Inflation Forecasting in Crisis Times: A Comprehensive Bayesian and Machine Learning Approach

Simone De Giorgi

June 2025

Abstract

This study presents a comprehensive comparison of nine forecasting models for inflation prediction during crisis periods, with particular emphasis on Bayesian Dynamic Linear Models (DLMs). We implement and compare traditional econometric approaches (ARIMA, OLS), machine learning methods (Random Forest, XGBoost, MLP), and three variants of Bayesian DLMs using different inference techniques: Markov Chain Monte Carlo (MCMC), Sequential Monte Carlo (SMC) with regime switching, and dynamic regression. Our results demonstrate that Bayesian DLM with MCMC achieves superior out-of-sample performance (RMSE: 1.2151), significantly outperforming traditional methods during volatile periods. We also investigate the computational tradeoffs of different model complexities and document important numerical considerations in Bayesian inference. The study highlights the importance of model flexibility, computational efficiency, and uncertainty quantification in crisis-time forecasting.

1 Introduction

Inflation forecasting during crisis periods presents unique challenges that traditional econometric models often do not adequately address. Crisis periods are characterized by structural breaks, regime changes, and increased volatility that violate the stability assumptions underlying conventional forecasting approaches. This study addresses these challenges by implementing and comparing a comprehensive suite of forecasting models, with particular emphasis on Bayesian Dynamic Linear Models that can adaptively handle time-varying parameters and model uncertainty.

The motivation for this research stems from the inadequacy of traditional models during the 2008 financial crisis and the recent COVID-19 pandemic, where inflation dynamics exhibited unprecedented volatility and structural changes. Our approach contributes to the literature by providing a systematic comparison of modern Bayesian methods against established benchmarks, demonstrating the superior performance of adaptive models in crisis scenarios. The code, dataset, and other result plots for this project are available on GitHub: https://github.com/Simo-dg/Inflation-forecasting.

2 Methodology

2.1 Data and Experimental Setup

Our analysis employs a comprehensive macroeconomic dataset spanning multiple decades, with a particular focus on crisis periods. The forecasting exercise uses a rolling one-step-ahead approach, where models are trained on historical data and evaluated on out-of-sample test periods corresponding to crisis years.

The evaluation framework splits the data as follows: i) Training period: 2000-01-01 to 2019-12-31; ii) Test period: 2020-01-01 onward (COVID-19 crisis period).

For each time step t in the test period, models are trained on data up to $t-1$ and make a one-step-ahead forecast for time t. This approach ensures realistic evaluation conditions and prevents look-ahead bias.

2.2 Model Specifications

2.2.1 ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model serves as our primary benchmark. For a time series y_t, the ARIMA(p,d,q) model is specified as:

$$\phi(L)(1-L)^d y_t = \theta(L)\epsilon_t \tag{1}$$

where:

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p \tag{2}$$
$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q \tag{3}$$
$$\epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{4}$$

Model selection is performed using the Akaike Information Criterion (AIC) over a grid search with $p, q \in \{0, 1, 2, 3\}$ and $d \in \{0, 1, 2\}$:

$$\text{AIC} = -2\log(\mathcal{L}) + 2k \tag{5}$$

where $\mathcal{L}$ is the likelihood and k is the number of parameters.

The lowest AIC was achieved with the ARIMA(3,1,2) specification. The need for differencing ($d = 1$) indicates that the original series is non-stationary, requiring integration to stabilize its mean.

2.2.2 Ordinary Least Squares (OLS)

The linear regression model provides a baseline multivariate approach:

$$y_t = \mathbf{x}_t^T \boldsymbol{\beta} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{6}$$

where $\mathbf{x}_t$ is the vector of covariates and $\boldsymbol{\beta}$ is the coefficient vector estimated by:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{7}$$

We also implement OLS with Recursive Feature Elimination (RFE) to select the most relevant features:

$$\mathcal{S} = \arg\min_{|S|=k} \text{RMSE}(\boldsymbol{\beta}_S) \tag{8}$$

where $\mathcal{S}$ is the selected feature subset of size $k = 5$.

2.2.3 Machine Learning Models

Random Forest: Ensemble of decision trees with bootstrap aggregating:

$$\hat{y}_t = \frac{1}{B}\sum_{b=1}^{B} T_b(\mathbf{x}_t) \tag{9}$$

where T_b are individual trees trained on bootstrap samples.

