MacroEconVue: A Privacy-Preserving Approach for Microfounded Inflation Forecasting

Jawand Singh
Computer Science
William & Mary
Williamsburg, VA
jsingh07@wm.edu

***Abstract*—Accurate and timely insight into inflation dynamics is essential for effective monetary policymaking. Yet, official inflation indicators are typically released only on a monthly basis and lack demographic granularity, limiting policymakers' ability to react swiftly to economic shifts or assess heterogeneous impacts across population groups. The MacroEconVue project proposes a novel, privacy-preserving framework that leverages federated learning (FL) to integrate microeconomic data into inflation forecasting. By enabling decentralized training of machine learning models directly on user devices, FL allows for the incorporation of sensitive, high-frequency microeconomic data without compromising individual privacy. This paper establishes baseline forecasting performance using traditional autoregressive models and long short-term memory (LSTM) networks. github.com/JawandS/macroeconvue/tree/main/modeling/modeling**

***Index Terms*—Nowcasting Inflation, Machine Learning, Federated Learning, Computational Economics**

I. Introduction

A. Economic Relevance

Inflation refers to a sustained rise in the general price level of goods and services within an economy, diminishing the purchasing power of income and altering consumption behavior. Accurately measuring and forecasting inflation is critical for a range of stakeholders, including central banks, businesses, and consumers, as it directly influences monetary policy, investment decisions, and wage negotiations. In the United States, maintaining price stability is a core mandate of the Federal Reserve (Fed) [1], which relies primarily on the Personal Consumption Expenditures (PCE) price index to monitor inflation trends. In the aftermath of the COVID-19 pandemic and its associated economic volatility, demand has grown for more timely and granular inflation metrics. Notably, the Federal Reserve Bank of Cleveland now provides real-time inflation nowcasts to supplement lagging indicators [3], and recent research from Germany has demonstrated that machine learning models applied to high-frequency scanner data can yield competitive forecasting performance [4].

B. Machine Learning for Economics

Conventional economic models are typically grounded in domain expertise and are designed as simplified representations of complex real-world dynamics. In contrast, machine learning offers a data-driven alternative, wherein models are inferred from large volumes of observations, allowing for the approximation of intricate systems that may be difficult to capture through theoretical abstraction alone [5]. Within macroeconomics, several studies have demonstrated that machine learning techniques can replicate, and in some instances outperform, traditional econometric models in terms of forecasting accuracy and robustness [6]. However, machine learning should not be viewed as a wholesale replacement for traditional approaches. The integration of domain knowledge remains essential for the selection, implementation, and interpretation of machine learning models [5]. A particular challenge in macroeconomic forecasting arises from the prevalence of non-stationary time series data, where the statistical properties evolve over time. Standard machine learning models often struggle with such data, resulting in forecasting errors that must be carefully mitigated [7]. Additionally, the often opaque nature of machine learning models poses interpretability concerns, which are especially problematic in policy-oriented economic research where transparency and explainability are paramount [8].

C. Research Gap

While prior studies have applied machine learning techniques to nowcast macroeconomic indicators such as inflation, the potential of federated learning remains largely unexplored in this context. Existing approaches typically rely on centralized datasets, which pose significant limitations in terms of data privacy, representativeness, and timeliness. To date, there has been limited investigation into how federated learning, which enables decentralized model training on sensitive, high-frequency data, can be employed to enhance the accuracy and granularity of inflation forecasting. This paper addresses this gap by introducing a federated learning framework for macroeconomic nowcasting, offering a privacy-preserving alternative to conventional methods.

II. Related Works

A. Macroeconomic Random Forest

In the area of machine learning for economic modeling one promising innovation is the macroeconomic random forest (MRF) [9]. Fundamentally, the MRF seeks to create a model that is both accurate and interpretable through its modification of the traditional random forest (RF). Rather than estimate a single average value at the leaf of decision like a RF the MRF includes a linear regression in each leaf. As a result, the MRF's key output is generalized time-varying parameters (GTVP) which are economically interpretable values with economic significance. To elaborate, the MRF is able to identify the underlying structure of the time series data without a priori knowledge by automatically selecting the best feature to split the data on. In order to minimize the variance of each decision tree, they are combined into a random forest. Furthermore, each tree is grown from a random subsample (bagging) and a subset of factors is used at each step to prevent the same trees from emerging (decorrelation) [9]. In the context of MacroeconVue the MRF can be used to create accurate and interpretable nowcasts of consumer inflation while ingesting private data through federated learning.

