



**POLITECNICO**  
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE  
E DELL'INFORMAZIONE

# Bayesian Beats of the US Economy

BAYESIAN LEARNING AND  
MONTECARLO SIMULATION PROJECT  
COMPUTER SCIENCE AND ENGINEERING (T2I)

## Authors:

**Andrea Ferrarini**  
**Andrea Nardi**  
**Christian Giovanni Pesaturo**  
**Salim Salici**

## Student IDs:

10751746  
10777415  
10794871  
10640001

Professor: Federico Bassetti  
Academic Year: 2024-2025



# Contents

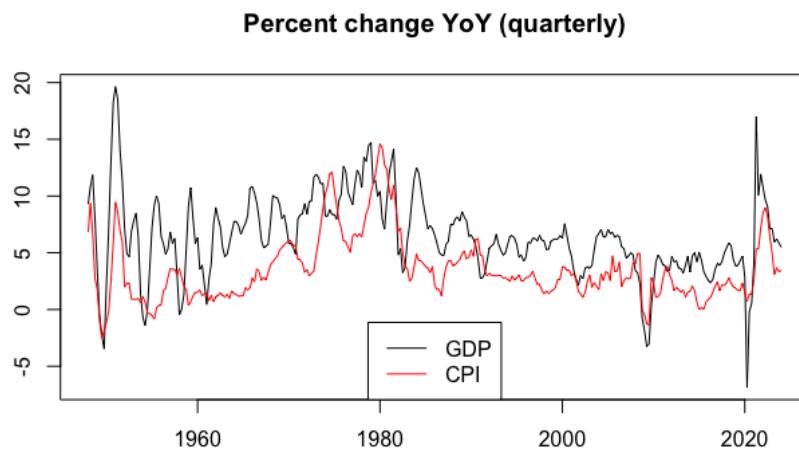
<b>Contents</b>	i
<b>Abstract</b>	i
<b>1 Empirical Dataset Overview: US Quarterly GDP and CPI Time Series</b>	1
1.1 Preliminary Exploratory Analysis . . . . .	1
1.2 Dataset Overview . . . . .	2
<b>2 Theoretical Foundations and Model Design in Time Series Inference</b>	3
2.0.1 Bayesian Modeling Rationale: Likelihood Construction and Prior Calibration . . . . .	3
<b>3 Bayesian Blueprint: Prior Beliefs and Model Structure</b>	7
3.0.1 Legend for Interpretation . . . . .	8
<b>4 Metrics of Truth: Evaluating Models Beyond the Horizon</b>	25
4.0.1 Performance Assessment of Univariate Time Series . . . . .	25
4.0.2 Into the Deeping Analysis . . . . .	28
<b>GDP Diagnostic Analisys</b> . . . . .	28
4.0.3 The <b>AR(9)</b> Process: Diagnostic Analysis . . . . .	28
4.0.4 The <b>MA(3)</b> Process: Diagnostic Analysis . . . . .	30
4.0.5 The <b>ARMA(2,3)</b> Process: Diagnostic Analysis . . . . .	32
<b>CPI Diagnostic Analisys</b> . . . . .	34
4.0.6 The <b>AR(9)</b> Process: Diagnostic Analysis . . . . .	34
4.0.7 The <b>MA(4)</b> Process: Diagnostic Analysis . . . . .	36
4.0.8 The <b>ARMA(3,3)</b> Process: Diagnostic Analysis . . . . .	38
4.0.9 The <b>VAR(1)</b> Process: Diagnostic Analysis . . . . .	40
4.0.10 Navigating the <b>ARMA</b> Landscape Heatmap . . . . .	42
<b>5 Bayesian Volatility Regime Detection for Time Series Anomalies</b>	45
<b>6 From Evidence to Implications: A Concluding Perspective</b>	49
<b>7 EXTRA</b>	51
7.0.1 Model Evaluation Metrics . . . . .	51



# Abstract

This project addresses the Bayesian modeling of **U.S. economic** time series through the application of autoregressive models with stochastic structure. Specifically, we study the **Gross Domestic Product** (GDP) growth rate and the **Consumer Price Index** (CPI) inflation rate, two critical indicators of macroeconomic health.

Using transformed series obtained from the **Federal Reserve Economic Data** (FRED), we construct a univariate probabilistic model for **GDP** growth. The core of the analysis is centered on Bayesian inference for  **$AR(n)$** ,  **$MA(n)$** ,  **$ARMA(p,q)$**  and  **$VAR(p)$**  type models, implemented via **JAGS**. Model estimation is carried out by simulating from the posterior distribution of the parameters, following a likelihood-prior structure tailored for time-dependent observations.



The report begins by exploring basic **AR**, **MA** and **ARMA** models, then transitions to a more robust **VAR** specification to account interdependencies between the 2 variables. The model is fitted to the first 85% of the data (training set), and predictive accuracy is evaluated on the remaining observations (test set). The methodology allows both in-sample reconstruction and out-of-sample forecasting, with uncertainty quantified via posterior credible intervals.

Visual diagnostics, posterior summaries, and prediction intervals support the evaluation of model adequacy. The results confirm the viability of Bayesian approaches in time series forecasting, especially when model flexibility and uncertainty quantification are key priorities.

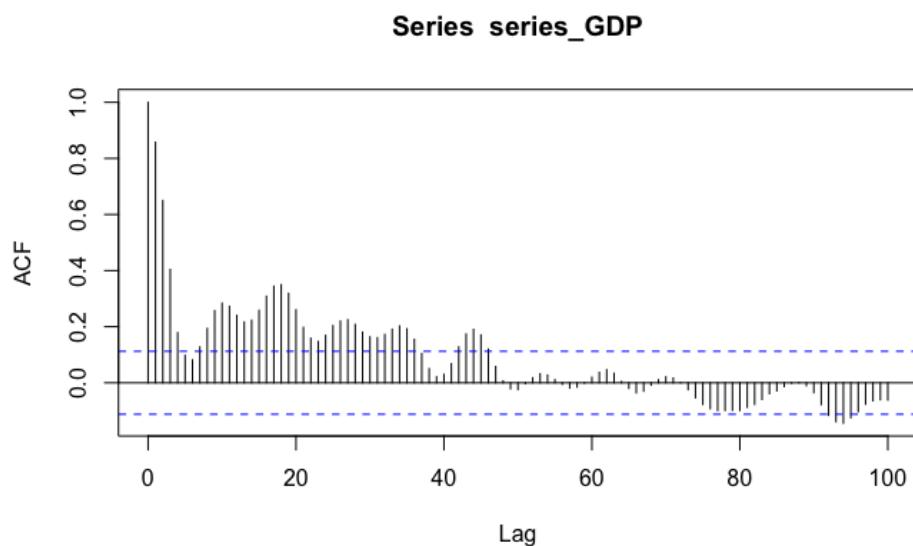
# 1 | Empirical Dataset Overview: US Quarterly GDP and CPI Time Series

## 1.1. Preliminary Exploratory Analysis

To evaluate the dynamic behavior of the time series, we compute and visualize their *Autocorrelation Functions* (ACFs).