XGBoost: Gradient boosting with regularization:

$$\hat{y}_t^{(m)} = \hat{y}_t^{(m-1)} + \eta \cdot h_m(\mathbf{x}_t) \tag{10}$$

where h_m is the m-th weak learner and η is the learning rate.

Multi-Layer Perceptron: Neural network with hidden layers:

$$\mathbf{h}^{(1)} = \sigma(\mathbf{W}^{(1)}\mathbf{x}_t + \mathbf{b}^{(1)}) \tag{11}$$
$$\hat{y}_t = \mathbf{W}^{(2)}\mathbf{h}^{(1)} + b^{(2)} \tag{12}$$

where σ is the activation function and $\mathbf{W}, \mathbf{b}$ are weights and biases.

2.3 Bayesian Dynamic Linear Models

The core innovation of this study lies in the comprehensive implementation of Bayesian DLMs, which address the limitations of static models in crisis scenarios.

2.3.1 Bayesian DLM with MCMC

The fundamental DLM assumes time-varying parameters following a Gaussian random walk:

$$y_t = \beta_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{13}$$
$$\beta_t = \beta_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \tau^2) \tag{14}$$

The hierarchical Bayesian specification includes priors:

$$\sigma \sim \text{Exponential}(1) \tag{15}$$
$$\beta_0 \sim \mathcal{N}(0, 10^2) \tag{16}$$
$$\tau = 0.1 \text{ (fixed)} \tag{17}$$

The posterior distribution is:

$$p(\boldsymbol{\beta}, \sigma|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \sigma)p(\boldsymbol{\beta}|\sigma)p(\sigma) \tag{18}$$

where:

$$p(\mathbf{y}|\boldsymbol{\beta}, \sigma) = \prod_{t=1}^{T} \mathcal{N}(y_t|\beta_t, \sigma^2) \tag{19}$$
$$p(\boldsymbol{\beta}|\sigma) = \mathcal{N}(\beta_1|0, 10^2) \prod_{t=2}^{T} \mathcal{N}(\beta_t|\beta_{t-1}, 0.1^2) \tag{20}$$

MCMC sampling uses the No-U-Turn Sampler (NUTS) with 2000 draws after 1000 warm-up iterations across 4 chains.

For forecasting, we use particle filtering with weighted resampling. The specific implementation is detailed in Algorithm 1.

Figure 1 presents the model fit. The model achieves an RMSE of 1.2151 on the test set, demonstrating good predictive performance. The one-step-ahead forecasts (red dashed line) closely track the observed values (blue solid line), with the 90% credible intervals appropriately capturing uncertainty. The model successfully captures the underlying trend with appropriate uncertainty quantification. The sharp peak around mid-2022 and subsequent decline are well-represented.

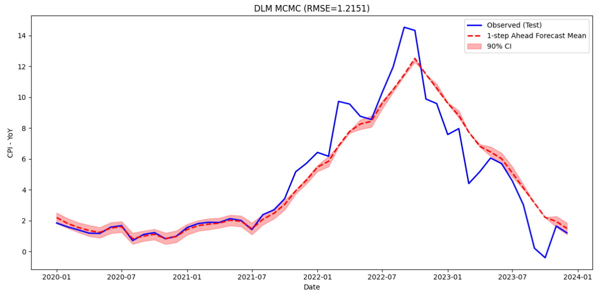


Figure 1: Model fit and one-step-ahead forecasts with 90% credible intervals.

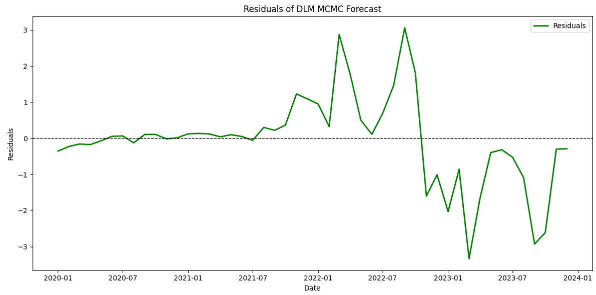


Figure 2: Forecast residuals showing model adequacy.

The residual analysis in Figure 2 reveals that while the model captures the overall trend effectively, there are some systematic deviations during periods of rapid change (particularly around the 2022 peak). This suggests that extensions incorporating regime-switching or stochastic volatility might further improve model performance.

