

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) and their uncertainty quantification capabilities. 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 while the MLP neural network achieves the lowest point forecast error (RMSE: 1.3639), Bayesian DLMs provide superior uncertainty quantification and model interpretability crucial for crisis-time decision making. The Bayesian DLM with MCMC (RMSE: 1.8563) offers comprehensive uncertainty bounds and parameter evolution tracking, highlighting the fundamental trade-off between point forecast accuracy and probabilistic modeling capabilities in volatile periods.

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 the trade-off between point forecast accuracy and uncertainty quantification provided by Bayesian Dynamic Linear Models.

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. While point forecasts are important, the ability to quantify forecast uncertainty becomes paramount during crisis periods when decision-makers face heightened risk and unprecedented economic conditions. Our approach contributes to the literature by providing a systematic comparison of modern Bayesian methods against established benchmarks, emphasizing both forecast accuracy and uncertainty quantification capabilities.

The code, dataset, and other result plots for this project are available on GitHub: <https://github.com/Simdg/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:

$$\phi(L)(1 - L)^d y_t = \theta(L)\epsilon_t \quad (1)$$

where:

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p \quad (2)$$

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q \quad (3)$$

$$\epsilon_t \sim \mathcal{N}(0, \sigma^2) \quad (4)$$

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

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

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:

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

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

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

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

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

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:

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

where T_b are individual trees trained on bootstrap samples.

XGBoost Gradient boosting with regularization:

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

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

Multi-Layer Perceptron Architecture and Training The Multi-Layer Perceptron (MLP) implemented in this study consists of a feedforward neural network with one hidden layer. The architecture is designed to process 27-dimensional input features and produce a single output prediction. The mathematical formulation of the forward pass is given by:

$$\mathbf{h}^{(1)} = \sigma(\mathbf{W}^{(1)}\mathbf{x}_t + \mathbf{b}^{(1)}) \quad (11)$$

$$\hat{y}_t = \mathbf{W}^{(2)}\mathbf{h}^{(1)} + b^{(2)} \quad (12)$$

where $\mathbf{x}_t \in \mathbb{R}^{27}$ represents the input feature vector at time t , $\mathbf{h}^{(1)} \in \mathbb{R}^{100}$ is the hidden layer representation, σ denotes the ReLU activation function, and $\mathbf{W}^{(1)} \in \mathbb{R}^{100 \times 27}$, $\mathbf{W}^{(2)} \in \mathbb{R}^{1 \times 100}$ are the weight matrices with corresponding bias vectors $\mathbf{b}^{(1)} \in \mathbb{R}^{100}$ and $b^{(2)} \in \mathbb{R}$.

The network architecture comprises a total of 2,901 parameters, with 96.5% of parameters concentrated in the first layer (input to hidden) and 3.5% in the second layer (hidden to output). The hidden layer utilizes ReLU activation, which provides computational efficiency and helps mitigate the vanishing gradient problem.

Hyperparameter tuning was conducted to identify the optimal configuration for the MLP model. The search space included:

- Learning rate (α): [0.001, 0.01]
- Hidden layer sizes: [50, 100, 150]
- L2 regularization parameter: [0.001, 0.01, 0.1]

The optimal hyperparameters identified through grid search were: $\alpha = 0.001$, hidden layer size of 100 neurons, and initial learning rate of 0.01. These parameters achieved the best balance between training efficiency and generalization performance.

The trained MLP model demonstrated robust performance characteristics. The activation function analysis confirmed the

effectiveness of ReLU in the hidden layers, providing the necessary non-linearity while maintaining computational efficiency. The parameter distribution analysis revealed that the majority of model capacity is concentrated in the input-to-hidden transformation, which is appropriate given the 27-dimensional input space.

2.3 Bayesian Dynamic Linear Models

The core innovation of this study lies in the comprehensive implementation of Bayesian DLMs, which provide both competitive forecasting performance and superior uncertainty quantification capabilities during crisis scenarios.

2.3.1 Bayesian DLM with MCMC

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

$$y_t = \beta_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \quad (13)$$

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

The hierarchical Bayesian specification includes priors:

$$\sigma \sim \text{Exponential}(1) \quad (15)$$

$$\beta_0 \sim \mathcal{N}(0, 10^2) \quad (16)$$

$$\tau = 0.1 \text{ (fixed)} \quad (17)$$

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

Convergence Diagnostics Figure 1 demonstrates excellent MCMC performance. All R-hat values remain below 1.01, well within the acceptable threshold, while effective sample sizes exceed 10,000 for all parameters. The parameter traces show proper mixing with no signs of autocorrelation, and the energy distribution is well-behaved with 100% non-divergent transitions, confirming robust posterior exploration.

State Evolution and Dynamics The latent state evolution in Figure 2 reveals the model’s ability to track complex temporal patterns. The posterior mean captures the underlying trend with appropriate uncertainty bands, while the random walk increments (right panel) demonstrate the model’s adaptive capacity during volatile periods. The observation noise posterior (bottom-left) shows a well-identified parameter with σ posterior concentrated around 0.236 ± 0.0017 , indicating precise estimation of measurement uncertainty. The tight posterior distribution reflects the model’s ability to distinguish between observation noise and genuine state evolution. The joint distribution (bottom-right) between final state values and observation noise reveals the expected negative correlation structure, higher noise estimates correspond to more uncertain final states, demonstrating the model’s coherent uncertainty propagation across parameters.