### GDP ACF

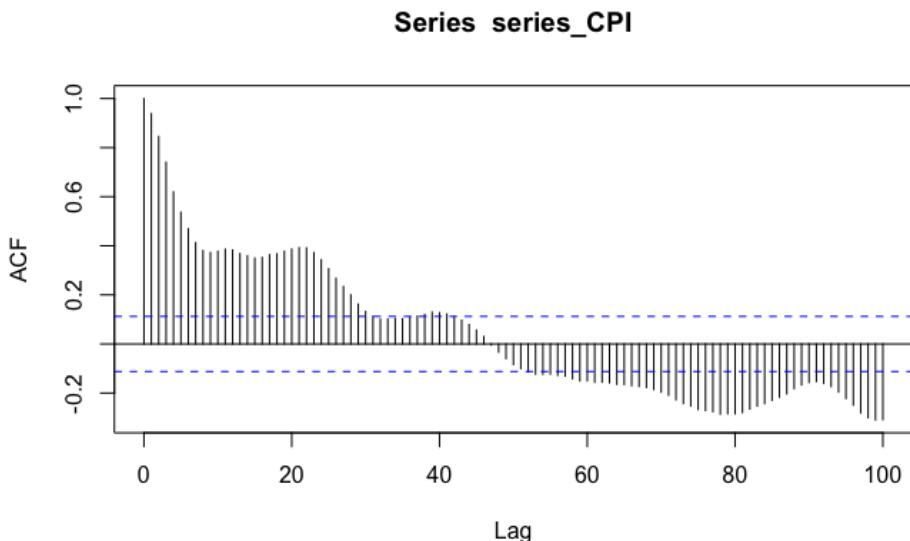
- The **GDP series** exhibits a strong spike at lag 1, followed by a slow decay with persistent positive autocorrelations across many lags.
- This behavior suggests non-stationarity and strong memory effects. The past information has a strong and persistent influence over time, meaning that current values in the series are highly dependent on a long sequence of past observations.



## 2 | Empirical Dataset Overview: US Quarterly GDP and CPI Time Series

### CPI ACF

- The **CPI series** also shows very high autocorrelation at initial lags, but the decay is smoother and slower compared to **GDP**.
- The gradual decline of the autocorrelation function indicates that the series may possess long memory characteristics, where the effect of past shocks decays slowly but remains influential.



### 1.2. Dataset Overview

The original file `gdp_inflation.csv` was pre-processed and aligned to guarantee a consistent length for both time series. The seasonally adjusted indeces measure price changes (as a percent change) from the previous year. In order to rigorously assess the predictive capabilities of the proposed models, the dataset was judiciously partitioned into two disjoint subsets: a training set and a test set.

- The training set, comprising 85% of the total observations, was employed to estimate the model parameters within a **Bayesian** framework.
- While the remaining 15% the test set was reserved for the purpose of out-of-sample forecasting and model validation.

This temporal split not only ensures the integrity of the inference but also permits a realistic evaluation of the models' forecasting performance on unseen data.

# 2 | Theoretical Foundations and Model Design in Time Series Inference

## 2.0.1. Bayesian Modeling Rationale: Likelihood Construction and Prior Calibration

### Likelihood: $\text{AR}(n)$

Rather than writing a separate JAGS block for each autoregressive order, we defined one generic  $\text{AR}(n)$  model string, where the order  $n$  is simply passed in as a data argument. This achieves two things:

- **Flexibility:** You can fit  $\text{AR}(1)$ ,  $\text{AR}(2)$ ,  $\text{AR}(3)$ , ... all in a single loop by varying the provided order parameter.
- **Maintainability:** All future changes for example to the prior on the  $\text{AR}$  coefficients need only be made once.

Recall that an AR(1) model is defined by

$$y(t) = a y(t - 1) + \eta(t), \quad \eta(t) \sim \mathcal{N}(0, \sigma^2) \quad (2.1)$$

Generalization of  $\text{AR}(n)$

$$y(t) = \left[ \sum_{i=1}^n \alpha_i y(t - i) \right] + \eta(t) \quad \eta(t) \sim \mathcal{N}(0, \sigma^2)$$

$$y(t) = \alpha_1 y(t - 1) + \alpha_2 y(t - 2) + \cdots + \alpha_n y(t - n) + \eta(t), \quad \eta(t) \sim \mathcal{N}(0, \sigma^2) \quad (2.2)$$

where  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ . In other words, it is simply a *linear combination* of  $y(t)$  on its previous  $n$  values, plus a noise.

## Prior of AR(n)

- **Intercept:**  $m_0 \sim \mathcal{N}(0, 10^4)$ :

Nearly flat Normal prior imparts almost no information on the mean level, letting the data dictate any non-zero drift.

- **Precision:**  $\sigma^2 = \tau^{-1}$ ,  $\tau \sim \text{Gamma}(0.1, 10)$ :

Gamma prior centers on a modest noise level while remaining diffuse enough to adapt to either low or high variability, avoiding overconfidence in the error term.

- **AR coefficient:**  $\alpha_i \sim \mathcal{U}(-1.5, 1.5)$ :

Uniform range slightly beyond  $\pm 1$  allows the sampler to explore near unit root behavior (long memory) if supported by the data, yet still constrains extreme, implausible dynamics.

## Likelihood: MA(n)

We now define an **MA(1)** (moving-average) process:

$$y(t) = \beta \eta(t-1) + \eta(t), \quad \eta(t) \sim \mathcal{N}(0, \sigma^2) \quad (2.3)$$

An **MA(1)** captures serial correlation by linearly combining the current white-noise shock  $\eta(t)$  with the immediately preceding shock  $\eta(t-1)$ . Such a model is particularly useful in time series analysis when short term dependencies (one lag) dominate and higher order moving averages add little explanatory power.

As with the **AR** case, we defined a generic JAGS implementation for the **MA(n)** process, enabling flexible evaluation across different orders by simply adjusting the order parameter.