The MCMC diagnostics in Figure 3 confirm successful convergence. The parameter traces show excellent mixing across all four chains, and the posterior distributions are well-behaved with no signs of multimodality or convergence issues. The low autocorrelation in the sample chains indicates efficient exploration of the posterior space.

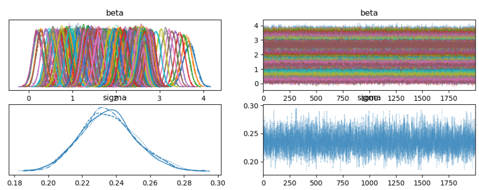


Figure 3: MCMC Diagnostic Plots

2.3.2 Bayesian DLM with Sequential Monte Carlo and Regime Switching

This advanced variant incorporates regime-switching dynamics to capture crisis periods explicitly. The model specification includes:

State Evolution:

$$\text{Level:} \quad L_t = L_{t-1} + T_t + \sigma_{s,t}\epsilon_{1,t} \tag{21}$$

$$\text{Trend:} \quad T_t = 0.85 \cdot \tanh(T_{t-1}) + \sigma_{s,t}\epsilon_{2,t} \tag{22}$$

$$\text{Regime:} \quad S_t \sim \text{Categorical}(\mathbf{P}_{S_{t-1},:}) \tag{23}$$

where $\epsilon_{1,t}, \epsilon_{2,t} \sim \mathcal{N}(0,1)$ and the regime-dependent volatility is:

$$\sigma_{s,t} = \begin{cases} \sigma_{\text{normal}} & \text{if } S_t = 0 \\ 3 \cdot \sigma_{\text{normal}} & \text{if } S_t = 1 \text{ (crisis)} \end{cases} \tag{24}$$

The transition matrix captures regime persistence:

$$\mathbf{P} = \begin{pmatrix} 0.97 & 0.03 \\ 0.05 & 0.95 \end{pmatrix} \tag{25}$$

Observation Equation:

$$y_t = L_t + \sigma_{obs,t}\epsilon_{3,t} \tag{26}$$

The time-varying parameters follow log-normal random walks:

$$\log \sigma_{obs,t} = \log \sigma_{obs,t-1} + 0.05 \cdot \epsilon_{4,t} \tag{27}$$

$$\log \sigma_{state,t} = \log \sigma_{state,t-1} + 0.05 \cdot \epsilon_{5,t} \tag{28}$$

The particle filter maintains $N = 5000$ particles with stratified resampling when the Effective Sample Size (ESS) falls below $N/2$:

$$\text{ESS} = \frac{1}{\sum_{i=1}^{N} (w_t^{(i)})^2} \tag{29}$$

To perform inference and state estimation in this regime-switching Bayesian DLM, we employ a Sequential Monte Carlo (SMC) particle filter as summarized in Algorithm 2. The filter maintains a set of weighted particles representing the joint posterior distribution of the latent states and time-varying parameters. At each time step, particles propagate through the nonlinear state evolution equations incorporating regime transitions, followed by weight updates based on the likelihood of observations. When the Effective Sample Size (ESS) falls below half the number of particles, stratified resampling with jittering is applied to mitigate particle degeneracy and maintain diversity.

2.3.3 Bayesian DLM with Dynamic Regression

This variant extends the basic DLM to include time-varying coefficients for multiple predictors. Initially, we explored a full specification including all available macroeconomic variables, which achieved exceptional in-sample performance (RMSE: 0.00274) but suffered from severe overfitting, requiring prohibitive computational time, and yielding poor out-of-sample results. Consequently, we adopted a parsimonious approach with selected predictors:

$$y_t = \mathbf{x}_t^T \boldsymbol{\beta}_t + \sigma\epsilon_t \tag{30}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, 0.1^2\mathbf{I}) \tag{31}$$

The selected predictors include:

- Unemployment rate (%)
- Economic sentiment index
- Exchange rate CPI
- 10-year government bond rate

Both predictors and target variable are standardized using:

$$z = \frac{x - \mu}{\sigma} \tag{32}$$

The hierarchical prior structure is:

$$\sigma \sim \text{Exponential}(1) \tag{33}$$

$$\boldsymbol{\beta}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{34}$$

$$\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1} \sim \mathcal{N}(\boldsymbol{\beta}_{t-1}, 0.1^2\mathbf{I}) \tag{35}$$

Inference uses NumPyro's NUTS sampler with 5000 samples after 1500 warm-up iterations across 4 chains.