B. Nowcasting World Trade

One example of the application macroeconomic random forest was in the three-step approach of nowcasting world trade [10]. In this study both tree-based and regression-based (such as the MRF) machine learning models were evaluated. Based on their findings, the MRF was found to outperform the other models tested and existing nowcasts of world-trade. Beyond the model selection the study also developed a framework to approach nowcasting of macroeconomic variables. First, they identified the most informative predictors through pre-selection. In this step they use the least angle regression to eliminate the least correlated predictors, which they found to boost the end model performance relative to no pre-selection. Next, the study performed factor extraction through principal component analysis in order to summarize information and reduce noise which further improves the overall model performance. In order to account for the usage of mixed-frequency data, lagged data is moved forward and multiple series are created from data with advanced data. Finally, the factors are used in a MRF which was found to significantly outperform more traditional tree based machine learning [10]. As noted in the paper, one future direction is to applying the three-step methodology on other macroeconomic variables which MacroeconVue seeks to do.

C. Federated Learning

Often times in real-world problems involving machine learning there is a key trade off between utility and privacy. This means that sensitive data is often essential in machine learning applications but is often too sensitive to share with a central entity. In order to address this, the framework of federated learning (FL) is presented as a way of using sensitive data without ever transmitting the private data [12]. Instead, the results of training the model (such as gradient changes for a neural network) are shared while training is actually done locally on the private data. Several key techniques are involved in the secure aggregation of such data and several protocols are implemented to ensure user privacy. For example, double-masks are applied to the results being shared so that the original results can not be reconstructed, even by a malicious server [12]. Furthermore, results are encrypted so that only the aggregate results of training can be decrypted, nothing individually. Beyond secure aggregation, distributed differential privacy is also applied to reinforce user privacy. With distributed differential privacy the chance of model memorization (where the model is over influenced by one data point) is reduced. This means that even if a user provides data that is a significant outlier, differential privacy means their results can not be recreated while improving the model performance. Overall, although there is some concern of a malicious server compromising the process, currently FL allows significant privacy reassurance [11]. As a result, MacroeconVue can leverage highly sensitive individual data (such as receipts, credit/debit transactions, bank statements, etc.) without that data ever leaving the user's device.

D. Federated Learning for Forecasting

While FL is a framework that allows for a variety of specific machine learning models to be used there are still specific concerns that need to be addressed when it comes to using FL for nowcasting. For example, one concern is the necessary use of highly heterogeneous data for a single global model [13]. Therefore, one proposed alternative is to cluster clients based on similarities in features and essentially create separate global models for each cluster [13]. The proposed federated averaging algorithm does have some challenges in feature selection that may be addressed based on some of the pre-selection and factor extraction techniques used in nowcasting world trade [10]. Furthermore, previous work has shown how federated learning can be used for the production of official statistics, with similar performance as centralized learning methods [14]. Finally, FL needs to be adapted to other machine learning architectures beyond neural networks or their variants. One approach is with Federated Forest, which creates a lossless learning framework for random forests while maintaining privacy [15]. In the context of MacroeconVue, this works hows how FL can be applied to official statistics and adapted for heterogeneous time series data for a tree-based machine learning architecture.

III. Data Selection

A. Inflation Overview

Inflation denotes a sustained rise in the overall price level of goods and services in an economy. A primary tool for tracking inflation is the Consumer Price Index (CPI), which functions not only as a key economic indicator but also as a benchmark for adjusting the dollar value of federal programs, including Social Security benefits [16]. However, the CPI has inherent limitations. It measures price changes based on a fixed basket of goods and services, which may not accurately reflect shifts in consumer behavior or actual changes in the cost of living [17]. Despite these shortcomings, the CPI remains a central and widely adopted metric with significant policy relevance, particularly in indexing government payments and contracts.

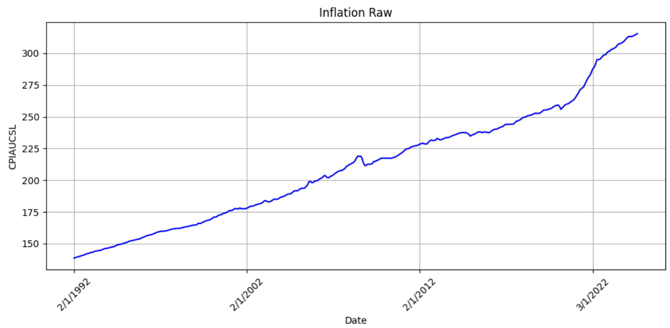

Fig. 1. Inflation level (based on the CPIAUCSL index) since 1992.