Generalization of **MA(n)**:

$$y(t) = \left[ \sum_{i=1}^n \beta_i \eta(t-i) \right] + \eta(t), \quad \eta(t) \sim \mathcal{N}(0, \sigma^2)$$

## Prior of MA(n)

- **Intercept:**  $m_0 \sim \mathcal{N}(0, 10^4)$

An almost flat Normal prior that leaves the series mean entirely data-driven.

- **Precision:**  $\tau \sim \text{Gamma}(0.01, 0.01)$

Highly diffuse Gamma prior on the noise precision, accommodating both low and

high variance regimes.

- **MA coefficient:**  $\beta \sim \mathcal{N}(0, 10)$

Normal prior on **MA** coefficients introduces slight shrinkage toward zero, offering a more regularized alternative to the uniform prior while preserving flexibility in capturing short-term shocks.

## Likelihood: ARMA(1,1)

After having explored both auto-regressive and moving average components independently, we now propose a more comprehensive model that integrates both effects: the **ARMA(1,1)** process. This specification combines the dynamics of **AR(1)** and **MA(1)** structures into a single model, capturing both short-term memory and random shock propagation. The model is expressed as follows:

$$y(t) = \alpha y(t-1) + \beta \eta(t-1) + \eta(t), \quad \eta(t) \sim \mathcal{N}(0, \sigma^2) \quad (2.4)$$

where  $\alpha$  governs the persistence of the process and  $\beta$  controls the contribution of the previous white noise term. This formulation allows the model to capture both autoregressive dynamics and noise-driven volatility.

The motivation for selecting an **ARMA(p,q)** model arises from the observation that higher-order autoregressive models (i.e., **AR(n)** with  $n > 1$ ) did not yield significant improvement in predictive accuracy. On the contrary, model comparison metrics (such as **WAIC**) suggested that a deeper memory structure was sufficient to capture the temporal dependencies in our dataset. The **ARMA(p,q)** model thus provides a parsimonious yet effective trade-off between model complexity and explanatory power.

Generalization of **ARMA(p,q)**:

$$y(t) = \left[ \sum_{i=1}^p \alpha_i y(t-i) \right] + \left[ \sum_{i=1}^q \beta_i \eta(t-i) \right] + \eta(t), \quad \eta(t) \sim \mathcal{N}(0, \sigma^2)$$

## Prior of ARMA(p,q)

- **Intercept:**  $m_0 \sim \mathcal{N}(0, 10^4)$

An almost flat Normal prior that leaves the series mean entirely data-driven.

- **AR coefficient:**  $\alpha \sim \mathcal{U}(-1.5, 1.5)$

Uniform prior enforcing the invertibility condition while remaining non-informative about the **MA** effect.

- **MA coefficient:**  $\beta \sim \mathcal{N}(0, 10)$

Normal priors on **MA** coefficients provide mild regularization while allowing flexible dynamics.

- **Noise precision:**  $\tau \sim \text{Gamma}(0.01, 0.01)$

Highly diffuse Gamma prior on the error precision, accommodating both low and high noise regimes.

## Likelihood: **VAR(n)**

To enrich the inferential scope of our analysis, we extend the univariate time series approach by jointly modeling the **U.S. GDP** growth and inflation dynamics through a **Vector Autoregressive** model **VAR(n)**. Unlike separate univariate models, a **VAR** framework allows us to account for contemporaneous feedback and temporal interdependencies between multiple economic indicators.

The model is specified as:

$$\vec{y}_t = \sum_{o=1}^n A_o \vec{y}_{t-o} + \vec{m}_0 + \vec{\varepsilon}_t, \quad \vec{\varepsilon}_t \sim \mathcal{N}(\vec{0}, \Sigma)$$

## Prior of **VAR(n)**

- **Transition coefficients:**  $A_{i,j,o} \sim \mathcal{U}(-1.2, 1.2)$

As with the **AR** process, we consider a uniform range slightly beyond  $\pm 1$  so as to allow the sampler to explore near unit root behavior while still constrains extreme values.

- **Innovation precision matrix:**  $\tau \sim \text{Wishart}(I, 3)$ ,  $\Sigma = \tau^{-1}$

Conjugate Wishart prior with scale  $I$  and minimal degrees of freedom  $\nu = 3$ , yielding a diffuse yet proper prior over the  $2 \times 2$  covariance.

- **Intercept vector:**  $m_{\text{gdp}}, m_{\text{infl}} \sim \mathcal{N}(0, 10^4)$ ,  $\mathbf{m}_0 = \begin{pmatrix} m_{\text{gdp}} \\ m_{\text{infl}} \end{pmatrix}$

Nearly flat Normal priors on each intercept component, ensuring the long-run means are driven by the data.

The others priors that we use during our comparison are specified significantly at 3.0.1.

# 3 | Bayesian Blueprint: Prior Beliefs and Model Structure

We now begin our discussion aimed at understanding the temporal dynamics within the **GDP** and, subsequently, **CPI** datasets through the application of **AR**, **MA**, **ARMA**, and **VAR** models. Our journey into the deeper structure of these models will unfold in two distinct stages: we will start with the analysis of **GDP**, followed by that of **CPI**. In this section, we present time series plots along with a discussion justifying the selected parameter values. We conclude with a brief qualitative comment on the observed dynamics.

The next chapter will focus on model comparison using several evaluation criteria, including both complexity-penalized metrics and prediction accuracy scores:

- **DIC** (Deviance Information Criterion),
- **WAIC** (Watanabe–Akaike Information Criterion),
- **BIC** (Bayesian Information Criterion),
- Classical test error metrics (on unseen data) metrics such as:
  - **MSE** (Mean Squared Error)
  - **MAE** (Mean Absolute Error)
  - **MAPE** (Mean Absolute Percentage Error)
  - **SMAPE** (Symmetric Mean Absolute Percentage Error)
  - **MASE** (Mean Absolute Scaled Error)

These indicators allow us to quantify both model fit and out-of-sample predictive accuracy.

Caveat lector: All the metrics used are explained in detail in Chapter 7

### 3.0.1. Legend for Interpretation

In order to guide the reader toward a robust and informed interpretation of the graphical evidence presented, we provide below the essential keys to correctly decode the underlying dynamics.

