Forecasting the Federal Funds Rate: A Comparison of Classical Statistical Models and Machine Learning Models

Simon Schömann, Enrique Lerma, Moritz Wendt, Alberto Marcos

Barcelona School of Economics, Data Science, Barcelona, Spain
December, 2024

Contents

1 Introduction

The COVID-19 pandemic posed unprecedented challenges to the global economy, disrupting well-established macroeconomic relationships and prompting central banks worldwide to adopt extraordinary monetary policy measures. In the United States, the Federal Reserve's federal funds rate became a critical instrument for navigating the economic turbulence. However, the pandemic's unique nature introduced significant uncertainty, as traditional macroeconomic models struggled to account for the highly dynamic and nonlinear economic environment. This project investigates the predictive capabilities of various modeling approaches for forecasting conventional monetary policy, specifically the federal funds rate, in such an extraordinary context.

Forecasting monetary policy and key economic variables during periods of economic disruption is not only academically challenging but also highly policy-relevant. Accurate predictions of the federal funds rate can help policymakers, investors, and businesses anticipate changes in economic conditions and make informed decisions. However, the pandemic has highlighted the limitations of traditional econometric models, which rely on the stability of historical relationships. As macroeconomic relationships evolve under such shocks, machine learning models like Random Forests, with their ability to capture nonlinear patterns, may offer complementary insights.

This project compares two distinct classes of forecasting models: traditional time series models, such as Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA), and Vector Autoregression (VAR), and machine learning techniques, exemplified by Random Forests. While AR, ARIMA, and VAR are grounded in economic theory and assume linear relationships among variables, Random Forests are data-driven and flexible, capable of modeling complex and nonlinear dependencies. However, these machine learning models face challenges, such as potential overfitting and the need for extensive data, particularly when applied to macroeconomic forecasting.

The research focuses on U.S. macroeconomic data, including inflation, GDP growth and unemployment, to generate forecasts during the COVID-19 period. The federal funds rate is the primary variable of interest, chosen for its central role in monetary policy. By comparing forecast accuracy across models using metrics like Root Mean Squared Error (RMSE), the study aims to assess the robustness and generalizability of these approaches.

This project's relevance lies in its ability to bridge traditional and modern forecasting methodologies, shedding light on their strengths and weaknesses in times of macroeconomic upheaval. While traditional time series models are reliable and interpretable, they may falter when textbook relationships break down. Conversely, although machine learning models are powerful in capturing nonlinearities, they require careful tuning and validation to avoid overfitting. By systematically comparing these approaches, the study contributes to understanding their suitability for forecasting in an era of heightened economic uncertainty.

Ultimately, this project offers a framework for forecasting conventional monetary policy during disruptive events, providing insights for researchers and policymakers navigating future crises.
There is also a section in this study that adopts a more systematic approach to feature selection by utilizing McCracken's (2016) FRED complete dataset. In this approach, we employed a LASSO-penalized regression to identify the most relevant predictors for the FFR variable. We also incorporated different powers of the variables to account for potential non-linear relationships in the data. By including polynomial transformations, we allowed the model to capture more complex interactions and dependencies that linear terms alone might miss. LASSO was particularly suited for this task due to its ability to handle high-dimensional data and perform both variable selection and regularization simultaneously.

After identifying the key predictors using LASSO, we implemented a default Random Forest model on the reduced set of features. This two-step process—systematic feature selection followed by machine learning estimation—ensures the most informative predictors will be selected.

2 Data Description

The data for this study were obtained from the Federal Reserve Economic Data (FRED) database. It initially spanned from January 1955 to December 2023; however, the dataset was truncated to begin in January 1985 for reasons discussed in the section on data cleaning and preparation. This adjustment results in a total of 468 monthly observations.