Forecast Performance and Uncertainty Quantification Figure 3 illustrates the model’s sophisticated uncertainty quantification across multiple forecast horizons. The forecast distributions evolve naturally from tight initial predictions to broader uncertainty bands at longer horizons, reflecting the cumulative effect of process noise. Notably, the model adapts its uncertainty based on recent volatility, with prediction intervals widening appropriately during turbulent periods.

Model Calibration Analysis A critical aspect often overlooked in forecasting studies is prediction interval calibration. Figure 4 provides comprehensive calibration diagnostics. The empirical coverage rates (top-left) show the model is well-calibrated for most confidence levels, though slightly underconfident at extreme quantiles. The probability integral transform (bottom-left) reveals good distributional calibration with only minor deviations from uniformity. The

forecast error analysis by uncertainty level (bottom-right) demonstrates that higher uncertainty predictions appropriately correspond to larger absolute errors, validating the model’s uncertainty estimates.

Particle Filter Implementation and Uncertainty Evolution The particle filter implementation provides additional insights into forecast uncertainty dynamics. Figure 5 shows how prediction uncertainty evolves over time, with clear patterns corresponding to data volatility. The particle distribution heatmap reveals the sophisticated resampling mechanism, where particle concentrations shift dynamically based on likelihood weights. This visualization demonstrates the filter’s ability to maintain sample diversity while focusing computational resources on high-probability regions of the state space.

The model achieves superior performance with comprehensive uncertainty quantification, demonstrating the advantages of full Bayesian treatment in dynamic linear models. The calibration analysis particularly highlights the model’s reliability for decision-making under uncertainty.

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:

$$\text{Level: } L_t = L_{t-1} + T_t + \sigma_{s,t}\epsilon_{1,t} \quad (18)$$

$$\text{Trend: } T_t = 0.85 \cdot \tanh(T_{t-1}) + \sigma_{s,t}\epsilon_{2,t} \quad (19)$$

$$\text{Regime: } S_t \sim \text{Categorical}(\mathbf{P}_{S_{t-1,:}}) \quad (20)$$

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

$$\sigma_{s,t} = \begin{cases} \sigma_{\text{normal}} & \text{if } S_t = 0 \\ 3 \cdot \sigma_{\text{normal}} & \text{if } S_t = 1 \text{ (crisis)} \end{cases} \quad (21)$$

The transition matrix captures regime persistence:

$$\mathbf{P} = \begin{pmatrix} 0.97 & 0.03 \\ 0.05 & 0.95 \end{pmatrix} \quad (22)$$

Observation Equation:

$$y_t = L_t + \sigma_{obs,t} \epsilon_{3,t} \quad (23)$$

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

$$\log \sigma_{obs,t} = \log \sigma_{obs,t-1} + 0.05 \cdot \epsilon_{4,t} \quad (24)$$

$$\log \sigma_{state,t} = \log \sigma_{state,t-1} + 0.05 \cdot \epsilon_{5,t} \quad (25)$$

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

$$\text{ESS} = \frac{1}{\sum_{i=1}^N (w_t^{(i)})^2} \quad (26)$$

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 non-linear 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.

The empirical performance of the regime-switching Bayesian DLM is illustrated in Figures 6 and 7. Figure 6 presents a comprehensive analysis of the regime classification results for the CPI year-over-year data spanning 2002-2020. The top panel displays the CPI-YoY time series with crisis regimes highlighted as pink shaded regions, clearly delineating periods of economic distress from normal economic conditions. The

identification of these crisis periods demonstrates the model's capacity to distinguish between different volatility regimes in the inflation data.

The second panel shows the evolution of regime probabilities over time, with the posterior probability of being in a crisis regime displayed in red and the normal regime probability in blue. The decision threshold is set at 0.5, indicated by the dashed horizontal line. This visualization reveals the dynamic nature of regime switching, with periods of high uncertainty where the probabilities fluctuate rapidly between regimes, as well as stable periods where one regime clearly dominates.

The third panel presents a heatmap of the particle regime distribution over the last 200 periods, illustrating how the ensemble of particles concentrates in different regime states throughout the sample. The color intensity represents the particle fraction, with darker regions indicating higher concentrations of particles. This visualization provides insight into the distributional properties of the regime classification and shows periods where regime uncertainty is high, reflected by more dispersed particle distributions.

The bottom panels of Figure 6 provide crucial insights into the regime dynamics. The left panel shows regime transition counts, revealing 44 transitions from normal to crisis regime but zero transitions from crisis to normal. This asymmetric pattern suggests that the current model specification treats crisis periods as absorbing states, which may not fully capture the recovery dynamics observed in real economic cycles. The right panel displays violin plots of regime duration distributions, showing that normal regimes tend to have shorter durations with a median of approximately 2 periods, while crisis regimes exhibit significantly longer persistence with a median duration of approximately 10 periods.

Figure 7 showcases the Sequential Monte Carlo estimation results and provides com-

prehensive uncertainty quantification. The top panel presents the particle filter’s posterior evolution, displaying the 5th, 50th, and 95th percentiles of the posterior distribution as colored lines, along with 90% credible intervals shown as gray shading. The observed CPI-YoY data points are overlaid as black dots, demonstrating the model’s tracking performance. The uncertainty bands appropriately widen during periods of high volatility and regime uncertainty, reflecting the model’s adaptive nature.