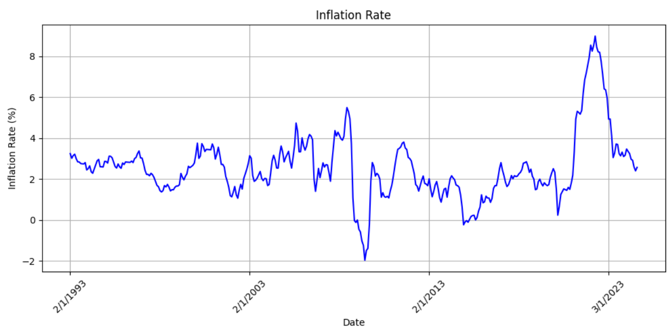

Fig. 2. Inflation Rate (percent change in CPAIUCSL) since 1993.

B. CPIAUCSL

Among the various CPI measures available, this paper focuses on the Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (CPIAUCSL), published by the Federal Reserve and accessible via the FRED database [18]. According to the Bureau of Labor Statistics, CPIAUCSL covers approximately 88% of the U.S. population, making it a highly representative and economically significant indicator. For this analysis, the monthly CPIAUCSL series is transformed into a year-over-year percentage change at each time point. This derived series is referred to throughout the paper as the "inflation rate."

C. Stationarity

Stationarity refers to the property of a time series in which statistical characteristics—such as mean and variance—remain constant over time. Non-stationary data can negatively affect model reliability, particularly in time series forecasting. While the raw CPIAUCSL series exhibits non-stationary behavior, its transformation into the year-over-year inflation rate yields a stationary series, making it more suitable for modeling purposes. The stationarity of the inflation rate series is evaluated using the Augmented Dickey-Fuller (ADF) test, a standard statistical procedure for assessing whether a time series exhibits constant mean and variance over time. The ADF test reports a test statistic and corresponding p-value to determine the presence of a unit root. For the raw CPIAUCSL series, the ADF statistic is 1.372 with a p-value of 0.997, strongly indicating non-stationarity. In contrast, the transformed inflation rate series yields an ADF statistic of -3.435 and a p-value of 0.01, providing strong evidence in favor of stationarity.

D. FRED-MD

A wide range of economic indicators is publicly available through the Federal Reserve Economic Data (FRED) portal, maintained by the Federal Reserve Bank of St. Louis [19]. Among these, a curated subset of monthly time series is compiled into the FRED-MD database [20], which is also managed by the St. Louis Fed and widely used in empirical macroeconomic research [21]. After preprocessing, specifically, removing rows with missing values—the resulting dataset consists of 126 economic indicators observed monthly over 380 periods, beginning in 1993.

IV. Modeling Experiment

A. Experiment Setup

To establish baseline performance for traditional and machine learning forecasting methods, four models were implemented: Autoregressive (AR), Vector Autoregressive (VAR), Long Short-Term Memory (LSTM), and Autoregressive LSTM (AR-LSTM). Each model was tasked with forecasting the inflation rate at five distinct horizons, 4, 16, 28, 40, and 52 months, to assess performance under both short- and long-term scenarios, including the post-COVID inflation spike. During training, each model had access only to data preceding the forecast window, defined as November 11, 2024 minus the respective horizon length. Model predictions were evaluated against actual inflation rate values using Root Mean Squared Error (RMSE), which is particularly useful for penalizing larger errors more heavily [22]. Although Mean Absolute Error (MAE) was also computed, RMSE served as the primary evaluation metric. To identify the best-performing configuration for each model at each forecast horizon, a grid search over relevant hyperparameters was conducted. The specific modeling approaches and their optimal configurations are detailed in the sections that follow.

B. Autoregressive Model

Description An AR model is a simple and foundational time series model often used in economics as a baseline when comparing to other models. In an AR model the current value of a variable is expressed as a linear combination of its past values, known as lags. Formally, an AR model of order p, denoted AR(p), uses the previous p observations as predictors in the regression equation [23]. A key assumption of AR models is that the underlying time series must be stationary, i.e., its statistical properties such as mean and variance must remain constant over time, for the model to yield valid and reliable forecasts.