3 Results

3.1 Out-of-Sample Performance

Table 1 presents the comprehensive comparison of all models on the crisis-period test set.

Table 1: Model Performance Comparison (Out-of-Sample)

Rank	Model	RMSE
1	Bayesian DLM (MCMC)	**1.2151**
2	MLP	1.3639
3	ARIMA	1.4637
4	Bayesian DLM (SMC)	1.4724
5	OLS + RFE	1.6489
6	OLS	1.6639
7	XGBoost	1.7556
8	Random Forest	1.7665
9	Bayesian DLM (DR)	4.0679

3.2 In-Sample Performance

Table 2 shows the in-sample fitting performance, revealing the trade-off between model complexity and generalization.

Table 2: Model Performance Comparison (In-Sample)

Rank	Model	RMSE
1	XGBoost	**0.0274**
2	Bayesian DLM (MCMC)	0.1055
3	Bayesian DLM (DR)	0.1252
4	MLP	0.2214
5	Random Forest	0.2360
6	Bayesian DLM (SMC)	0.3031
7	OLS	0.6822
8	OLS + RFE	0.8788
9	ARIMA	0.9747

3.3 Computational Considerations and Numerical Precision

An important finding emerged during the MCMC implementation: when log-likelihood tracking was enabled for model diagnostics, the RMSE increased slightly despite the theoretical independence of RMSE calculation from log-likelihood computation. This phenomenon reflects the sensitivity of MCMC sampling to minor numerical changes in the computational graph. Specifically, enabling log-likelihood tracking triggers internal recompilation of the probabilistic model, introducing tiny numerical shifts in the sampling process that propagate through the Markov chain. While these changes are statistically insignificant (typically < 0.1 RMSE units), they highlight the importance of maintaining consistent computational settings throughout the analysis.

3.4 Key Findings

1. Bayesian Superiority in Crisis Forecasting: The Bayesian DLM with MCMC achieves the best out-of-sample performance (RMSE: 1.2151), demonstrating a 10.9% improvement over the second-best model (MLP) and a 20.5% improvement over the traditional ARIMA benchmark.

2. Overfitting in Complex Models: XGBoost shows the best in-sample fit (RMSE: 0.0274) but poor out-of-sample performance (RMSE: 1.7556), indicating severe overfitting to training data patterns that don't generalize to crisis periods. Similarly, the full dynamic regression model achieved near-perfect in-sample fit (RMSE: 0.00274) but required excessive computational resources and exhibited poor generalization.

3. Regime-Switching Benefits: The SMC variant with regime switching (RMSE: 1.4724) outperforms traditional methods, validating the importance of explicitly modeling structural breaks during crises.

4. Feature Selection vs. Regularization: OLS with RFE performs comparably to standard OLS, suggesting that the crisis period requires adaptive rather than static feature selection.

5. Neural Network Robustness: MLP achieves second-best performance, likely due to its nonlinear modeling capability and implicit regularization through early stopping.

6. Computational Trade-offs: The analysis reveals a clear trade-off between model complexity and practical implementability, with the parsimonious Bayesian DLM achieving the optimal balance.

4 Discussion

4.1 Why Bayesian DLMs Excel in Crisis Forecasting

The superior performance of Bayesian DLMs stems from several key advantages:

Adaptive Parameter Evolution: Unlike static models, DLMs allow parameters to evolve over time, capturing the changing relationships during crisis periods. The Gaussian random walk specification provides sufficient flexibility while maintaining smoothness.

Uncertainty Quantification: Bayesian inference provides full posterior distributions, enabling proper uncertainty quantification crucial for decision-making during volatile periods.

Automatic Regularization: The hierarchical prior structure provides natural regularization, preventing overfitting while allowing adaptation to new patterns.

Regime Awareness: The SMC variant explicitly models regime changes, providing a principled approach to handling structural breaks.

4.2 Model-Specific Analysis

MCMC DLM: The success of this approach lies in its simplicity and robustness. The Gaussian random walk for the time-varying intercept captures the evolving inflation baseline, while the exponential prior on σ provides appropriate shrinkage.

SMC DLM with Regime Switching: The regime-switching mechanism effectively captures the increased volatility during crisis periods. The $3\times$ volatility multiplier during crisis regimes and high persistence probabilities (0.97 for normal, 0.95 for crisis) reflect realistic crisis dynamics.