The dataset focuses on five variables:

- **CPI Inflation**: A measure of the rate at which the general level of prices for goods and services is rising. The mean value of CPI is 138.10, with a median of 138.3. The standard deviation is 81.13, indicating significant variation over the observed period. The interquartile range (IQR) is 155.645, and values range from a minimum of 29.37 to a maximum of 309.685.
- **Composite Leading Indicator (CLI) Normalized**: This variable serves as a proxy for GDP, capturing economic activity trends. The mean CLI is 99.96, with a median of 100.09. The standard deviation is 1.29, reflecting relatively minor fluctuations in economic activity over time. The IQR is 1.54, and values range from a minimum of 93.48 to a maximum of 103.41.
- **Unemployment Rate**: The percentage of the labor force that is unemployed but actively seeking employment. The mean unemployment rate is 5.91%, with a median of 5.6%. The standard deviation is 1.70%, and the IQR is 2.3. Unemployment rates range from a low of 3.4% to a high of 14.8%.
- **Effective Federal Funds Rate**: The interest rate at which depository institutions lend balances at the Federal Reserve to other depository institutions overnight. The mean rate is 4.79%, with a median of 4.73%. The standard deviation is 3.67%, and the IQR is 4.65. Rates range from a minimum of 0.05% to a maximum of 19.1%.
- **Brave-Butters-Kelley Leading Index**: This variable was included as part of a robustness check, offering an alternative proxy for GDP. The BBK Leading Index has a mean value of 2.95 and a median of 3.06. The standard deviation is 5.57, and the IQR is 4.22. Values range from a minimum of -71.48 to a maximum of 45.39.

Using GDP proxies such as the Composite Leading Indicator and the Brave-Butters-Kelley Leading Index allowed us to significantly increase the number of available observations compared to using quarterly GDP data. This step was particularly crucial for the Random Forest model, which requires a larger number of observations to avoid overfitting and improve the robustness of results.

This selection of variables captures key aspects of the U.S. economy, providing a comprehensive basis for analyzing and forecasting macroeconomic dynamics.

3 Data Cleaning and Preparation

Our dataset included no missing values. The primary concern in our empirical analysis was the issue of non-stationarity. Non-stationarity can be problematic for certain modeling approaches, particularly traditional time series models such as AR, ARIMA, and VAR. These models assume stationarity in the data, as non-stationary inputs can lead to spurious relationships and unreliable forecasts. To address this, stationarity tests, specifically the Augmented Dickey-Fuller (ADF) test, were conducted on the variables. Non-stationarity is less of a concern for Random Forest models, as they do not rely on statistical assumptions such as stationarity. However, ensuring stationarity in the variables for comparative purposes is essential, as it allows consistent preprocessing across models and avoids potential biases in forecasting evaluations.

The ADF test suggested that the federal funds rate was likely non-stationary. We initially applied first-order differencing to address this; however, after further consideration, we concluded that the dataset should be truncated to begin in 1985. This decision was driven by two key factors: first, the period from 1985 onward still provides a sufficiently large number of observations (468 monthly data points); and second, we aimed to avoid capturing patterns of monetary policy that might no longer be relevant due to structural breaks in the earlier data. Structural breaks are well-documented in the literature. Similarly, the ADF test indicated that CPI inflation and the unemployment rate were likely non-stationary on the updated dataset. To address this, we applied first-order differencing to these variables, ensuring that they met the stationarity requirement for AR, ARIMA, and VAR models. The Composite Leading Indicator (CLI) Normalized and the Brave-Butters-Kelley (BBK) Leading Index were found to be stationary and required no further transformation.

Following the transformation guidelines from McCracken's dataset, appropriate transformations (e.g., first differences, logarithmic transformations) were applied based on predefined codes. Missing values were imputed with column means, and redundant variables were removed, ensuring a robust dataset for modeling. Additionally, observations prior to 1985 were excluded to align with the temporal scope of the analysis.

4 Methodology and Results

4.1 Overview of the Models

The analysis employs a variety of time series models, including Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA), Vector Autoregression (VAR), and Random Forest. Pesaran (2015) lays the theoretical foundation for our macroeconometric modelling. Below, we briefly introduce each model:

- **Autoregressive (AR) Model**: The AR model expresses the current value of a time series as a linear combination of its past values. For an AR(p) model, the equation is:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t,$$

  where y_t is the value at time t, ϕ_i are the coefficients, and ε_t is a white noise error term. AR models are useful for capturing the momentum or mean-reverting behavior of time series data.

- **Autoregressive Integrated Moving Average (ARIMA) Model**: The ARIMA model extends the AR framework by incorporating differencing to achieve stationarity and a moving average (MA) component. An ARIMA(p, d, q) model can be expressed as:

$$\Delta^d y_t = \phi_0 + \sum_{i=1}^{p} \phi_i \Delta^d y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t,$$

  where d denotes the degree of differencing and θ_j are MA coefficients. ARIMA models are versatile for handling trends and seasonality in data.