Hyperparameters As the AR model is univariate, only the inflation rate series was used as input. Forecast horizons ranging from 1 to 52 months and lag lengths from 1 to 24 were initially explored. For consistency with the other models, a subset of lag values, 1, 2, 4, 6, 12, and 24, and forecast

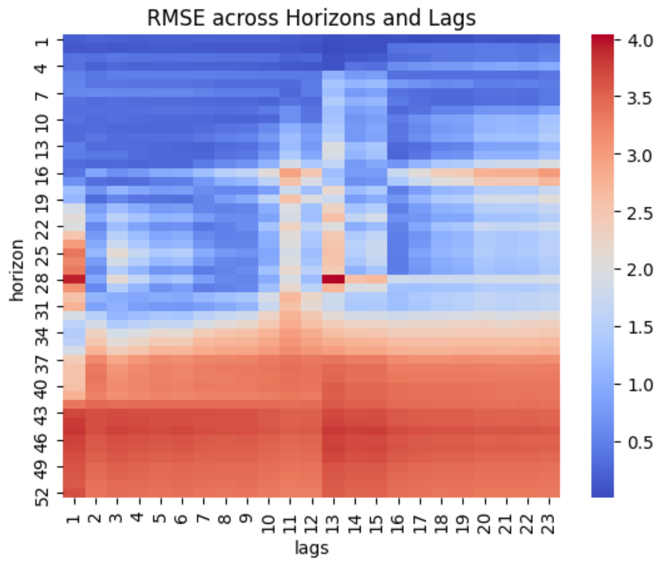

Fig. 3. Heatmap of RMSE Values for AR Model at various Lags and Horizons.

horizons, 4, 16, 28, 40, and 52 months, were selected for comparative analysis.

Several standard techniques exist for determining the optimal lag length, including examination of the partial autocorrelation function, information criteria (e.g., AIC, BIC), and manual search methods. This paper adopts a manual grid search approach [24]. Importantly, the optimal lag structure cannot be known in advance for unseen data. As such, each model is evaluated based on its best-performing configuration, its "ideal" version, at each forecast horizon. Analysis of average RMSE across lag lengths reveals no single lag value that consistently outperforms others, highlighting the context-specific nature of lag selection.

C. Vector Autoregressive Model

Description Similar to the AR model, the VAR model is a canonical model in macroeconomics, often used as a benchmark when comparing other models. In particular, VARs are used for forecasting and have an economically interpretable structure. Like an AR model, a VAR uses past values to predict the present. However, a VAR model is also multivariate meaning each variable is modeled as a linear function of its own lags as well as the lagged values of all other variables being considered. [25].

Principal Component Analysis (PCA) For the VAR model, the FRED-MD dataset was further processed using Principal Component Analysis (PCA). PCA reduces the dimensionality of the dataset by transforming the original, potentially correlated variables, into a smaller set of uncorrelated components known as principal components [26]. This transformation serves to simplify the input space and mitigate multicollinearity, high correlation among explanatory variables, which can negatively impact the stability and interpretability of certain time series models, including VAR [27]. Due to the nature of the FRED-MD dataset high multicollinearity is very likely since many economic indicators would be highly correlated. As a result, PCA was tested with the expectation that a reduction in multicollinearity would boost model performance.

Hyperparameters For the VAR model, three hyperparameters were considered: (1) the number of lags, (2) the proportion of variance retained during PCA transformation, and (3) whether the input variables were normalized prior to applying PCA. To analyze the impact of these hyperparameters on model performance, outliers, defined as observations exceeding 1.5 times the interquartile range, were removed. A regression analysis was then conducted with RMSE as the dependent variable and the three hyperparameters, along with forecast horizon, as independent variables. Notably, the results indicate that normalizing the data prior to PCA is associated with a statistically significant decline in model performance. Additional tests were conducted using lag values of 1, 2, 4, 5, 12, and 24, along with PCA variance thresholds of 1.0, 0.95, 0.9, and 0.8. Analysis of the best-performing hyperparameter combinations across forecast horizons suggests that lower lag values generally yield better predictive performance.

TABLE I
OLS REGRESSION OF RMSE ON VAR MODEL HYPERPARAMETERS

Variable	Coefficient	Std. Error	p-value
Intercept	2.110	(1.276)	0.100
Forecast Horizon	0.062	(0.006)	0.000
Lags	0.057	(0.014)	0.000
Variance in PCA	-2.569	(1.424)	0.073
Normalized (1=Yes)	0.891	(0.217)	0.000
Observations	162		
R-squared	0.459		
Adj. R-squared	0.445		
F-statistic	33.30		

D. Long Short-Term Model

Description LSTMs are frequently used when processing longer sequential data and is often used in time series analysis. Furthermore, LSTMs as a type of Recurrent Neural Network, are designed to effectively capture long-term dependencies through a gated feedback mechanism. With memory cells that allow the model to retain and update information from prior time steps [28]. This paper employs a sequence-to-sequence LSTM architecture, in which a fixed-length input sequence, referred to as the context window, is used to generate a multi-step forecast over a specified horizon. The implemented model consists of two stacked LSTM layers, each with 64 hidden units.