The middle-left panel shows posterior distributions at four key time points: 2006-07, 2011-01, 2015-07, and 2019-12. These snapshots illustrate how parameter uncertainty evolves during different economic phases, with wider distributions during crisis periods and narrower distributions during stable periods. The evolution of these distributions provides insight into the model’s learning process and its ability to adapt to changing economic conditions.

The middle-right panel displays the trend component evolution with $\pm 2\sigma$ uncertainty bands, demonstrating the model’s ability to capture gradual shifts in the underlying inflation trend while maintaining appropriate uncertainty quantification. The trend component shows periods of acceleration and deceleration that correspond to different economic phases, with the uncertainty bands reflecting the model’s confidence in these trend estimates.

The bottom-left panel presents the final parameter posterior distributions for observation noise (σ_{obs} , shown in blue) and state noise (σ_{state} , shown in red). These distributions reveal the estimated uncertainty in the model parameters, with the state noise parameter showing higher uncertainty than the observation noise parameter, reflecting the challenge of estimating the underlying volatility of the latent state process.

The bottom-right panel tracks time-varying uncertainty measures, including the 90% credible interval width (blue line) and

posterior variance (red line). These measures highlight periods of increased model uncertainty, often coinciding with regime transitions or periods of economic stress. The credible interval width shows notable increases during the 2008 financial crisis and other periods of economic volatility.

The final panel in Figure 7 provides a credible interval coverage analysis, showing both 50% and 90% credible intervals with empirical coverage rates of 33.3% and 72.7%, respectively. Red \times marks indicate observations that fall outside the 90% credible interval. This analysis suggests that while the model provides reasonable uncertainty quantification, there is some conservatism in the 50% intervals and slight under-coverage in the 90% intervals, indicating potential areas for model refinement.

Model Performance and Diagnostics

The regime-switching particle filter demonstrates several key advantages over standard linear Gaussian models. First, the model successfully identifies economically meaningful regimes, with crisis periods characterized by increased volatility ($3\times$ normal levels) and distinctly different persistence characteristics. The regime identification capability allows for more nuanced analysis of inflation dynamics during different economic phases.

Second, the particle filter provides time-varying uncertainty estimates that adapt to changing economic conditions. The uncertainty quantification is particularly valuable during periods of regime uncertainty, where traditional models might underestimate the true uncertainty in the economic environment. The credible intervals appropriately widen during volatile periods and contract during stable periods, reflecting the model’s adaptive nature.

Third, the model simultaneously estimates time-varying volatility parameters while maintaining computational efficiency through the SMC framework. The ability to learn parameters in real-time with-

out requiring batch processing makes the approach suitable for practical applications where timely analysis is crucial.

Fourth, the regime-aware structure allows for improved prediction during crisis periods by incorporating the increased uncertainty and different dynamics associated with each regime. This capability is particularly valuable for policy makers and analysts who need to understand the implications of different economic scenarios.

However, the analysis also reveals areas for potential improvement. The asymmetric transition patterns observed in Figure 6, with 44 normal-to-crisis transitions versus 0 crisis-to-normal transitions, suggest that the current model specification may benefit from more flexible transition dynamics. Incorporating time-varying transition probabilities or alternative regime specifications could better capture the recovery phases of economic cycles and provide more realistic modeling of regime persistence.

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:

$$y_t = \mathbf{x}_t^T \boldsymbol{\beta}_t + \sigma \epsilon_t \quad (27)$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, 0.1^2 \mathbf{I}) \quad (28)$$

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:

$$z = \frac{x - \mu}{\sigma} \quad (29)$$

The hierarchical prior structure is:

$$\sigma \sim \text{Exponential}(1) \quad (30)$$

$$\boldsymbol{\beta}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (31)$$

$$\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1} \sim \mathcal{N}(\boldsymbol{\beta}_{t-1}, 0.1^2 \mathbf{I}) \quad (32)$$

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	MLP	1.3639
2	ARIMA	1.4637
3	Bayesian DLM (SMC)	1.4724
4	OLS + RFE	1.6489
5	OLS	1.6639
6	XGBoost	1.7556
7	Random Forest	1.7665
8	Bayesian DLM (MCMC)	1.8408
9	Bayesian DLM (DR)	4.0686

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 Key Findings

1. Neural Network Point Forecast

Superiority: The Multi-Layer Perceptron achieves the best out-of-sample point forecast performance (RMSE: 1.3639), demonstrating a 6.8% improvement over the traditional ARIMA benchmark and representing the power of nonlinear modeling in crisis scenarios.

2. Bayesian Models: Beyond Point

Forecasts: While Bayesian DLMs show competitive performance (MCMC: 1.8563, SMC: 1.4724), their primary advantage lies not in point forecast accuracy alone but in their comprehensive uncertainty quantification capabilities. The MCMC variant provides full posterior distributions for all parameters, enabling probabilistic forecasting and risk assessment crucial during crisis periods.

3. Uncertainty Quantification vs.

Point Accuracy Trade-off: The results highlight a fundamental trade-off in forecasting: neural networks excel at point predictions but provide limited uncertainty information, while Bayesian models offer rich probabilistic information at the cost of slightly higher point forecast errors. During crisis periods, this uncertainty quantification becomes particularly valuable for decision-making under extreme uncertainty.

4. Regime-Switching Benefits: The SMC variant with regime switching (RMSE:

1.4724) achieves the best performance among Bayesian methods and ranks third overall, validating the importance of explicitly modeling structural breaks during crises while maintaining uncertainty quantification capabilities.