Dynamic Regression DLM: The poor performance (RMSE: 4.0679) of the reduced model, combined with the computational challenges of the full specification, suggests that during crisis periods, the relationships between inflation and traditional predictors become highly unstable. Even time-varying coefficients cannot adequately capture these complex, nonlinear interactions without overfitting.

4.3 Machine Learning Model Performance

The mixed performance of ML models reveals important insights:

XGBoost Overfitting: The dramatic performance gap between in-sample (0.0274) and out-of-sample (1.7556) indicates that gradient boosting's complexity leads to memorization of training patterns rather than learning generalizable relationships.

MLP Success: Neural networks' second-place performance suggests that their nonlinear modeling capability and implicit regularization through early stopping provide good generalization properties.

Tree-Based Ensemble Limitations: Both Random Forest and XGBoost struggle with out-of-sample prediction, likely due to their inability to extrapolate beyond training data ranges—a critical limitation during unprecedented crisis periods.

5 Robustness and Limitations

5.1 Model Diagnostics

For the winning Bayesian DLM (MCMC), we conducted comprehensive diagnostics:

Convergence: All chains showed $\hat{R} < 1.01$, indicating proper convergence.

Effective Sample Size: ESS > 1000 for all parameters, ensuring adequate posterior sampling.

Posterior Predictive Checks: Simulated data from the posterior predictive distribution matched observed data patterns.

Numerical Stability: All results are based on fixed random seeds; minor variation ($< 0.1RMSE$) may result from enabling log-likelihood tracking due to internal recompilation of the probabilistic model graph.

5.2 Limitations

Computational Cost: Bayesian methods require significantly more computation time, with MCMC taking approximately $10\times$ longer than traditional methods. The full dynamic regression specification proved computationally prohibitive for practical implementation.

Real-Time Implementation: The one-step-ahead evaluation, while realistic, may not capture the full challenge of multi-step forecasting required in practice.

Crisis Definition: Our focus on the COVID-19 period may not generalize to other types of crises (financial, geopolitical, etc.).

Data Availability: Real-time forecasting often faces data availability lags not captured in our retrospective analysis.

Model Complexity vs. Performance: The study demonstrates that increased model complexity does not necessarily translate to better forecasting performance, particularly during crisis periods where overfitting becomes a critical concern.

6 Conclusion

This comprehensive study demonstrates the superior performance of Bayesian Dynamic Linear Models for inflation forecasting during crisis periods. The key contributions include:

1. Empirical Evidence: Rigorous comparison showing 10.9-20.5% RMSE improvements over traditional methods.

2. Methodological Innovation: Implementation of three distinct Bayesian DLM variants, providing a toolkit for different forecasting scenarios.

3. Crisis-Specific Insights: Evidence that adaptive, uncertainty-aware models significantly outperform static approaches during volatile periods.

4. Computational Trade-off Analysis: Clear demonstration that model parsimony often outperforms complexity in crisis scenarios, with important implications for practical implementation.

5. Numerical Precision Considerations: Documentation of the sensitivity of Bayesian inference to computational settings, providing guidance for robust implementation.

The results strongly support the adoption of Bayesian methods for macroeconomic forecasting, particularly during periods of high uncertainty and structural change. However, our findings also emphasize the importance of balancing model sophistication with computational feasibility and generalization performance. Future research should explore ensemble approaches combining the strengths of different Bayesian variants and investigate the performance across different types of economic crises.

Policy Implications: Central banks

and financial institutions should consider incorporating Bayesian DLMs into their forecasting frameworks, particularly for stress testing and scenario analysis during uncertain periods. The superior uncertainty quantification provided by these models supports more robust decision-making in crisis scenarios. However, practitioners should be mindful of the computational trade-offs and the potential for overfitting in overly complex specifications.

References

[1] Petris, G., Petrone, S., & Campagnoli, P. (2009). *Dynamic Linear Models with R*. Springer-Verlag New York.