Legend	
— red	Observed data
* blue	In-sample predictions (mean)
— blue	95% credible interval (in-sample)
* orange	Out-of-sample forecasts (mean)
— orange	95% credible interval (out-of-sample)
--- orange	50% credible interval (out-of-sample)
magenta	End of training set
— green	Posterior mean of $m_0$

- **Full-range plot (in-sample + out-of-sample)**

The first panel illustrates the entire time series dynamics. The observed data are shown in red, over which the following elements are superimposed:

- *In-sample predictions*, plotted as blue asterisks, accompanied by 95% credible intervals represented by — solid blue lines connecting the 2.5% and 97.5% quantiles.
- *Out-of-sample forecasts*, plotted as \* orange asterisks, with:
  - \* 95% credible bands: — solid orange lines between the 2.5% and 97.5% quantiles;
  - \* 50% credible bands: dashed - - - dashed orange lines between the 25% and 75% quantiles .
- A | dashed magenta vertical line marks the end of the training set, clearly separating the in-sample from the out-of-sample region.
- A - - - dashed green horizontal line represents the posterior mean estimate of  $m_0$ , i.e., the average level of the series as inferred by the model.

- **Zoom on the out-of-sample forecast**

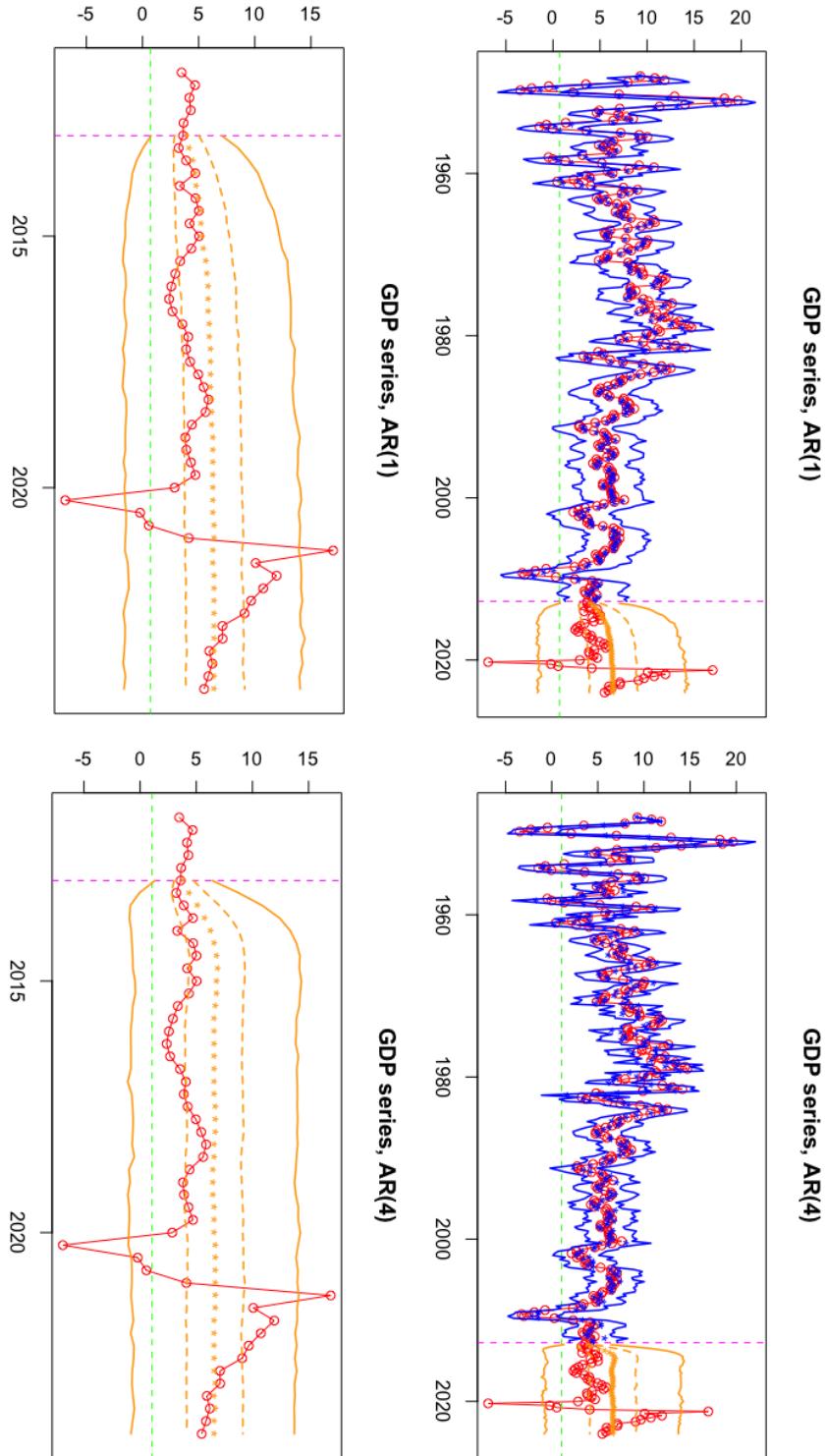
The second panel zooms into the out-of-sample interval. Again:

- The real series is plotted in red;

- The same **magenta** and **green** reference lines are maintained;
- The out-of-sample forecasts and their associated **95%** and **50%** **credible intervals** are shown as in the previous panel, providing a more focused view on predictive performance.

This two-panel visualisation allows for a comprehensive assessment of the model's fit to historical data and the reliability of future projections. The overlap between observed and predicted values enables immediate visual comparison, highlighting both the pointwise accuracy and the uncertainty estimated by the model.