- **Vector Autoregression (VAR) Model**: The VAR model generalizes the AR framework to multiple interdependent time series. A VAR(p) model is specified as:

$$\mathbf{y}_t = \mathbf{c} + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \varepsilon_t,$$

  where $\mathbf{y}_t$ is a vector of variables, $\mathbf{A}_i$ are coefficient matrices, and ε_t is a vector of error terms. VAR models capture the joint dynamics among variables, making them valuable for multivariate forecasting.

- **Random Forest**: Random Forest is a non-parametric ensemble learning method that combines multiple decision trees to improve predictive accuracy and control overfitting. It does not rely on explicit parametric equations like ARIMA or VAR but learns patterns in the data through recursive partitioning. For the tuning we use time series cross validation from the caret package in R. Normal cross-validation randomly splits the data into training and testing sets, which can lead to data leakage in time series problems by allowing future data to inform the model during training. This violates the temporal structure of time series data, as it assumes independence between observations, which is not valid in sequential datasets. Time series cross-validation in caret accounts for the sequential nature of time series data. Instead of random folds, it creates rolling or sliding training and testing sets. Each fold uses a continuous block of earlier observations for training and a subsequent block for testing. This ensures that test data is always after the training data in time. The configuration of the RF under the Time Slice method is outlined in section 4.2.

4.2 Model Selection

The selection of AR, ARIMA, and VAR models was guided by information criteria, specifically the Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC). In cases where the two criteria suggested different lag orders, we struck a balance between selecting a sufficient number of past values to capture the dynamics of the data and maintaining parsimony to avoid overfitting. This balance was feasible given our relatively large sample size, which allowed for robust estimation while minimizing the risk of over-parameterization. We refer to the code for detailed results of the AIC and SBC calculations. We estimated an AR(10), an ARIMA(5,0,3) and a VAR(5). Under the baseline model we also estimate an AR(1) as a reference.

In this study, we utilized two approaches for the Random Forest model: a default Random Forest and a Time-Slice Cross-Validated Random Forest. These approaches were designed to ensure robust evaluation of model performance while preserving the temporal structure of the data. Similar to the AR, ARIMA, and VAR models, the Random Forest models were iteratively estimated under various modifications to the dataset. Notably, the Zero Lower Bound adjustment was not applied, as the Random Forest models did not predict negative Federal Funds rates, rendering this step unnecessary.

For the default Random Forest, the dataset was divided into three subsets: training, validation, and testing. Three hyperparameters were tuned using the validation dataset: the bootstrap size, representing the proportion of the training dataset used for sampling; the number of variables selected (mtry), which corresponds to the number of predictors considered at each split in the decision trees; and the number of lags, which, while not an inherent hyperparameter of Random Forest, was included to capture potential temporal dependencies in the data.

In the Time-Slice Cross-Validation approach, a fixed-size rolling window was employed, ensuring that the size of the training set remained constant as the window moved forward over time. The hyperparameters tuned in this approach included the size of the training set, which ranged from 70% to 100% of the available observations, and the models were tested on a forecast horizon of 12 periods.
Both Random Forest approaches were configured with 100 trees. The number of lags was set to 5 for the Time-Slice Cross-Validated Random Forest, which was higher than the 3 lags identified as optimal by the default Random Forest. These parameters were deliberately kept relatively low to mitigate the already considerable computational demands of the analysis. The primary tuning parameter, mtry, was optimized over a range from 1 to the total number of predictors in both approaches.

4.3 Model Specifications and Robustness Checks

To ensure robustness, we re-estimated the models with different specifications and imposed additional constraints. For the AR, ARIMA, and VAR models, a post-estimation zero lower bound (ZLB) was applied to the federal funds rate forecasts. The ZLB reflects the empirical reality that nominal interest rates cannot fall below zero under normal circumstances, a constraint that became especially relevant following the global financial crisis and subsequent monetary policy adjustments. Incorporating the ZLB improved the performance of these models by preventing unrealistic negative interest rate forecasts, which would otherwise distort the results and reduce their policy relevance.