Hyperparameters Prior to training, the input data for the LSTM model was normalized using a min-max scaler to improve convergence and overall model performance. A grid search was then conducted over context window sizes of 1, 2, 4, 6, 12, and 24, and dropout rates of 0.0, 0.1, 0.2, and 0.3, in order to identify the optimal configuration at each forecast horizon. Dropout regularization was employed to mitigate overfitting, an especially important consideration

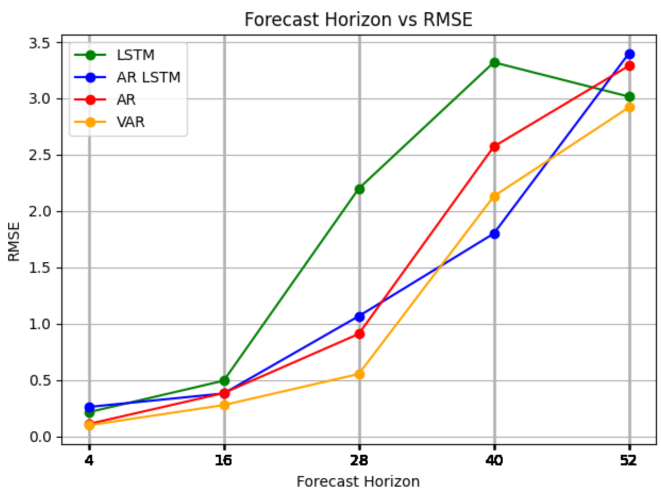

Fig. 4. Comparison of RMSE as Varying Horizons for the Optimal Version of Each Model (AR, VAR, LSTM, AR-LSTM).

given the relatively small dataset used in this study. However, after removing statistical outliers, no consistent relationship was observed between dropout rate and model performance. In contrast, shorter context windows, particularly sizes of 1 or 2, tended to produce better results across all forecast horizons.

Autoregressive LSTM By simplifying the input to a single series (inflation rate) the AR-LSTM was included as a way of contrasting input complexity on model performance. The overall experiment closely followed the setup of the standard LSTM model, with two key modifications. First, instead of using the multivariate FRED-MD dataset, the AR-LSTM was trained exclusively on the univariate inflation rate series. Second, rather than employing a sequence-to-sequence approach, the model was configured to perform one-step-ahead forecasting: the predicted value at each step was recursively fed back as input to generate subsequent forecasts, mimicking the behavior of a traditional autoregressive process. The same hyperparameter grid, context window sizes and dropout rates, was used for model selection.

V. Modeling Results

A. RMSE Comparison

Figure 4 presents the Root Mean Squared Error (RMSE) for the best-performing configuration of each model across all forecast horizons (4, 16, 28, 40, 52). These horizons were selected starting at a four month horizon since forecasting is often done quarterly and at a monthly frequency, the full quarter's effects would be noticeable in the fourth month. The following horizons were selected at year intervals for longer-term projections. The optimal hyperparameters identified at the four month horizon are as follows (with additional results available in the online code appendix):

- AR: lags = 12
- VAR: lags = 4, PCA variance = 0.8, normalization = True
- LSTM: context window = 2, dropout rate = 0.3
- AR-LSTM: context window = 24, dropout rate = 0.1

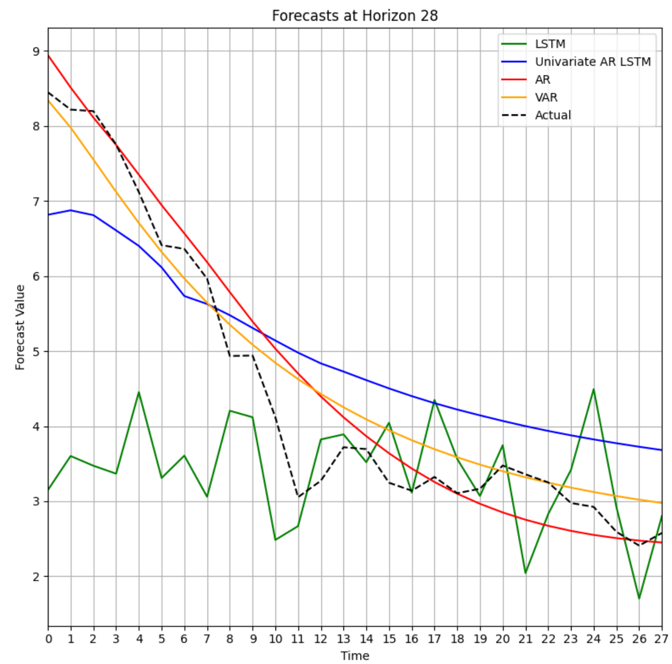

Fig. 5. Forecasts at a 28 Month Horizon for Each Model (AR, VAR, LSTM AR-LSTM).