## GDP Dynamics under AR Modeling



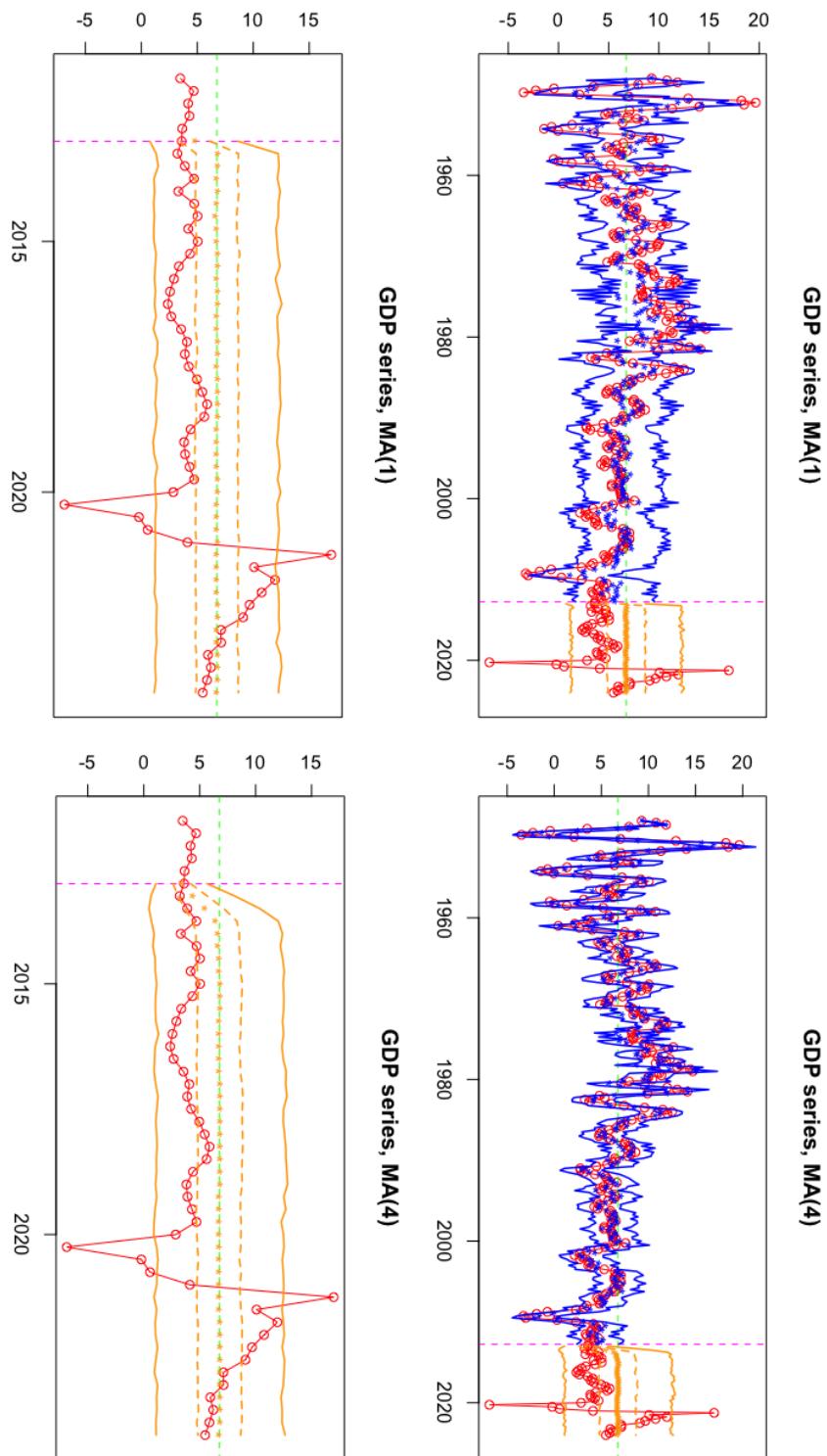
## Comparison of AR(1) and AR(4)

From the analysis of the plots in the figure, two clearly distinct behaviors emerge. The ***AR(1)*** model, while being extremely responsive to the most recent movements, struggles to capture medium-term oscillations, which is reflected in wider uncertainty bands (yellow lines) immediately after the training/test split.

In contrast, the ***AR(4)*** model incorporates information over a longer horizon, that is, the observed values of the time series in the four previous time points, absorbing peaks and drops more gradually: the forecast trajectory is smoother and the uncertainty bands are overall narrower in the out-of-sample.

In this way, the ***AR(4)*** provides a less volatile forecast, although at the cost of a more delayed response to recent shocks. These elements suggest that, for our quarterly **U.S. GDP dataset**, an intermediate order (such as 4) represents a good compromise between responsiveness and stability of the forecasts.

## GDP Dynamics under MA Modeling



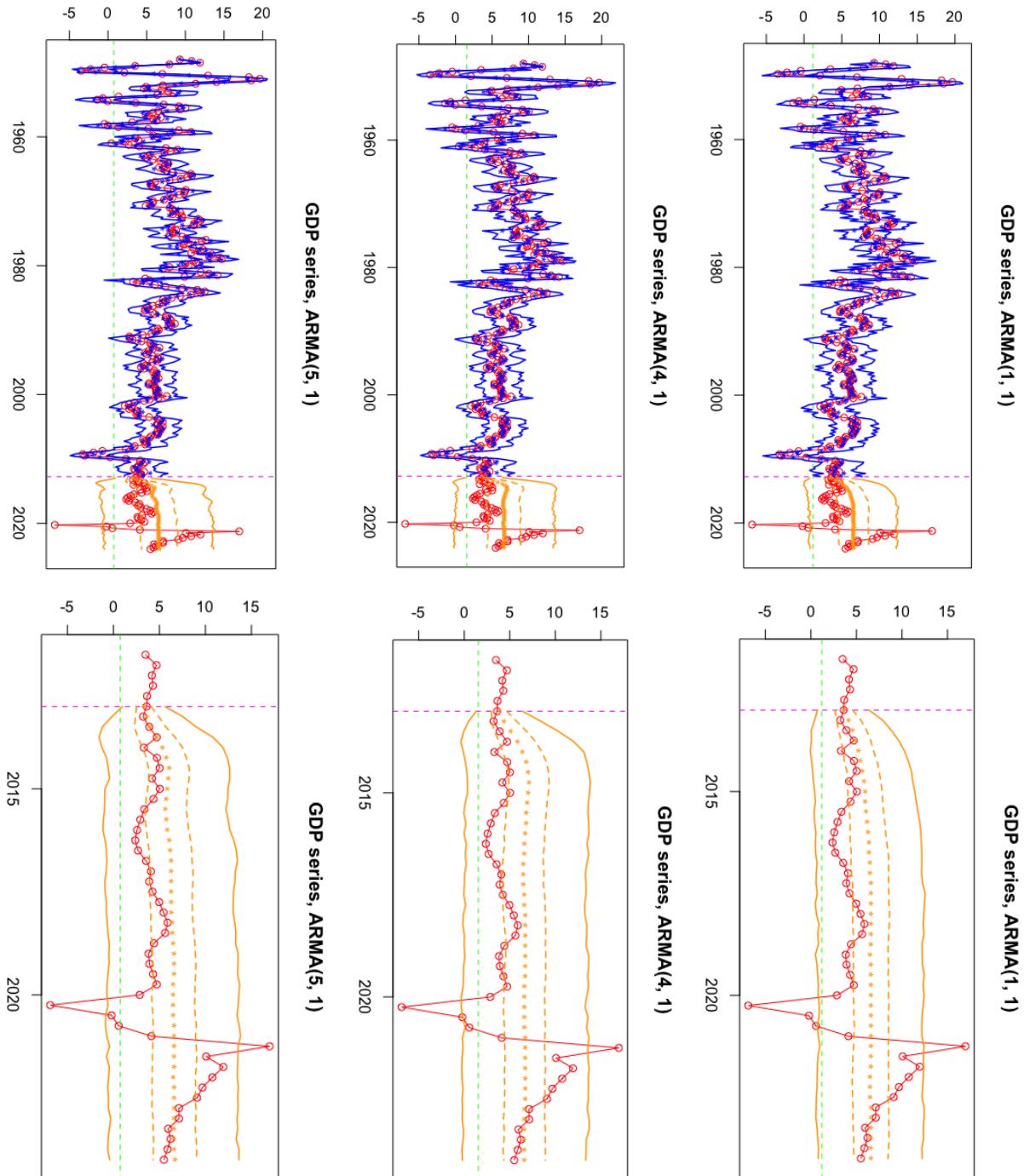
## Comparison of MA(1) and MA(4)