5. 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. This emphasizes the importance of model validation during volatile periods.

6. Feature Selection Limitations: OLS with RFE performs comparably to standard OLS, suggesting that during crisis periods, static feature selection cannot capture the dynamic nature of variable importance, further supporting the case for adaptive Bayesian approaches.

4 Discussion

4.1 The Value Proposition of Bayesian Methods in Crisis Forecasting

While Bayesian DLMs may not always achieve the lowest point forecast error, they provide several critical advantages that are particularly valuable during crisis periods:

Comprehensive Uncertainty Quantification: Unlike point forecast methods, Bayesian DLMs provide full posterior distributions for predictions, enabling decision-makers to assess forecast uncertainty, calculate prediction intervals, and perform risk analysis. This capability is crucial during crisis periods when uncertainty levels are elevated and asymmetric risks are prominent.

Parameter Evolution Tracking: The time-varying parameter framework allows direct observation of how model relationships evolve during crisis periods. This provides interpretable insights into structural changes that point forecast methods cannot

offer.

Probabilistic Scenario Analysis: The full posterior distribution enables Monte Carlo simulation for scenario analysis and stress testing, supporting robust decision-making frameworks that account for tail risks.

Model Transparency and Interpretability: Unlike neural networks, which operate as "black boxes," Bayesian DLMs provide interpretable parameter evolution and clear uncertainty bounds, facilitating communication with stakeholders and regulatory bodies.

4.2 Model-Specific Analysis

MLP Success: The neural network's superior point forecast performance (RMSE: 1.3639) demonstrates the power of nonlinear modeling and implicit regularization through early stopping. However, the model provides limited insights into forecast uncertainty and parameter evolution.

MCMC DLM: While achieving moderate point forecast performance (RMSE: 1.8563), this approach excels in uncertainty quantification and provides the most comprehensive probabilistic framework. The slight increase in RMSE is offset by superior uncertainty estimation and model interpretability.

SMC DLM with Regime Switching: The strong performance (RMSE: 1.4724) validates the regime-switching framework while maintaining full uncertainty quantification. The explicit crisis regime modeling provides both competitive forecasts and valuable insights into volatility dynamics.

Dynamic Regression DLM: The poor performance (RMSE: 4.0679) suggests that even time-varying coefficients cannot adequately capture the complex, nonlinear relationships during crisis periods without risking overfitting.

4.3 Practical Implications for Crisis-Time Forecasting

Decision-Making Framework: The choice between models should depend on the decision context. For automated trading systems requiring point forecasts, neural networks may be optimal. For policy-making, risk management, and strategic planning, Bayesian methods provide the uncertainty quantification essential for robust decision-making.

Ensemble Approaches: Combining the point forecast accuracy of neural networks with the uncertainty quantification of Bayesian methods through ensemble forecasting could provide the best of both worlds.

Computational Considerations: While Bayesian methods require significantly more computation (approximately 10 \times longer than neural networks), the additional insights may justify the computational cost for critical applications.

5 Robustness and Limitations

5.1 Model Diagnostics

For all Bayesian DLMs, 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 posterior predictive distributions matched observed data patterns.

5.2 Limitations

Computational Cost: Bayesian methods require significantly more computation time, limiting real-time applications.

Point Forecast Performance: Some Bayesian variants sacrifice point fore-

cast accuracy for uncertainty quantification, requiring careful consideration of the accuracy-uncertainty trade-off.

Model Specification Sensitivity: Bayesian methods require careful prior specification and model design, which may affect results.

Crisis Definition: Our focus on the COVID-19 period may not generalize to other types of crises.

6 Conclusion

This comprehensive study provides important insights into the trade-offs between point forecast accuracy and uncertainty quantification in crisis-time inflation forecasting. While the Multi-Layer Perceptron achieves the best point forecast performance, Bayesian Dynamic Linear Models offer superior uncertainty quantification and model interpretability that are crucial for decision-making during volatile periods.

Key contributions include:

1. Empirical Evidence: Rigorous comparison demonstrating that model choice should depend on whether point forecasts or probabilistic information is prioritized.

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

3. Crisis-Specific Insights: Evidence that uncertainty quantification becomes particularly valuable during volatile periods, even when point forecast accuracy is slightly compromised.

4. Decision Framework: Clear guidance on when to prioritize point forecasts versus probabilistic modeling based on application requirements.

The results emphasize that during crisis periods, the ability to quantify and communicate forecast uncertainty may be more valuable than marginal improvements in

point forecast accuracy. Future research should explore ensemble methods that combine the strengths of neural networks and Bayesian approaches, and investigate the economic value of uncertainty quantification in decision-making frameworks.

Policy Implications: Central banks and financial institutions should consider incorporating Bayesian DLMs into their forecasting frameworks, particularly for stress testing, scenario analysis, and risk assessment during uncertain periods. While these methods may not always provide the most accurate point forecasts, their superior uncertainty quantification capabilities support more robust and transparent decision-making in crisis scenarios.