The performance of the models under different variable specifications was also evaluated:

- **Replacing the Composite Leading Indicator (CLI):** Substituting the CLI with the Brave-Butters-Kelley (BBK) Leading Index yielded similar forecasting results. This finding suggests that the model outcomes are not spurious and that the results are robust to the choice of GDP proxy.

- **Dropping the Unemployment Rate:** Excluding the unemployment rate improved the performance of the VAR (using the BBK proxy) and it did not diminish the performance of the Time-Slice Cross-Validation Random Forest model (although it did increase the Regular Random Forest RMSE). This aligns with the hypothesis that the rapid increase in unemployment during the COVID-19 pandemic may have introduced substantial forecasting error. The Federal Reserve's response to this surge in unemployment appears to deviate from historical patterns, possibly due counter-cyclical measures taken by the fiscal side. Further examination of this hypothesis should be conducted, potentially including monetary aggregate variables to investigate whether the counteracting effects of monetary expansion could help resolve this puzzle.

4.4 Performance and Results

The models were evaluated using Root Mean Squared Error (RMSE). Across all specifications, the simple AR and ARIMA models consistently delivered the best forecasting performance. This result underscores their ability to effectively capture the underlying dynamics of the federal funds rate, particularly when supported by a long time series and parsimonious parameterization.
Imposing the ZLB post-estimation significantly improved the performance of the AR, ARIMA, and VAR models. By preventing negative forecasts of the federal funds rate, the ZLB constraint not only enhanced accuracy but also ensured the theoretical plausibility of the results.
The inclusion of the BBK Leading Index instead of the CLI produced comparable forecasting results, reinforcing the robustness of the models and confirming that the GDP proxy does not introduce bias into the forecasts. Similarly, excluding the unemployment rate improved forecasting accuracy for the VAR (BBK Version) while not increasing the RMSE of the Time-Slice Cross-Validated Random Forest model. This finding suggests that unemployment dynamics during the pandemic may have behaved atypically, contributing to forecast errors when included.

5 Conclusions

This project has provided insights into the challenges and opportunities of forecasting macroeconomic variables, particularly during periods of significant economic uncertainty. By comparing traditional time series models with machine learning techniques, we have deepened our understanding of the trade-offs between simplicity, interpretability, and flexibility in forecasting approaches.

One key takeaway is that while time series models such as the VAR are able to capture inter-dependencies between several variables, they struggle to perform under rapidly changing macroeconomic conditions or periods marked by structural breaks. The FED not reacting to the unprecedented unemployment increase in the early months of the pandemic posed great challenge to this more structural model class. We showed that simply dropping variables can improve forecasting performance. This is similar to what Lenza and Primiceri (2021) found. They concluded that simply dropping observations from the extraordinary pandemic times helps when estimating VAR models. Conversely, machine learning techniques like Random Forests offer greater flexibility but require careful tuning to avoid overfitting and ensure robust out-of-sample performance. This tension between adaptability and reliability suggests a need for more refined methods that can effectively integrate the strengths of both approaches.

Moreover, this study has highlighted the importance of thoughtful data preparation and model specification, including addressing issues like non-stationarity and structural breaks. These considerations are critical for ensuring that forecasting models remain both theoretically grounded and empirically robust. Looking ahead, future work could explore more advanced hybrid models that combine the interpretability of traditional approaches with the flexibility of machine learning. Additionally, integrating external data sources and incorporating real-time economic indicators may further enhance forecasting accuracy. Ultimately, this project serves as a foundation for refining methodologies and expanding the toolkit for macroeconomic forecasting in increasingly complex and uncertain economic environments.

The LASSO-based model for feature selection on a broader dataset demonstrates superior performance compared to our earlier approach. Its systematic methodology provides a consistent and objective means of selecting variables from large datasets, such as the FRED, making it particularly valuable in high-dimensional settings.

The initial step of feature selection using LASSO highlights the inherent sparsity of the dataset, as only a small subset of the original variables was retained for the optimal LASSO estimation. This reinforces the notion that many predictors in the dataset contribute little to the prediction of the target variable, with only a few being truly relevant.