From this comparison we observe:

- **MA(1)** relies solely on the single most recent residual, producing highly *reactive* forecasts: its out-of-sample uncertainty band is wide and adjusts immediately to the last training point's fluctuation.
- **MA(4)**, by incorporating the four latest residuals, yields *smoothed* predictions: its point forecasts fluctuate less and its credibility intervals narrow, indicating greater stability at the cost of a slower response to abrupt shocks.

In essence, **MA(4)** averages extreme events over several past lags dampening their influence on future predictions, whereas **MA(1)** reflects each recent spike directly in its forecast.

## GDP Dynamics under ARMA Modeling



## Comparison of ARMA(1,1), ARMA(4,1), ARMA(5,1)

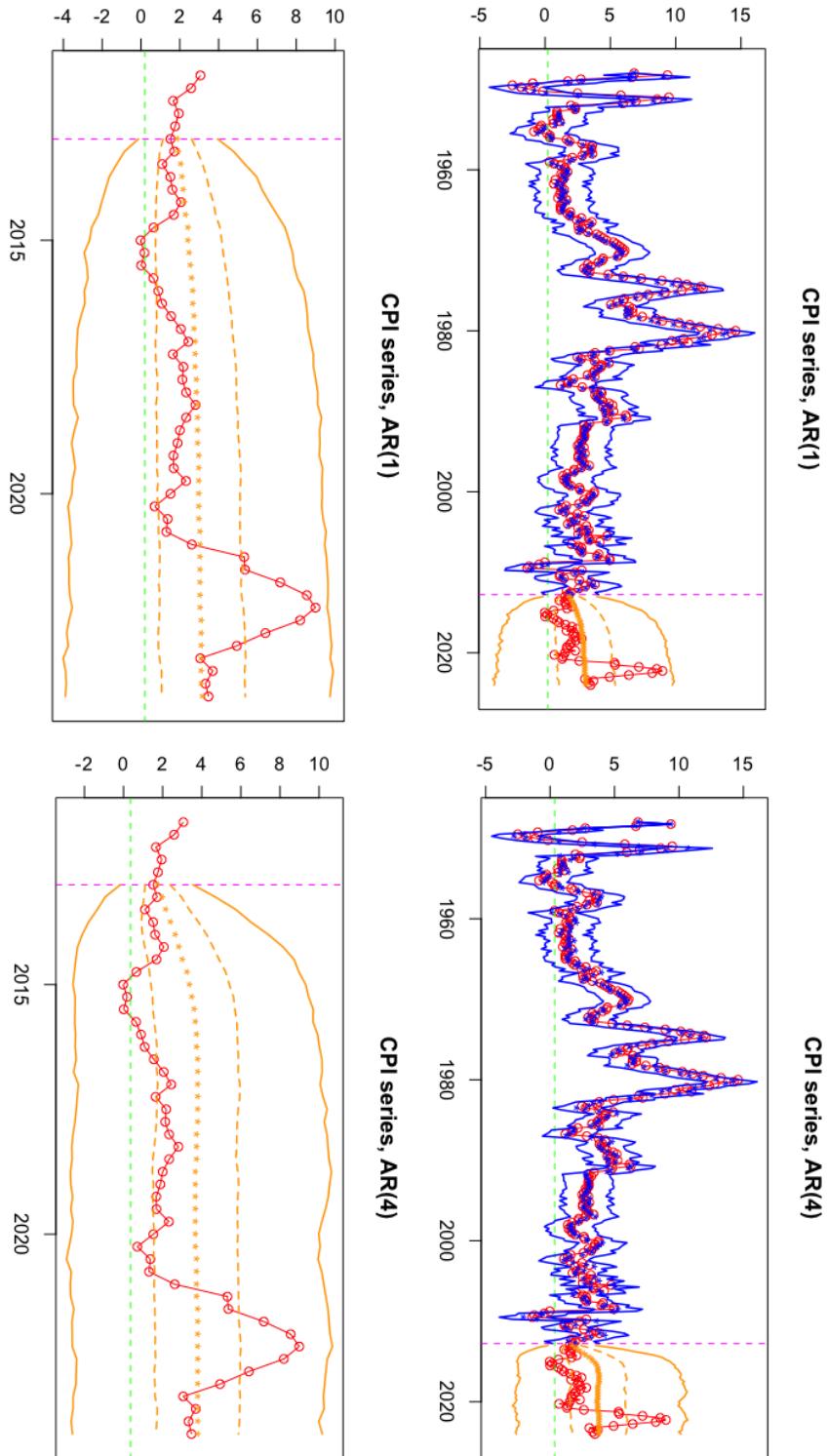
In describing the dynamics of the **GDP** via **ARMA** models, the interplay between autoregressive and moving-average terms becomes immediately apparent. The **ARMA(1,1)** model tracks very recent fluctuations with remarkable agility thanks to its single lag **AR** component but this narrow horizon comes at the cost of notably wider credible bands as soon as one moves beyond the training sample.

Extending the autoregressive order to four lags, as in **ARMA(4,1)**, provides a smoothing effect: peaks and troughs are absorbed more gradually, the prediction path appears more polished, and the credibility intervals narrow significantly, achieving greater stability without sacrificing responsiveness. Pushing further to **ARMA(5,1)** offers only marginal benefits, indicating that beyond a certain point additional lags offer only partial returns.

Ultimately, for our quarterly sample of **U.S. GDP**, the specification **ARMA(5,1)** achieves the best equilibrium: it “remembers” medium-term fluctuations while readily adjusting for recent shocks and offers narrower uncertainty bands overall.