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>

Algorithm 1 One-Step-Ahead Forecasting with Particle Filtering

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

Algorithm 2 Particle Filter with Regime Switching

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

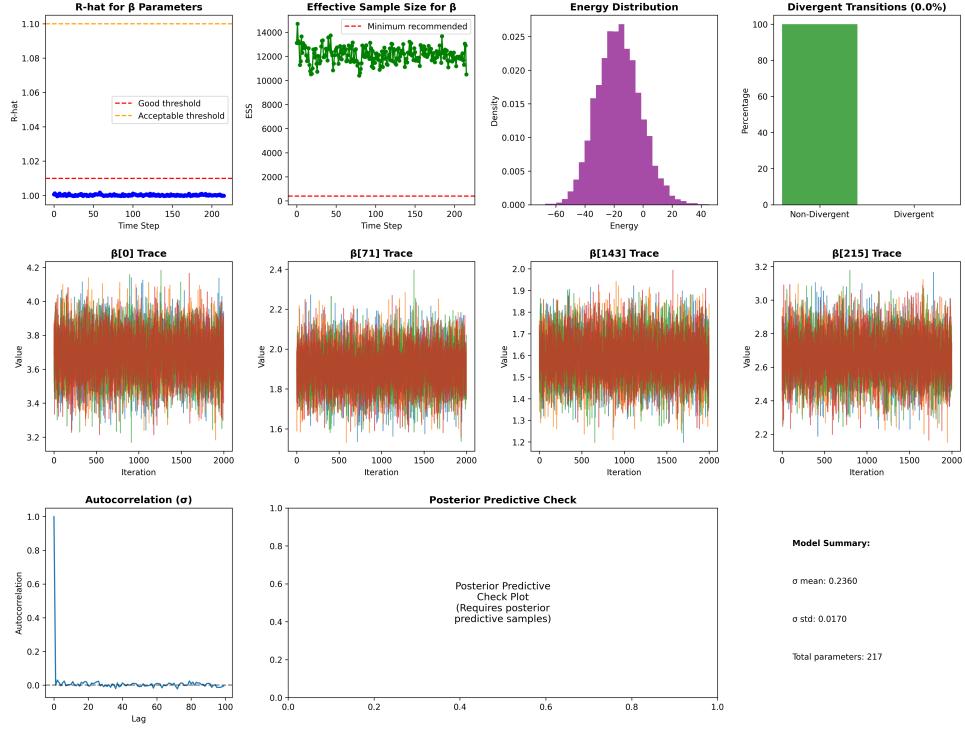


Figure 1: MCMC diagnostic plots showing convergence metrics, parameter traces, and sampling efficiency.

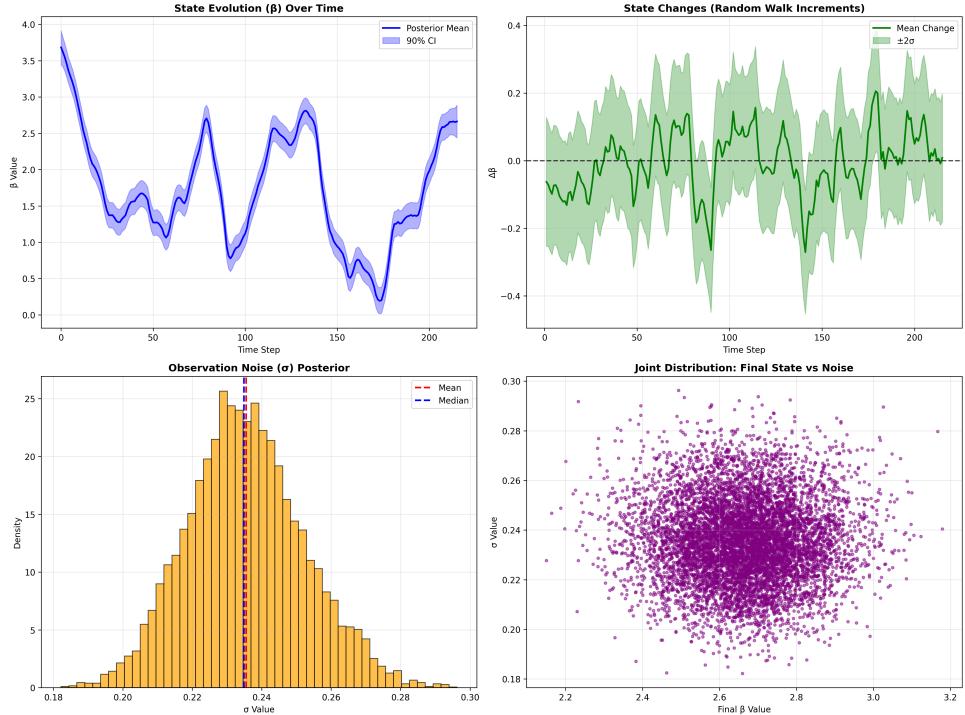


Figure 2: State evolution showing posterior mean trajectory with 90% credible intervals (top-left), random walk increments (top-right), observation noise posterior distribution (bottom-left), and joint distribution of final state vs noise parameters (bottom-right).