Overall, the VAR model demonstrates consistently strong performance across forecast horizons. In contrast, the LSTM model exhibits a steep increase in RMSE as the horizon lengthens, indicating diminished predictive capability over longer intervals. For example, at horizons of 28 and 40 months the optimal LSTM had approximately 300% and 80% greater RMSE respectively than the optimal VAR. The AR model performs competitively at shorter horizons but, like the LSTM, shows noticeable degradation as the forecast horizon extends.

B. Forecasts

Figure 5 illustrates the forecasts generated at a 28-month horizon, capturing the inflation trajectory during the post-COVID-19 recovery period following widespread supply chain disruptions. Both the AR and VAR models generally track the direction and magnitude of actual inflation, demonstrating reasonable alignment with observed trends. The AR-LSTM model partially captures the downward trajectory, though with less precision. In contrast, the LSTM model fails to maintain coherent structure, with its forecasts exhibiting erratic fluctuations that resemble noise, indicating a loss of predictive capacity at this extended horizon.

At a forecast horizon of 52 months, the full extent of the inflation shock is clearly observable, as seen in Figure 6. At this time horizon none of the models were able to capture the sharp and rapid increase in inflation. Although the LSTM briefly exhibits an initial upward spike, it quickly devolves into erratic fluctuations, suggesting that it fails to meaningfully capture the magnitude or persistence of the shock.

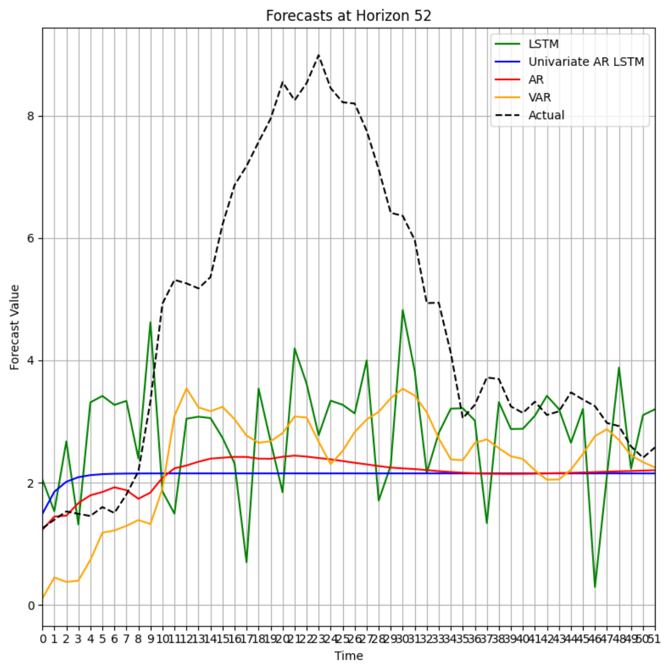

Fig. 6. Forecasts at a 52 Month Horizon for Each Model (AR, VAR, LSTM AR-LSTM).

C. Discussion

It is important to emphasize that, in most real-world forecasting applications, shorter time horizons are typically of greater practical relevance and reliability. Predictive accuracy tends to degrade as the forecast horizon increases, owing to the compounding of uncertainty and noise. Nonetheless, this paper deliberately explores longer horizons, including 28 and 52 months, to examine the robustness of traditional and neural forecasting methods in the face of rare but consequential macroeconomic shocks. The results clearly indicate that none of the evaluated models, AR, VAR, LSTM, or AR-LSTM, were able to anticipate or adapt to the sharp inflation spike following the COVID-19 supply chain disruptions. Even when analyzing forecasts at the early onset of the shock, model outputs remained anchored to historical patterns and failed to adjust meaningfully to emerging trends.

This limitation stems from a fundamental challenge in time series modeling: exogenous shocks are, by definition, events not present in historical data and thus unobservable to purely data-driven models. Without access to real-time or leading indicators that correlate with such disruptions, models are effectively blind to their onset and trajectory. This highlights a critical need for integrating high-frequency or alternative data sources, such as transaction-level financial data, supply chain analytics, or news sentiment, to enrich traditional macroeconomic datasets. It is likely that the lack of high-frequency data played a key part in the LSTM's performance degradation, particularly compared to other models that might be more robust when trained on smaller datasets. From a systems design perspective, future forecasting frameworks may need to incorporate dynamic, multi-resolution inputs or hybrid models that combine statistical learning with structured domain knowledge. Absent such innovations, even the most sophisticated algorithms will remain limited in their capacity to predict and respond to abrupt regime changes.