The medellum **ARMA(5,1)** should be considered the best within this comparison set. For additional and general results view the table 4.0.1.

## CPI Dynamics under AR Modeling



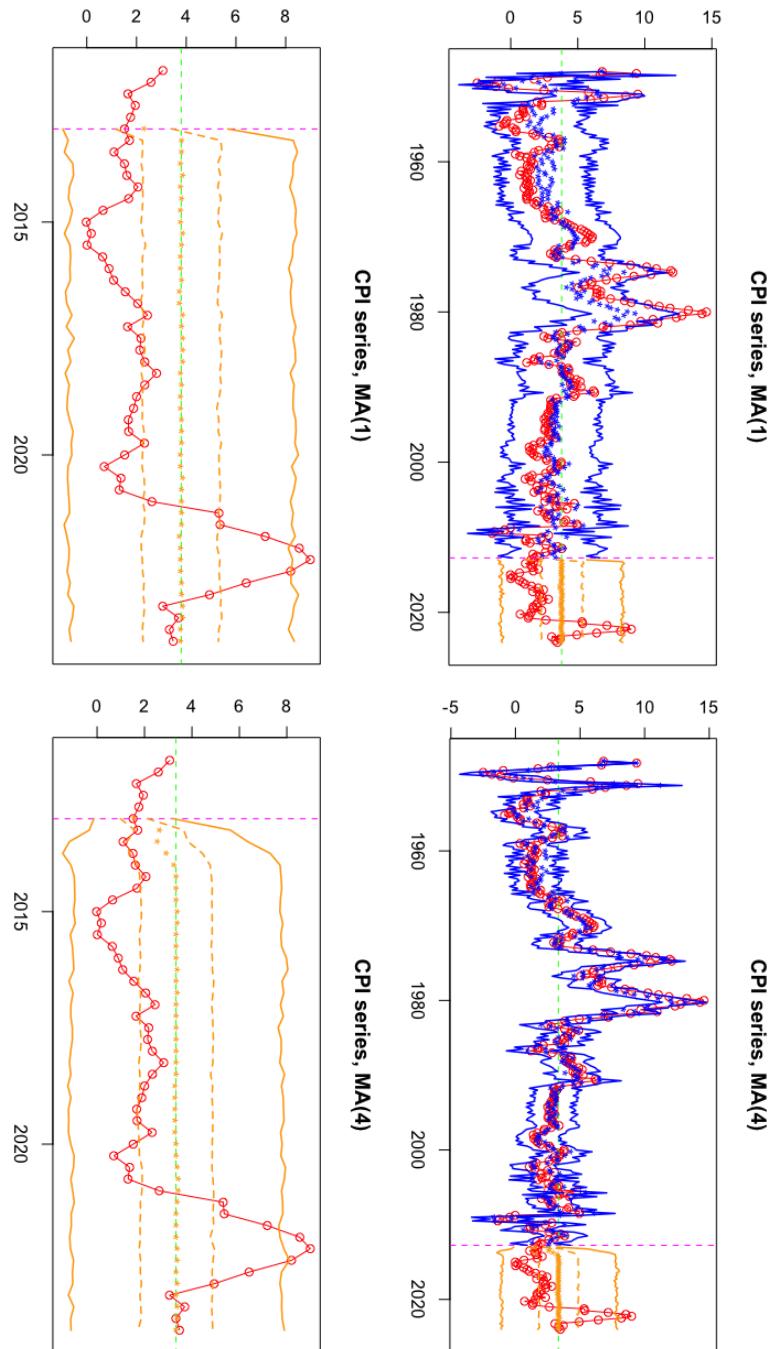
## Comparison of AR(1) and AR(4)

For the **CPI**, the comparison between **AR(1)** and **AR(4)** reveals dynamics analogous to those observed for the **GDP**, but on a more pronounced price-change scale:

- The **AR(1)** model reacts very quickly to the most recent monthly inflation swings, yet produces relatively wide uncertainty bands beyond the training window, since it “forgets” all but the last observation.
- The **AR(4)** model incorporates information from the past four months, naturally smoothing sharp spikes and tightening the out-of-sample credible intervals. The result is a more stable forecast that still retains reasonable responsiveness to inflation peaks.

Overall, even for the **CPI**, an intermediate lag order (4) offers an effective balance: it captures short-term movements without unduly inflating forecast uncertainty.

## CPI Dynamics under MA Modeling



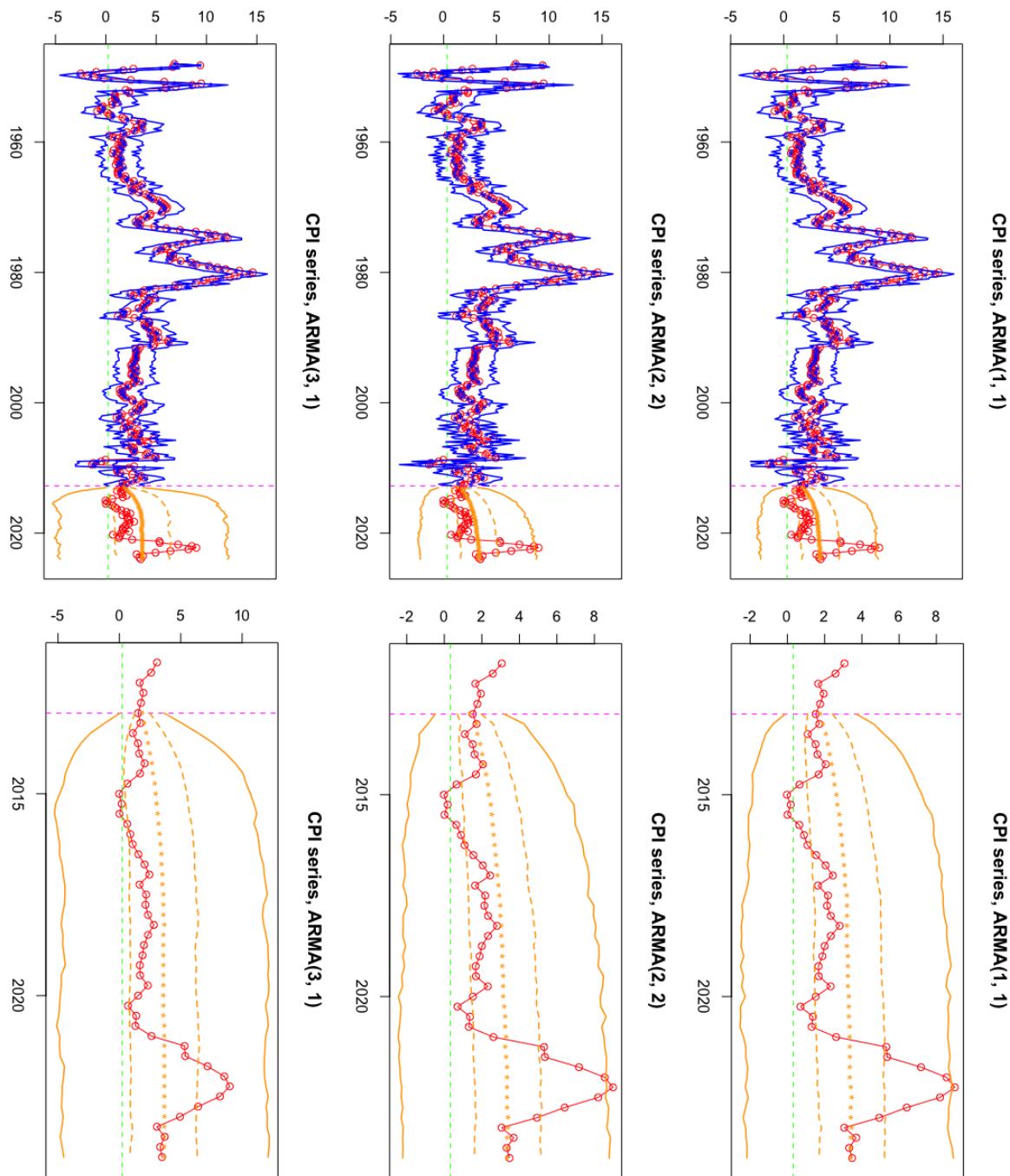
## Comparison of MA(1) and MA(4)