The selected subset of variables significantly outperformed all other Random Forest estimations in terms of predictive accuracy. However, there remains important theoretical work to be undertaken to better understand the underlying reasons for the predictive power of the selected features. A deeper investigation into the economic or statistical rationale for these variables' influence could provide valuable insights and enhance the interpretability of the model

This approach demonstrates the value of integrating regularization techniques with machine learning models for robust and scalable analysis of large, complex datasets.

References

Lenza, M., and Primiceri, G. E. (2021). How to estimate a vector autoregression after March 2020. *Department of Economics, Northwestern University*
McCracken, M. W., Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. Journal of Business Economic Statistics, 34(4), 574-589.
Pesaran, H. (2015). Time Series and Panel Data Econometrics. *Oxford University Press*

Graphical Appendix

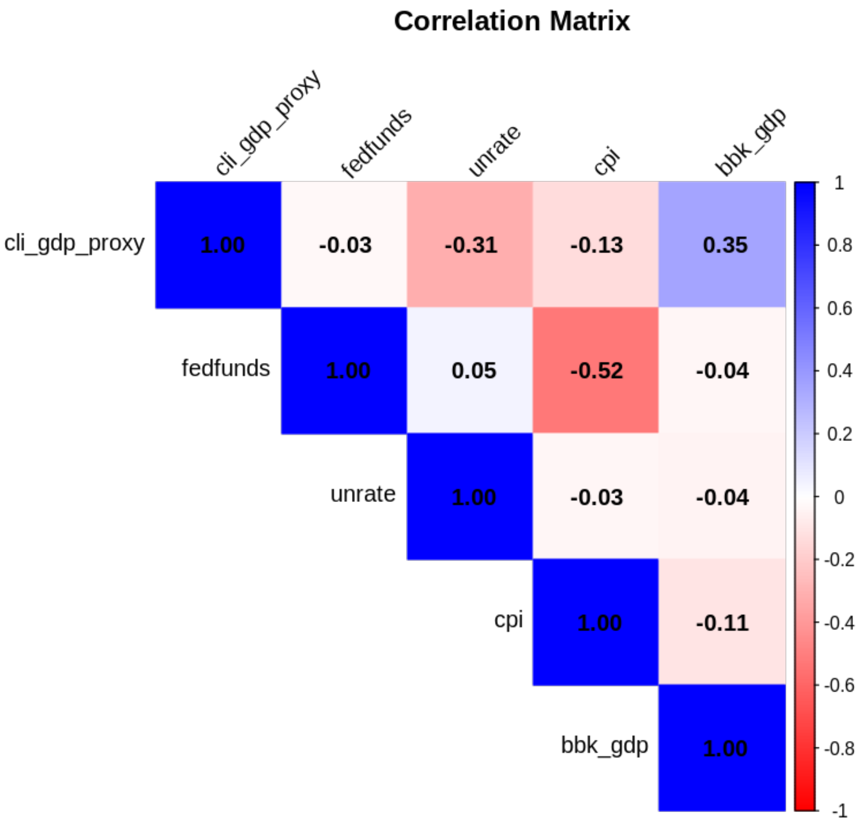


Figure 1: Correlation Matrix

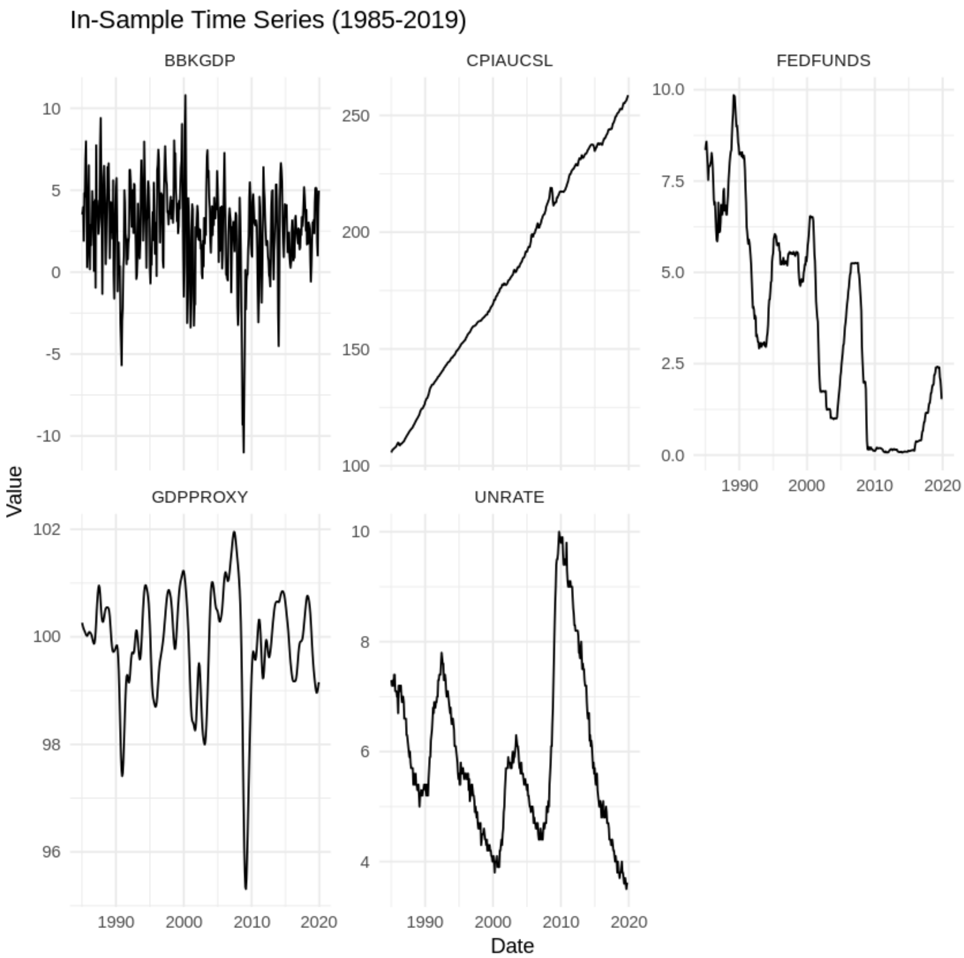
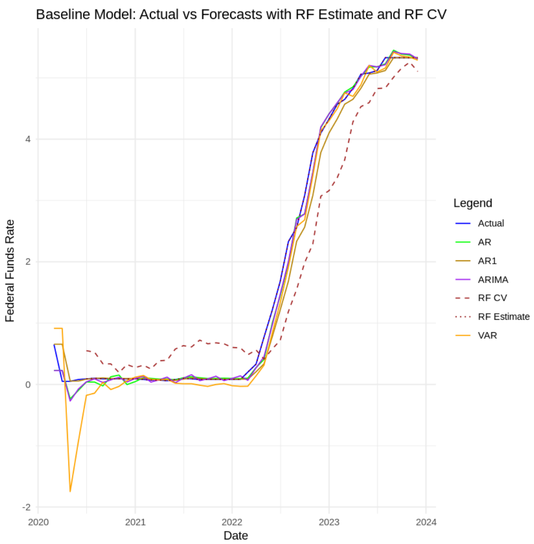
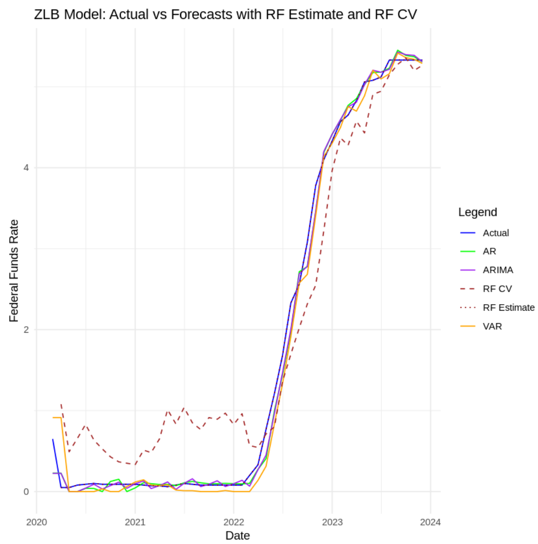


