Grok 2") to forecast Czech inflation, comparing them to each other and to human analysts. During the stable pre-pandemic period, both the AI and human forecasts were similar (all expecting inflation to stay near target). But when inflation surged in 2021–2022, neither humans nor AI predicted the magnitude of the spike accurately – confirming that this shock fooled everyone. Interestingly, one AI model (OpenAI o1) did start signaling persistent inflationary pressure earlier than the consensus; its forecasts in 2021 indicated inflation would keep rising, whereas many human analysts thought the rise was temporary. In retrospect, the AI's direction was right (inflation was not as transitory as hoped), though it still underestimated the degree of the increase, as did essentially all forecasters. The second AI model (Grok 2) behaved more like the humans, treating the initial surge as largely temporary. By late 2023, Grok 2 expected inflation to fall quickly, correctly matching the actual decline, whereas OpenAI o1 expected more persistence. The differences between AI models highlight that not all AI is the same – their architectures and training can yield different biases. The CNB [2025] analysis concluded that these AI models can offer useful "alternative perspectives" – for example, an early warning that inflation might stick around – so they could serve as indicators or supplementary forecasts alongside traditional ones. However, their black-box nature means they aren't replacing human forecasters; instead, they act as an additional input, much like a survey of models.

Aside from neural networks and LLMs, regularization and dimensionality-reduction techniques from machine learning have also proven valuable. A common challenge in macro forecasting is having too many potential predictors relative to the number of observations (since quarterly data records are limited). ML offers solutions like LASSO (Least Absolute Shrinkage and Selection Operator) and ridge regression, which automatically select or shrink coefficients on less useful variables, effectively performing variable selection. During the pandemic, when forecasters threw dozens of new indicators into the mix, these techniques helped avoid overfitting. In fact, an IMF [2024] paper titled *"Mending the Crystal Ball: Enhanced Inflation Forecasts with Machine Learning"* found that a simple LASSO regression outperformed more complex models and benchmarks for short-term inflation forecasting in the volatile 2022–23 period, in a study on Japan. The flexibility to incorporate many predictors (oil prices, output gap, global prices, etc.) and automatically shrink irrelevant ones gave LASSO an edge over both a small-scale AR model and even nonlinear ML like random forests in that case. The authors note that ML models' flexibility and focus on pure forecasting (rather than structural explanation) were key advantages in capturing the evolving inflation dynamics. Essentially, ML could adapt more quickly to the fact that, for example, import prices and supply shocks had become the dominant drivers of inflation during 2021–22, whereas traditional models anchored on past low-inflation regimes struggled.

Another striking example comes from the Bank of England: researchers there used a high-dimensional approach with disaggregated price data. Joseph et al. [2024] built models to forecast UK CPI inflation using hundreds of sub-component price series (essentially breaking inflation into all its item categories) along with various ML algorithms. They found that when they exploit this "big data" of item-level prices, they achieve strong improvements in forecast accuracy – up to 70% reduction in error at a 1-year horizon – compared to a standard aggregate model. In particular, shrinkage methods like ridge and LASSO performed best in handling the item-level data, indicating that the combination of a large, granular information set and appropriate regularization is key to good performance. This result underscores how ML enables forecasters to use the richness of micro data (which items are driving inflation? Are some prices spiking while others are stable?) to better predict the macro outcome. Traditional models could not handle hundreds of inputs or would suffer from overfitting, whereas ML made it feasible and even provided ways (like Shapley values and variable importance measures) to interpret which categories mattered most. In effect, this approach blends economic domain knowledge (knowing that disaggregates contain signal) with ML's data-crunching power.

INSTITUTIONAL ADOPTION & REAL-WORLD APPLICATIONS

There have been numerous other case studies of AI in forecasting across various institutions. To highlight a few: the IMF has applied deep learning to nowcast GDP by feeding satellite imagery and other unconventional data into neural networks (expanding coverage of countries where data is sparse). The BIS (Bank for International Settlements) has explored ML for financial cycle prediction and early warning of crises (finding, for example, that tree-based models can improve predictions of banking stress by capturing nonlinear interactions among credit, asset prices, and global factors [Aldasoro 2025]). The Bank of Canada published guidance on when and how to use ML in economics, showing cases like housing price prediction where ML (e.g. gradient boosting) outperforms because of nonlinear effects of income, interest rates, and regional factors [Desai 2023]. At the Federal Reserve, researchers have incorporated ML into their forecasting toolkits on an experimental basis – for instance, using random forests to improve recession probability models and using textual analysis (with NLP algorithms) on the Fed's Beige Book and news to augment economic forecasts. These examples illustrate a broad trend: AI/ML methods are no longer theoretical exercises but are being deployed in real forecasting contexts, especially to tackle problems of large data and structural change [Bareith et al. 2024].