VI. Modeling Next Steps

Several directions may further strengthen and extend the findings of this study:

1) **Expanded Hyperparameter Search**: The current grid search explores a limited set of hyperparameter values for each model. Future work could employ more comprehensive search strategies, such as randomized search, Bayesian optimization, or evolutionary algorithms, to explore larger and more complex hyperparameter spaces, potentially yielding improved model performance across horizons.
2) **Rolling Window Evaluation**: Rather than evaluating models at fixed forecast points, future experiments could implement a rolling-window framework, averaging prediction errors over multiple overlapping test periods. This would provide a more robust measure of model stability and generalization over time.
3) **Model Diversification and Architectural Variation**: Additional models, including ensemble methods, Transformer-based architectures, or state space models, should be introduced to benchmark performance under varying data conditions. Existing neural models could also be extended by modifying layer depth, recurrent unit types (e.g., GRU vs. LSTM), or attention mechanisms to better capture long-term dependencies.
4) **Integration of Mixed-Frequency and High-Frequency Data**: A critical limitation identified in this study is the inability of models to capture exogenous shocks. Future research should explore the use of mixed-frequency datasets (e.g., weekly commodity prices, real-time financial transactions) and privacy-preserving learning frameworks such as federated learning, which can facilitate access to sensitive, high-resolution data without compromising individual-level privacy.

VII. Conclusion

The MacroEconVue project investigated the comparative performance of traditional statistical models and machine learning (ML) approaches for inflation forecasting, while also exploring the potential of federated learning (FL) to incorporate privately held data. Through a systematic evaluation of AR, VAR, LSTM, and AR-LSTM models, we found that traditional models—particularly AR and VAR—consistently outperformed ML-based models at longer forecast horizons. This result likely stems from the sensitivity of deep learning models to limited data availability and their tendency to overfit small datasets. The integration of a FL framework with an LSTM model represents a novel contribution, illustrating how decentralized learning can facilitate the use of sensitive microeconomic data without compromising user privacy. Overall, the project demonstrates a viable path toward improving macroeconomic forecasting by expanding access to traditionally inaccessible data through privacy-preserving, distributed learning methods.

APPENDIX

A. Modeling Code Repository

The full implementation of data preprocessing, model training, hyperparameter tuning, and evaluation scripts is available at the following GitHub repository: https://github.com/JawandS/macroeconvue/tree/main/modeling/modeling
This repository includes Jupyter notebooks, source code for all models (AR, VAR, LSTM, AR-LSTM), data transformation pipelines, and visualizations corresponding to the results presented in this paper.

REFERENCES

[1] Federal Open Market Committee, "Statement on Longer-Run Goals and Monetary Policy Strategy," Board of Governors of the Federal Reserve System, Washington, DC, USA, Jan. 2024. [Online]. Available: https://www.federalreserve.gov/monetarypolicy/files/FOMC_LongerRunGoals.pdf [Accessed: February 18, 2025]

[2] Board of Governors of the Federal Reserve System, "Inflation (PCE)," Washington, DC, USA, Aug. 2024. [Online]. Available: https://www.federalreserve.gov/economy-at-a-glance-inflation-pce.htm [Accessed: February 18, 2025]

[3] Federal Reserve Bank of Cleveland, "Inflation Nowcasting," Cleveland, OH, USA, Aug. 2024. [Online]. Available: https://www.clevelandfed.org/indicators-and-data/inflation-nowcasting [Accessed: February 18, 2025]

[4] G. W. Beck, K. Carstensen, J.-O. Menz, R. Schnorrenberger, and E. Wieland, "Nowcasting consumer price inflation using high-frequency scanner data: evidence from Germany," European Central Bank, Frankfurt am Main, Germany, Working Paper No. 2930, Apr. 2023. [Online]. Available: https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2930~05cff276eb.en.pdf

[5] O. Simeone, "A Very Brief Introduction to Machine Learning With Applications to Communication Systems," arXiv, Aug. 2018. [Online]. Available: https://arxiv.org/pdf/1808.02342

[6] T. Atashbar, R. A. Shi, "AI and Macroeconomic Modeling: Deep Reinforcement Learning in an RBC model," IMF Working Paper No. 2023(040), Feb. 2023. [Online]. Available: https://www.elibrary.imf.org/view/journals/001/2023/040/article-A001-en.xml

[7] A. Dixit, S. Jain, "Effect of Stationarity on Traditional Machine Learning Models: Time Series Forecasting Perspective," in *Proceedings of the 14th International Conference on Contemporary Computing (IC3)*, Noida, India, Nov. 2021 [Online]. Available: https://dl.acm.org/doi/10.1145/3474124.3474167