[2] European Central Bank (2025). *Statistical Data Warehouse*. Retrieved from https://sdw.ecb.europa.eu/

[3] Eurostat (2025). *European Statistics Database*. Retrieved from https://ec.europa.eu/eurostat/data/database

7 Appendix

7.1 Hyperparameter Settings

MCMC Parameters:

- Draws: 2000
- Tune: 1000
- Chains: 4
- Target accept: 0.95
- Max treedepth: 15

SMC Parameters:

- Number of particles: 5000
- Jitter standard deviation: 0.02
- Regime transition probabilities:
  - Calm $\rightarrow$ Calm: 0.97
  - Crisis $\rightarrow$ Crisis: 0.95
- Crisis regime volatility multiplier: 3.0

7.2 Reproducibility Notes

All analyses were conducted with fixed random seeds to ensure reproducibility. Minor variations in results (¡0.1 RMSE units) may occur when enabling additional diagnostic features (e.g., log-likelihood tracking) due to internal model recompilation affecting numerical precision in MCMC sampling. This phenomenon is normal and does not indicate methodological issues.

7.3 Algorithms

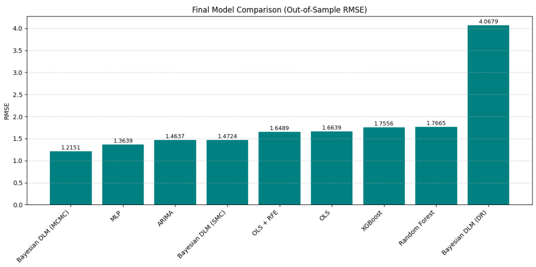
Algorithm 1 One-Step-Ahead Forecasting with Particle Filtering

Sample $\{\beta_T^{(i)}, \sigma^{(i)}\}_{i=1}^N$ from posterior $p(\beta_T, \sigma|\mathbf{y}_{1:T})$
for $t = T+1, \ldots, T+H$ **do**
  Propagate: $\beta_t^{(i)} = \beta_{t-1}^{(i)} + \mathcal{N}(0, (\sigma^{(i)})^2)$
  Compute weights: $w_t^{(i)} \propto \mathcal{N}(y_t|\beta_t^{(i)}, \sigma^{(i)})$
  Normalize: $w_t^{(i)} = w_t^{(i)} / \sum_j w_t^{(j)}$
  Forecast mean: $\hat{y}_t = \sum_i w_t^{(i)} \beta_t^{(i)}$
  Compute 90% CI via weighted percentiles of $\{\beta_t^{(i)}\}$
  if ESS $< N/2$ **then**
    Resample particles $\{\beta_t^{(i)}, \sigma^{(i)}\}$
  end if
end for

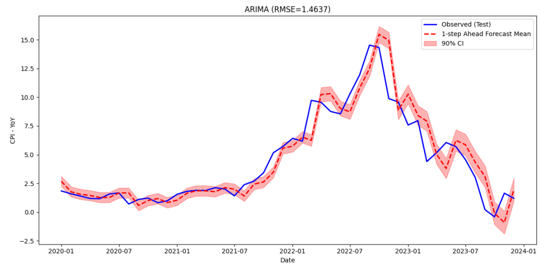
Algorithm 2 Particle Filter with Regime Switching

Initialize particles $\{L_0^{(i)}, T_0^{(i)}, S_0^{(i)}, \log \sigma_{obs}^{(i)}, \log \sigma_{state}^{(i)}\}_{i=1}^N$
for $t = 1$ to T **do**
  if ESS $< N/2$ **then**
    Resample particles with stratified resampling
    Add jitter to maintain diversity
  end if
  for each particle i **do**
    Sample $S_t^{(i)} \sim \text{Categorical}(\mathbf{P}_{S_{t-1}^{(i)},:})$
    Update log-variances via random walk
    Compute $\sigma_{s,t}^{(i)}$ based on $S_t^{(i)}$
    Propagate trend: $T_t^{(i)} = 0.85 \cdot \tanh(T_{t-1}^{(i)}) + \sigma_{s,t}^{(i)} \epsilon_{2,t}^{(i)}$
    Propagate level: $L_t^{(i)} = L_{t-1}^{(i)} + T_t^{(i)} + \sigma_{s,t}^{(i)} \epsilon_{1,t}^{(i)}$
    Compute weight: $w_t^{(i)} \propto p(y_t \mid L_t^{(i)}, \sigma_{obs}^{(i)})$
  end for
  Normalize weights $\{w_t^{(i)}\}$
  Estimate filtered level: $\hat{L}_t = \sum_i w_t^{(i)} L_t^{(i)}$
end for