How do these AI-driven forecasts compare to traditional approaches in the uncertain environment of recent years? The evidence so far is that AI/ML can substantially improve predictive performance in certain dimensions, but they are not uniformly superior in all cases and still work best in conjunction with human insight. In "normal" times, simpler models often hold their own – indeed, a well-known result from the forecasting literature is that with stable relationships, a simple linear model or even a random walk can be hard to beat. But in the chaotic period of 2020–2022, there have been clear instances where ML models captured turning points or complex drivers better. For example, as noted, a gradient-boosted tree was able to account for time-varying, nonlinear relationships during the U.S. pandemic recovery and beat linear models in forecasting unemployment. Similarly, the use of large information sets combined with ML (factor models with penalization) yielded improved accuracy for inflation when old models failed to anticipate the surge. However, ML is no panacea: some central bank evaluations found that when structural breaks are extreme, no model – not even ML – had enough training data to predict the shift. For instance, during the initial COVID shock, many ML models trained on pre-2020 data also failed until they could be retrained with pandemic data included. Moreover, certain ML models like standard neural networks require enormous data to train effectively, and macroeconomic sample sizes are limited. This is why methods like tree ensembles and regularized linear models (which are more data-efficient) often performed better than very complex deep learning in macro forecasting competitions (the IMF study where simple LASSO beat newer nonlinear models in 2022 inflation forecasting is telling).

The consensus emerging is that AI/ML methods are powerful complements to, rather than outright replacements for, traditional models and economist judgment. They excel at pattern recognition and handling complexity – providing new "eyes" on the data. They might pick up an incipient trend or correlation that a human or simple model misses (for example, an ML model might detect that an uptick in online job postings plus rising used car prices is an early predictor of broader inflation). But human forecasters still play a crucial role in guiding these models, choosing sensible input features, and interpreting outputs (especially when the outputs seem counterintuitive). In practice, many institutions now run hybrid forecasting processes: a structural model might produce a baseline, and an ML model produces an alternative forecast, and experts will compare the two, understand why they differ, and often come up with a reasoned synthesis. In uncertain times, having these diverse model outputs enriches the information set for decision-makers. The rise of AI thus should be seen as an expansion of the forecasting arsenal. As Mullainathan and Spiess [2017] pointed out, machine learning excels at pure prediction tasks and can discover complex predictive signals, whereas traditional econometric models excel at interpretation and theory-consistent structure – blending the two can yield the best of both worlds. Recent experience validates this: the best forecasting performance has often come from combining a large and relevant information set (where ML helps sift signals) with sound judgment and parsimony (where human economists ensure interpretability and avoid nonsense).

RISKS AND LIMITATIONS OF AI

Although AI/ML models are a powerful tool, especially in terms of data processing, one should not forget the limitations that accompany their use, such as interpretability and transparency, overfitting and robustness, data quality and availability, big data dependency, and transparency, accountability, and ethical concerns [IMF 2024]. A central challenge lies in the interpretability of complex AI models, which often function as "black boxes". Unlike traditional econometric approaches, where the influence of each variable can be clearly identified and explained, many AI models provide little insight into the mechanisms behind their predictions, what is problematic in policy contexts where decision-makers must justify forecasts to stakeholders and base decisions on traceable reasoning. For example, if an AI model is showing a rise in inflation, then policymakers should know whether this reflects wages, commodity prices, or other factors, and how these drive overall prices.

Efforts to improve interpretability, such as the use of Shapley values or feature attribution methods, can be employed to decompose model outputs into input contributions. However, these approaches tend to detect correlation more than causation, limiting their explanatory power, and thus most institutions have responded by embracing hybrid approaches, combining AI insights with both structural models and expert judgment, in order to maintain both performance and intelligibility [Desai 2023]. Robustness of the model is also a critical issue, as highlighted above, given that ML models are highly susceptible to overfitting, especially in data-poor environments like macroeconomics. Model predictions that perform well on historical data can be disastrous when faced with structural breaks or regime shifts, as was seen in the COVID-19 pandemic. Data quality and availability is another concern, especially if we know that the quality and comparability of data is often questionable, and AI relies on the use of large databases. Many high-frequency or alternative datasets are noisy, short in duration, or subject to revision.

Finally, concerns about accountability, ethical data use, and model management remain. That is why institutions are increasingly emphasizing explainability, rigorous validation, and in particular human oversight to ensure the responsible application of AI in forecasting.

CONCLUSIONS

In conclusion, the upheavals of the past few years have ushered in a new paradigm for macroeconomic forecasting – one that is more data-intensive, probabilistic, and cognizant of uncertainty. Forecasting in the post-COVID world is undeniably more challenging; unprecedented shocks taught us to expect the unexpected and to be humble about our models' limitations. But at the same time, those challenges acted as a catalyst for innovation. Forecasters have broadened their toolkit to include high-frequency indicators, non-traditional data, and powerful AI/ML algorithms, enabling them to adapt in real time and capture complex dynamics better than before. Traditional models like DSGEs and VARs have not been thrown out – they still provide valuable theoretical consistency and a baseline for thinking – but they are now augmented by adaptive methods that can handle regime changes and nonlinearities [Liu 2024]. The integration of AI and machine learning has been a central development: as we have detailed, these methods have improved forecasting of nowcasted GDP, inflation, and other variables in volatile conditions, often outperforming legacy approaches and providing new insights (for example, through analysis of text or disaggregate data). Major institutions from the ECB to the Fed and IMF have not only acknowledged this shift but actively incorporated it, blending human expertise with machine computations in their forecasting process.