[8] A. Desai, "Machine Learning for Economics Research: When, What and How," Bank of Canada, Staff Analytical Note 2023-16, Oct. 2023. [Online]. Available: https://www.bankofcanada.ca/2023/10/staff-analytical-note-2023-16

[9] P. G. Coulombe, "The Macroeconomy as a Random Forst," Mar. 2021. [Online]. Available: https://arxiv.org/pdf/2006.12724

[10] M. D. Chinn, B. Meunier, S. Stumpner, "Nowcasting World Trade with Machine Learning: a Three-Step Approach," Nationa Buerau of Economic Research, Jun. 2023. [Online]. Available: https://www.nber.org/papers/w31419.

[11] F. Hartmann, P. Kairouz, "Distributed differential privacy for federated learning," Google Research, Mar. 2023. [Online]. Available: https://research.google/blog/distributed-differential-privacy-for-federated-learning/.

[12] K. Bonawitz et al., "Practical Secure Aggregation for Federated Learning on User-Held Data," Nov. 2016. [Online]. Available: https://arxiv.org/pdf/1611.04482

[13] F. D. Gonzalez, " Federated Learning for Time Series Forecasting Using LSTM Networks: Exploiting Similarities Through Clustering," KTH Royal Institute of Technology, 2019. [Online]. Available: https://kth.diva-portal.org/smash/get/diva2:1334598/FULLTEXT01.pdf

[14] J. Stock, O. Hauke, J. Weißmann, H. Federrath, "The Applicability of Federated Learning to Official Statistics," Sep. 2023. [Online]. Available: https://arxiv.org/abs/2307.15503

[15] Y. Liu, Y. Liu, Z. Liu, J. Zhang, C. Meng, Y. Zheng, "Federated Forest," IEEE, May 2019. [Online]. Available: https://arxiv.org/abs/1905.10053

[16] Bureau of Labor Statistics. "CPI Questions and Answers." https://www.bls.gov/cpi/questions-and-answers.htm#:~:text=As%20a%20means%20of%20adjusting,to%20millions%20of%20American%20workers.

[17] New York Federal Reserve. "Understanding the CPI." https://www.newyorkfed.org/medialibrary/media/research/current_issues/ci3-6.html

[18] Federal Reserve Bank of St. Louis. "Consumer Price Index for All Urban Consumers (CPIAUCSL)." https://fred.stlouisfed.org/series/CPIAUCSL

[19] Federal Reserve Bank of St. Louis. "FRED Economic Data." https://fred.stlouisfed.org/

[20] Federal Reserve Bank of St. Louis. "FRED-MD Database." https://www.stlouisfed.org/research/economists/mccracken/fred-databases

[21] McCracken, M. W., & Ng, S. (2015). "FRED-MD: A Monthly Database for Macroeconomic Research." Federal Reserve Bank of St. Louis Working Paper. https://s3.amazonaws.com/real.stlouisfed.org/wp/2015/2015-012.pdf

[22] Bellprat, O., et al. (2022). "Reliability of RMSE as an Evaluation Metric." *Geoscientific Model Development*, 15(15), 5481–5500. https://gmd.copernicus.org/articles/15/5481/2022/

[23] ScienceDirect. "Autoregressive Models in Time Series Analysis." https://www.sciencedirect.com/topics/mathematics/autoregressive-model

[24] Enders, W. (1997). "Lag Length Selection in Time Series Models." LSU Working Paper. https://www.lsu.edu/business/economics/files/workingpapers/pap97_27.pdf

[25] Penn State Department of Statistics. "Introduction to VAR Models." https://online.stat.psu.edu/stat510/lesson/11/11.2

[26] IBM. "Principal Component Analysis Explained." https://www.ibm.com/think/topics/principal-component-analysis

[27] IBM. "Multicollinearity in Machine Learning." https://www.ibm.com/think/topics/multicollinearity

[28] NVIDIA Developer. "Long Short-Term Memory (LSTM)." https://developer.nvidia.com/discover/lstm

[29] Bureau of Labor Statistics. "Handbook of Methods" https://www.bls.gov/opub/hom/cpi/presentation.htm?

[30] Flower Framework, "What is Federated Learning"? https://flower.ai/docs/framework/tutorial-series-what-is-federated-learning.html

[31] Flower Framework, "Flower Android Tutorial", https://flower.ai/docs/examples/android.html

[32] Bureau of Labour Statistics, "CPI Databases" https://www.bls.gov/cpi/data.htm