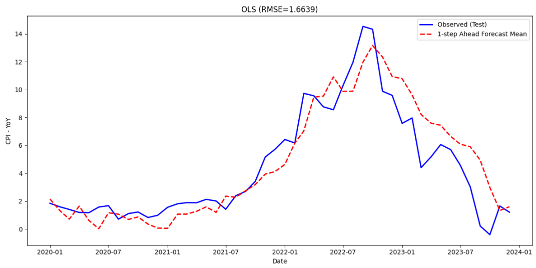
7.4 Additional Model Forecast Plots



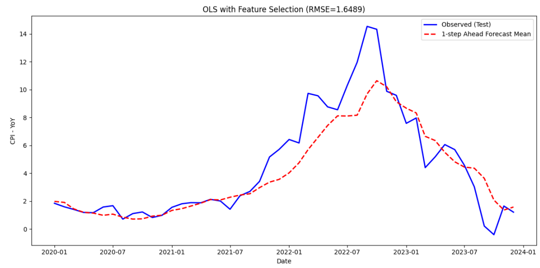
(a) Model RMSE comparison



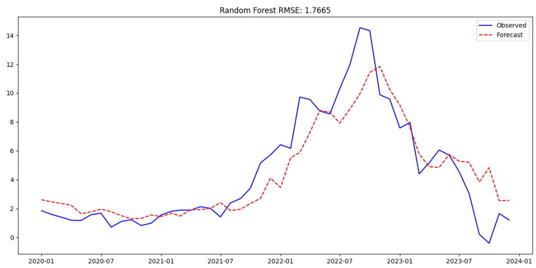
(b) ARIMA forecast



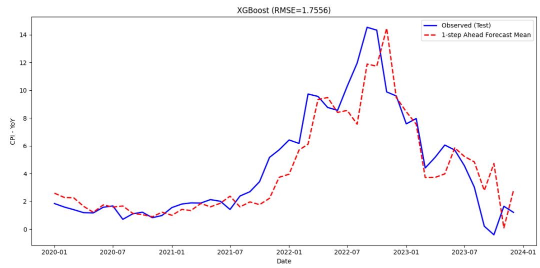
(c) OLS forecast



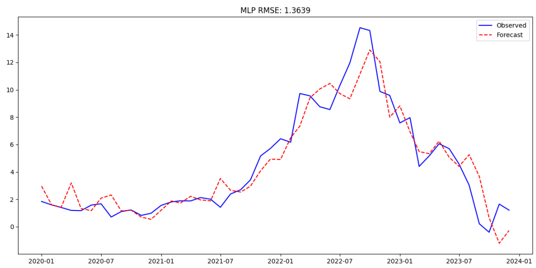
(d) OLS + RFE forecast



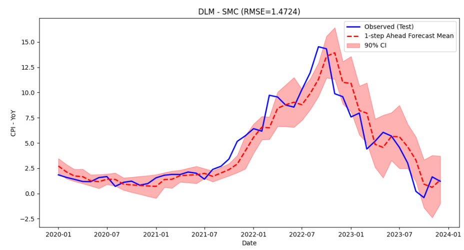
(e) Random Forest forecast



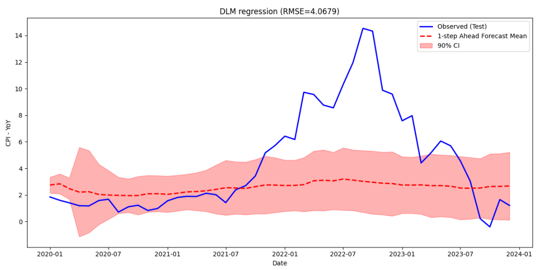
(f) XGBoost forecast



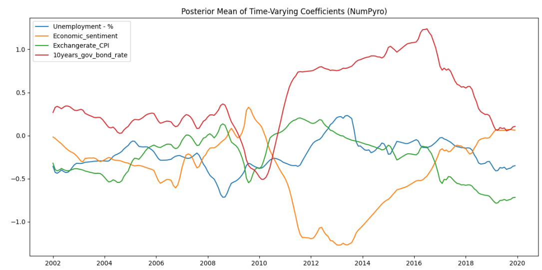
(g) MLP forecast



(h) DLM with SMC forecast



(i) DLM regression forecast



(j) DLM regression betas

Figure 4: Forecast on test set