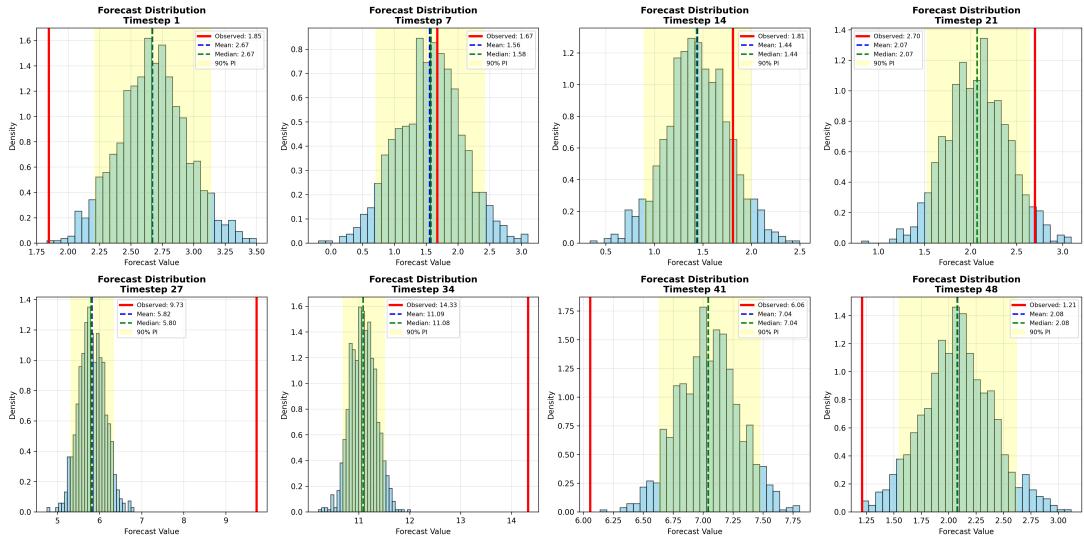


Figure 3: Forecast distribution evolution across timesteps showing adaptive uncertainty quantification.

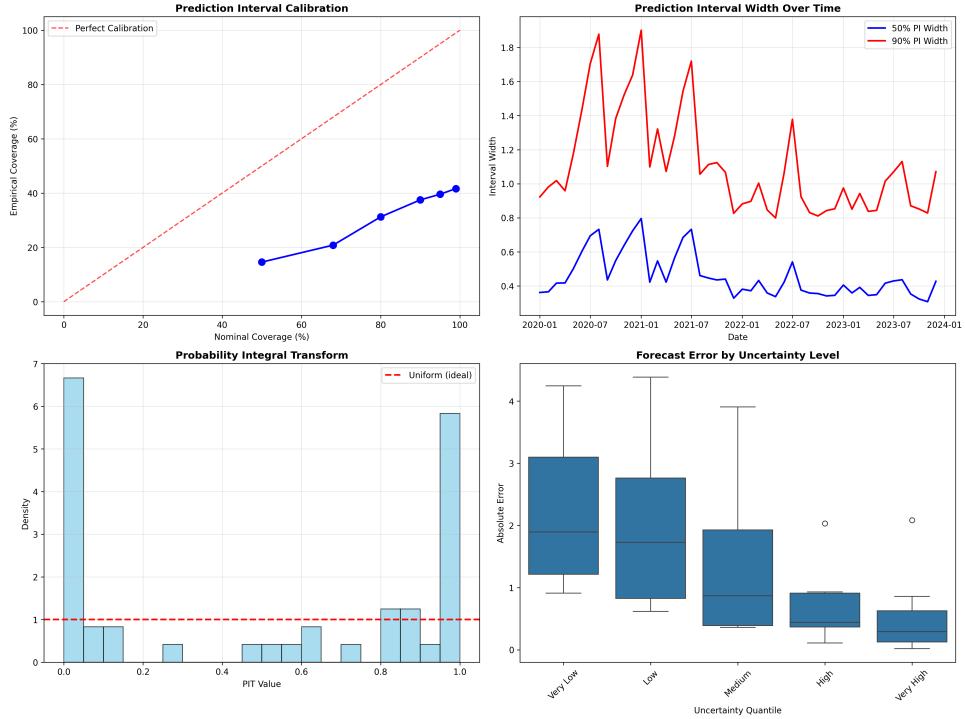


Figure 4: Comprehensive calibration analysis including empirical coverage, prediction interval evolution, and forecast error stratification.

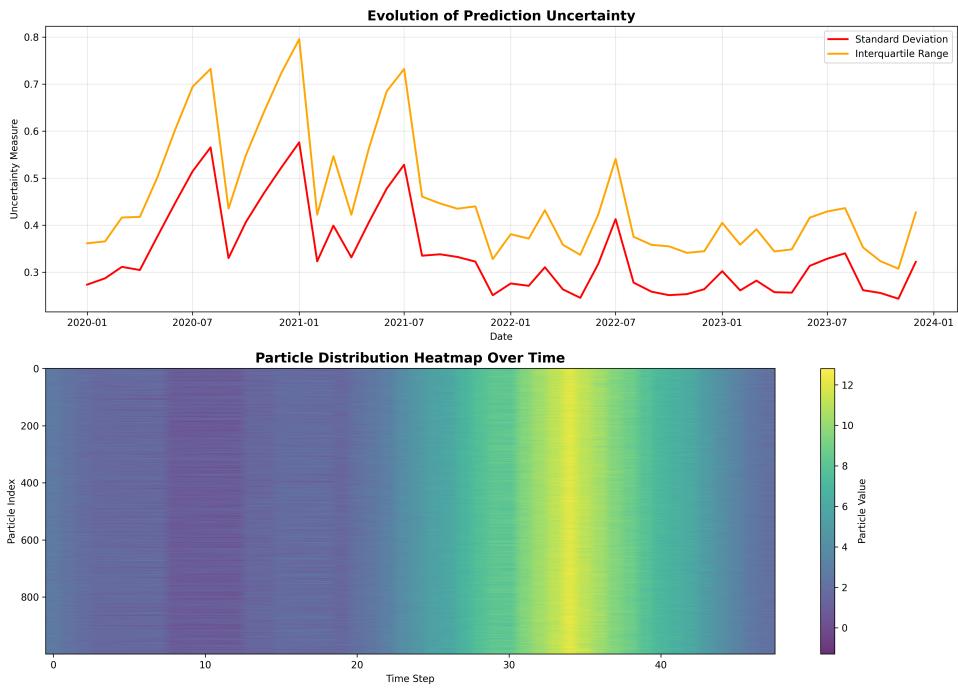


Figure 5: Evolution of prediction uncertainty metrics and particle filter distribution heatmap showing dynamic resampling behavior.