The net effect is that macroeconomic forecasts today are arguably more robust and information-rich than those prior to 2020. A central bank forecasting round now might include input from an AI model flagging an uptick in online job postings that presages hiring, a textual analysis summarizing thousands of business reports into a sentiment index, and scenario drills of various tail risks – all alongside the conventional outlook. Forecasters have transformed from simply extrapolating trends to managing an ensemble of models and data streams, navigating through uncertainty with a combination of technological assistance and seasoned judgment. This evolution bodes well for dealing with future shocks: whether it is a climate-related event, a technological disruption, or another pandemic, the forecasting community is now better equipped to respond quickly and flexibly.

Of course, challenges remain. As we've argued, care must be taken to ensure models remain transparent, generalizable, and anchored in economic reality [Almosova, Andresen 2022]. Ongoing research will be needed to improve explainability and to prevent the misapplication of AI (for example, avoiding false confidence in predictions). Yet, the trajectory is clear – macroeconomic forecasting is becoming more of a high-tech endeavor, continuously updated with real-time data and refined with intelligent algorithms, all under the guidance of human expertise. AI/ML outputs should complement, not replace, structural models and expert judgment. Such a hybrid approach combines flexibility with theoretical grounding. Far from making human forecasters obsolete, the AI revolution in forecasting is reshaping their role – from solitary practitioners of "the dismal science" into collaborative teams of economists and machines working together to decipher an ever more complex economic landscape.

REFERENCES

Aldasoro I., Hördahl P., Schrimpf A., Zhu X. S. (2025) Predicting Financial Market Stress with Machine Learning. BIS Working Papers 1250, BIS.

Almosova K. Andresen M. (2022) Non-linear Inflation Forecasting with Recurrent Neural Networks. Journal of Forecasting, 41(8), 1250-1269. https://doi.org/10.1002/for.2901

Bareith H., Adjei E., Hassan M. Kokot W. (2024) Navigating Inflation Challenges: AI-Based Portfolio-Management Insights. Risks, 12(3), Art. 46. https://doi.org/10.3390/risks12030046

Cárdenas-Cárdenas J. A., Cristiano-Botia D. J., Martínez-Cortés N. (2023) Colombian Inflation Forecast using Long Short-Term Memory Approach. Borradores de Economía, 1241.

Czech National Bank (2025) First Use of AI in Inflation Forecasting at the CNB. CNB Research Note.

Desai A. (2023) Machine Learning for Economics Research: When What and How? Staff Analytical Note 2023-16. Bank of Canada.

Faria-e-Castro M., & Leibovici F. (2024) Artificial Intelligence and Inflation Forecasts. Federal Reserve Bank of St. Louis Review, 106(12), 1-14.

IMF (2024) Mending the Crystal Ball: Enhanced Inflation Forecasts with Machine Learning. IMF Working Paper.

Joseph A., Potjagailo G., Chakraborty C., Kapetanios G. (2024) Forecasting UK Inflation bottom up. International Journal of Forecasting, 40(4), 1521-1538.

Jouilil M., Iaousse A. (2023) Comparing the Accuracy of Classical and Machine-Learning Methods in Time-Series Forecasting: a Case Study of USA Inflation. Statistics, Optimisation & Information Computing, 11(3), 1767-1784. https://doi.org/10.19139/soic-2310-5070-1767

Lenza M., Moutachaker I., Paredes J. (2023a) Density Forecasts of Inflation: a Quantile Regression Forest Approach. European Central Bank, Working Paper Series, 2830.

Lenza M., Moutachaker I., Paredes J. (2023b) Forecasting Euro Area Inflation with Machine-Learning Models. European Central Bank, Research Bulletin, 112.

Li J. (2024) Inflation Forecasting Using a Hybrid LSTM-SARIMA Model Based on Discrete Wavelet Transform. Advances in Economics, Management and Political Sciences, 73, 1419-1434. https://doi.org/10.54254/2754-1169/73/20231419

Liu Y. (2024) Forecasting China's Consumer Price Index (CPI) based on Combined ARIMA-LSTM Models. Frontiers in Business Economics and Management, 3(2), 49-67. https://doi.org/10.54097/v49dwv67

Lopez-Lira A., Tang Y. (2024) Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models, arXiv preprint, arXiv:2304.07619.

Mullainathan S., Spiess J. (2017) Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives, 31(2), 87-106.

Tchoketch-Kebir A. Madouri A. (2024) Research Leadership and High Standards in Economic Forecasting: Neural-Network Models Compared with Etalon ARIMA Models. Business Ethics and Leadership, 8(1), 220-233. https://doi.org/10.61093/bel.8(1).220-233.2024