

# MacroeconVue: A Privacy-Preserving Microfounded Approach for Demographic-Based Consumer Inflation Nowcasting

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

Alex Anderson  
Computer Science  
William & Mary  
Williamsburg, VA  
atanderson@wm.edu

**Abstract**—Although many key monetary decisions made by central banks require real-time understanding of market conditions, official inflation metrics are only published on a monthly frequency. Furthermore, there is increased interest in how inflation affects different groups of Americans. Therefore, creating a nowcast (estimates of the present) that includes insights by demographics would further empower policy makers, central banks, and macroeconomists. MacroeconVue proposes a privacy-preserving method of leveraging private data using federated learning, meaning no data would ever leave user devices. This data would be used in a macroeconomic random forest machine learning model to create an accurate and interpretable model of consumer inflation. As a result, the MacroeconVue system would create a high-frequency consumer inflation index that allows for deeper understanding of inflation by subgroup without compromising individual privacy.

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

## I. INTRODUCTION

### A. Economic Relevance

Inflation represents a sustained increase in the general price level of goods and services. This increase often means consumers cannot make their usual purchases with the same amount of income. Understanding inflation is key for policymakers, businesses, consumers, and central banks in order to accomplish their respective goals. In particular, one of the United States Federal Reserve (Fed) mandates is for stable prices [1], meaning the Fed has a vested interest in understanding inflation. Specifically, the Fed uses the Personal Consumption Expenditures (PCE) price index, which shows how Americans allocate their spending. The Fed prefers to use PCE over alternative measures of inflation precisely because it is quicker to adapt to changes in spending patterns [2]. Over time, and particularly after COVID-19's extreme economic changes, greater interest was taken in high-frequency metrics of inflation. For example, the Cleveland Fed provides a public nowcast (estimates of the present) of inflation in order to better inform other models [3]. and a study done in Germany on nowcasting inflation using machine learning on high-frequency scanner data which yielded "competitive results" [4].

### B. Machine Learning for Economics

Traditionally, economic models are constructed using domain knowledge and serve as simplified abstractions of complex phenomena. With machine learning, a sufficient number of observations can be used to construct models of complex systems that may not be easily abstracted otherwise [5]. In the domain of macroeconomics, previous work has been done to replicate existing models with machine learning, in some cases creating models even more robust than traditional versions [6]. It is important to note that machine learning does not always serve as an improvement over traditional methods. Domain knowledge is still relevant when selecting a machine learning model in any context [5]. Furthermore, macroeconomics often uses non-stationary time series data, meaning that the underlying statistical properties of the data change with time. Traditional machine learning models used to forecast non-stationary data can often face significant errors that need to be accounted for [7]. Finally, machine learning models can often be difficult to interpret, which is critical in an economic context [8].

### C. Research Gap

Although previous work has been done on leveraging machine learning for nowcasting of macroeconomic variables, such as inflation, the integration of federated learning with such models has not been explored.

### D. Research Questions

**RQ 1: Privacy** Can we use federated learning to include individual data without that data ever leaving a user's device?

**RQ 2: Accuracy** Can we find and select data sources to create an accurate model of consumer inflation?

**RQ 3: Interpretability** Can we use a macroeconomic random forest model and interpret the resulting coefficients in an economically significant manner?

**RQ 4: Performance** Can we use edge machine learning techniques to minimize any performance impact on user devices?

## 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 MacroconVue 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 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 MacroconVue 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, MacroconVue 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 MacroconVue, 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.

## 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](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.wp2930~05cfe276eb.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," Naciona 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>