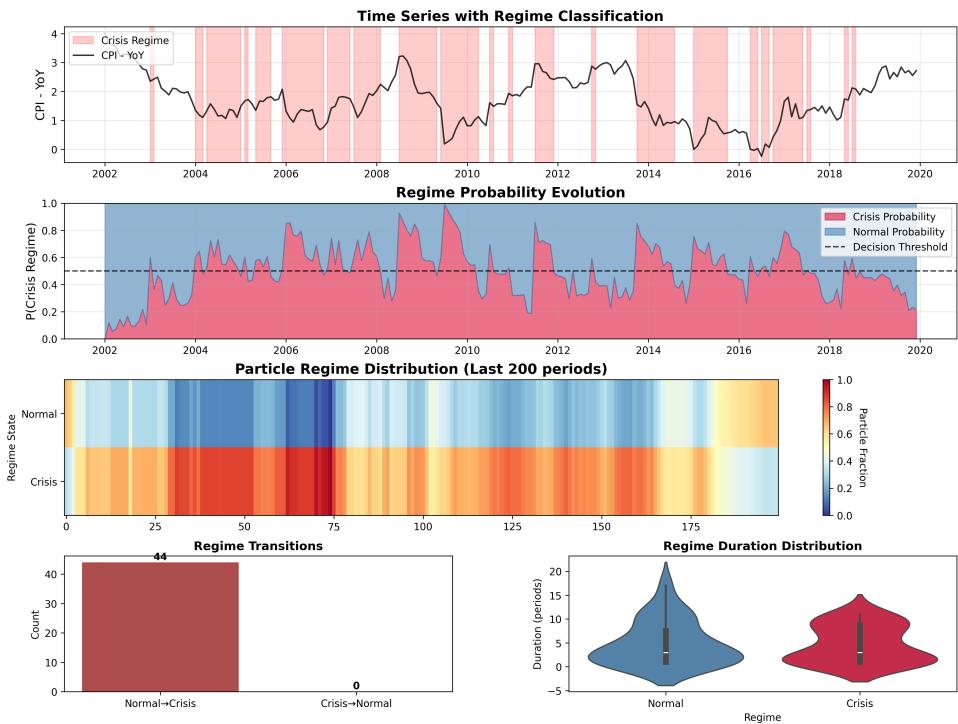


Figure 6: Regime Classification Analysis: Time series with regime identification, probability evolution, particle distribution, transition counts, and duration distributions.