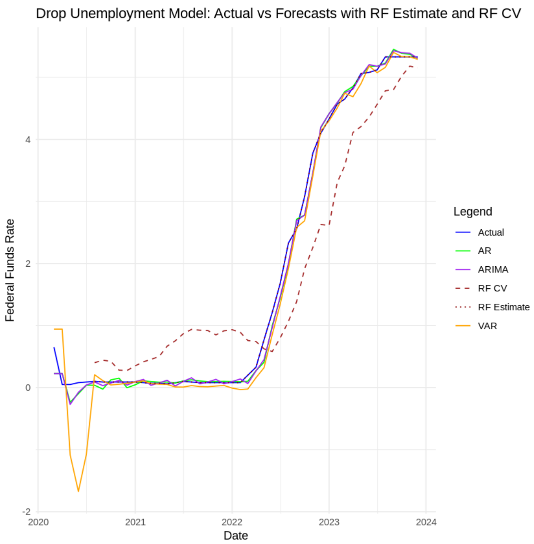
Figure 2: In-Sample Time Series (1985-2019)



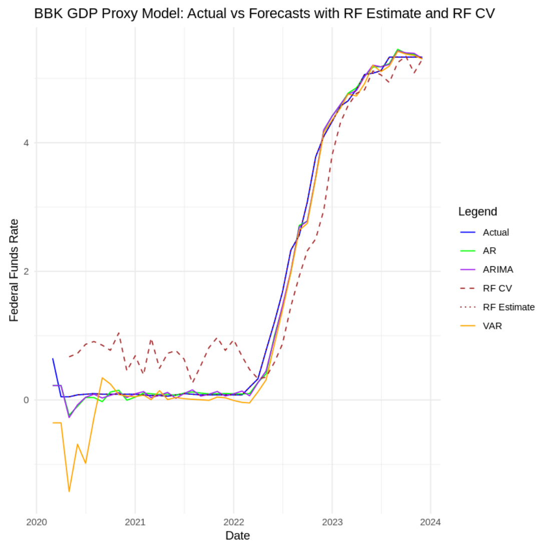
(a) Baseline Model



(b) ZLB Model

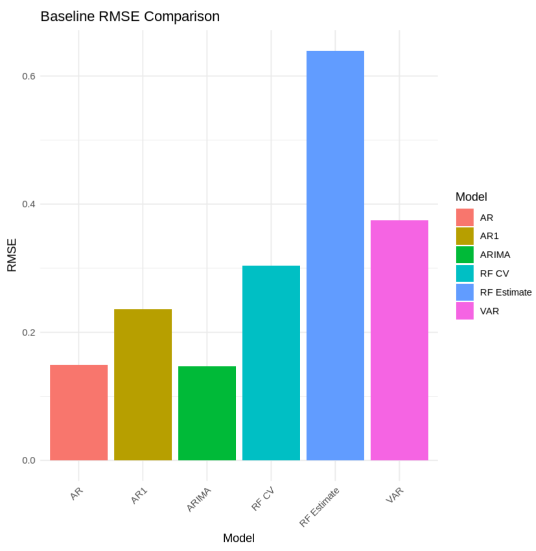


(c) Drop Unemployment Model

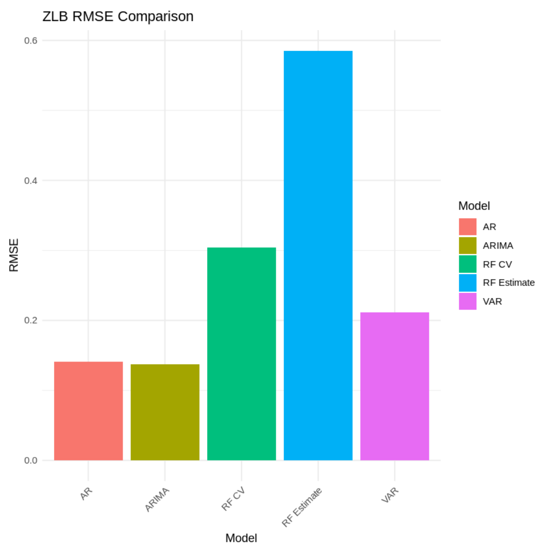


(d) BBK GDP Model

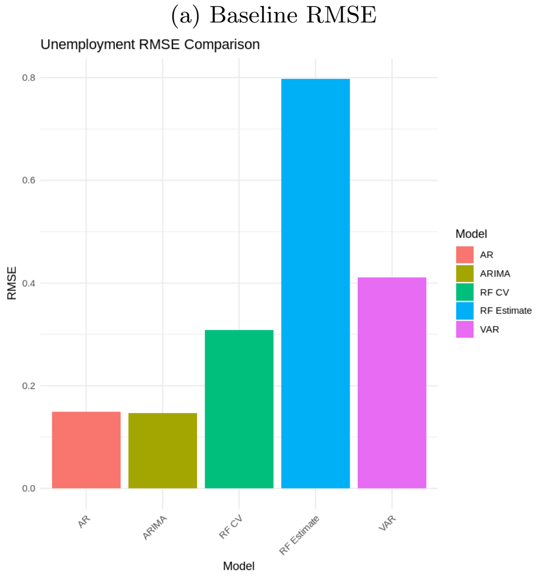
Figure 3: Actual values against predicted models