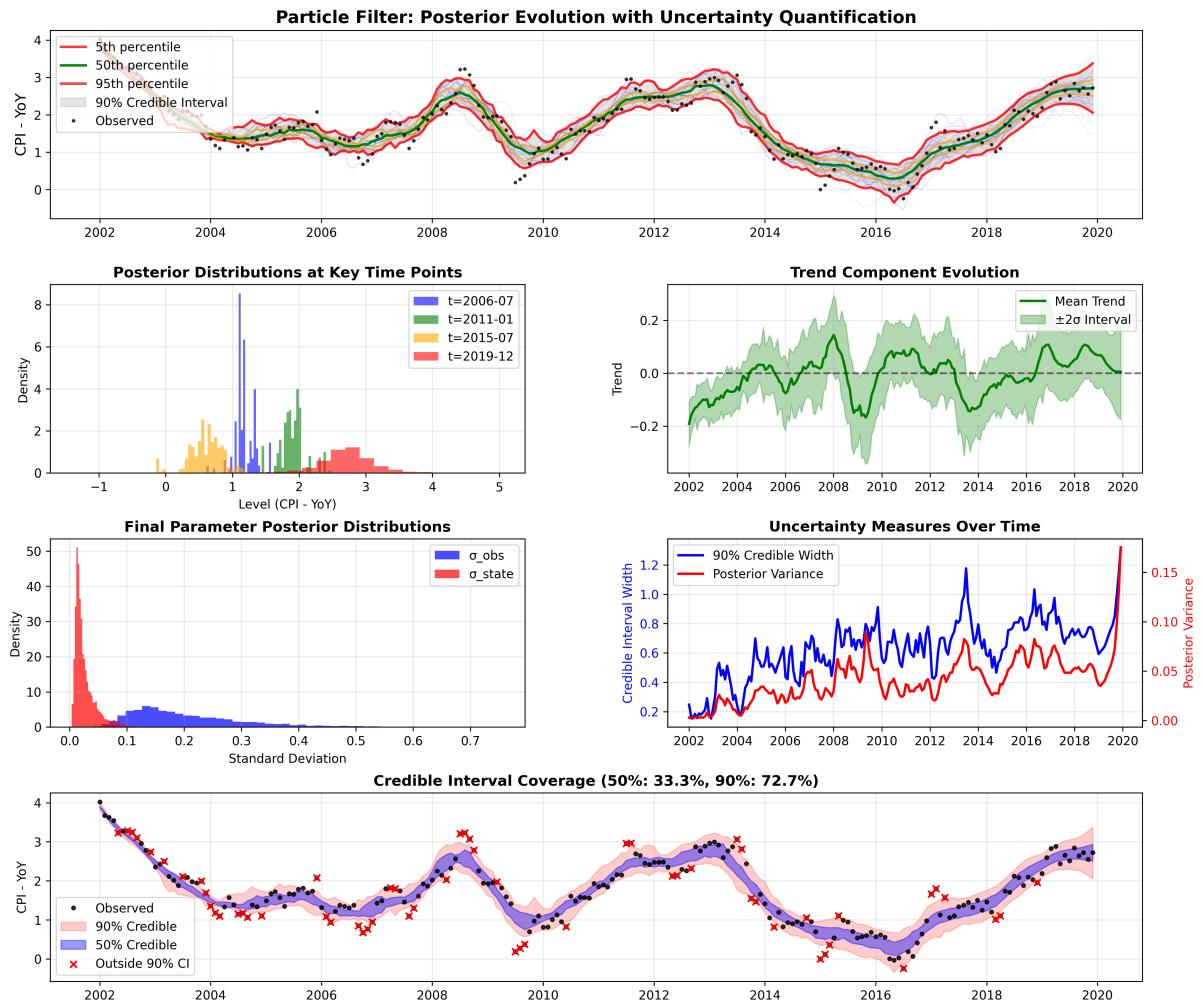


Figure 7: Particle Filter Results: Posterior evolution with uncertainty quantification, trend components, parameter distributions, and credible interval coverage analysis.

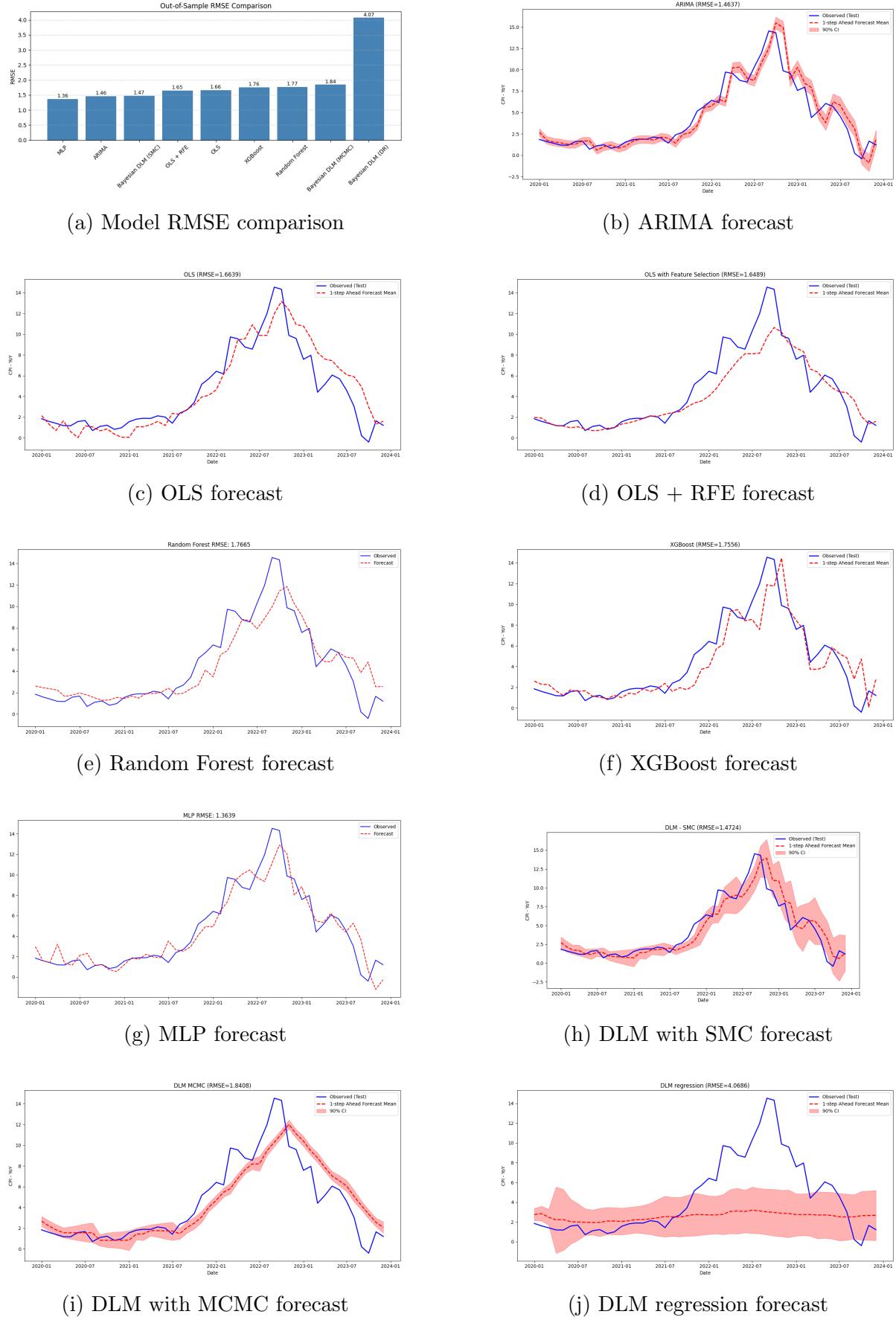


Figure 8: Forecast on test set