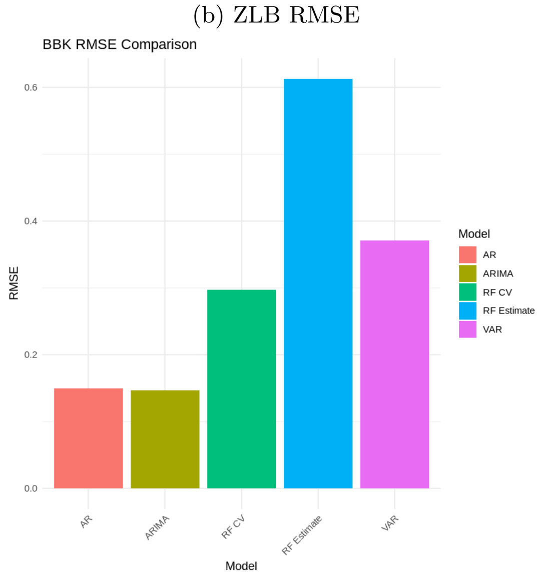
(a) Baseline RMSE



(b) ZLB RMSE



(c) Drop Unemployment RMSE



(d) BBK GDP RMSE

Figure 4: RMSE Comparison

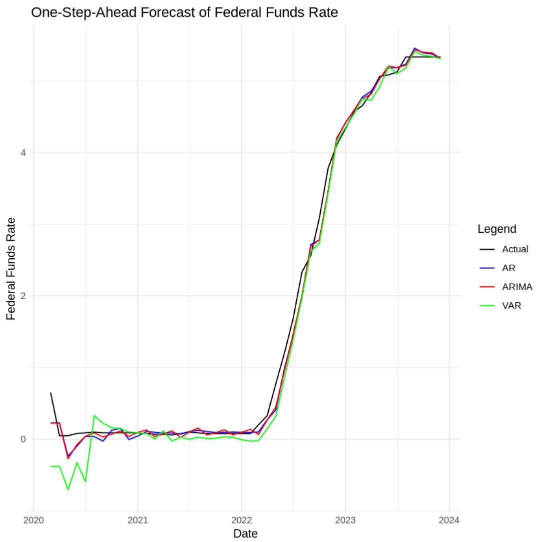


Figure 5: TS Models with BBK GDP Proxy and Without Unemployment: Actual vs Forecasts

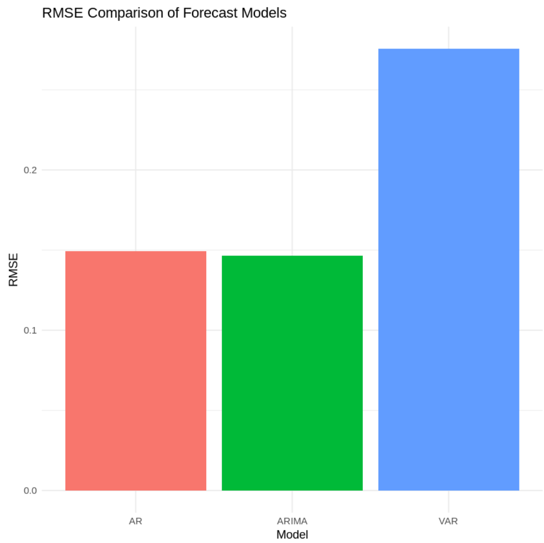


Figure 6: TS Models with BBK GDP Proxy and Without Unemployment: RMSE