In these panels we compare two moving-average specifications applied to the **CPI** series:

- **MA(1)** relies on a single past innovation, yielding high sensitivity to monthly spikes but relatively wide out-of-sample credible bands.
- **MA(4)** incorporates the four most recent residuals, which smooths abrupt jumps and tightens the credible intervals, resulting in more stable forecasts while still accommodating short-term shocks.

Overall, even for the **CPI**, using four past observations achieves a balance between adaptability and predictive accuracy.

## CPI Dynamics under ARMA Modeling



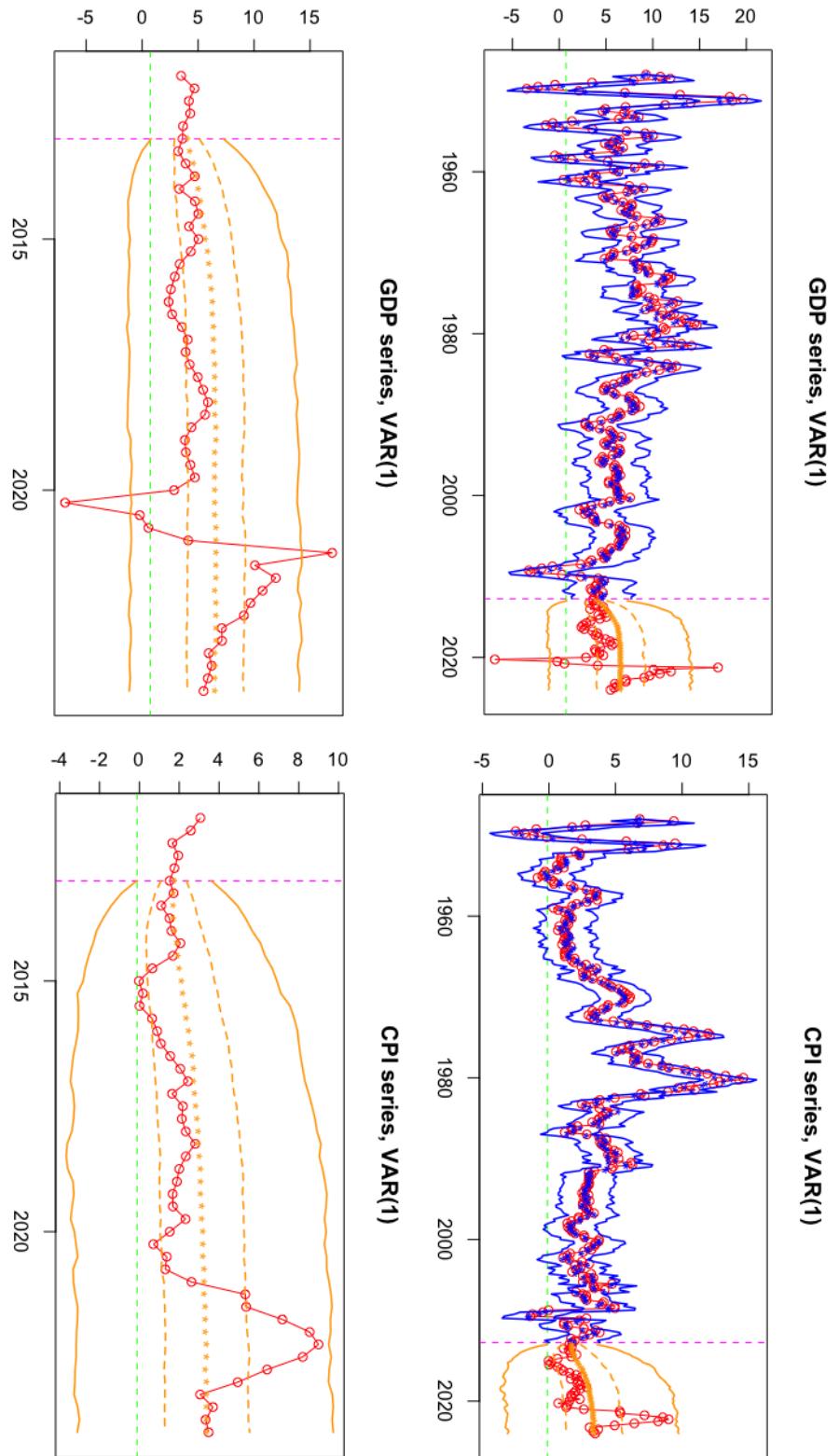
## Comparison of ARMA(1,1), ARMA(2,2), ARMA(3,1)

In the panels below we compare the in-sample fit and out-of-sample forecasts for three **ARMA** specifications applied to the **CPI** series:

- **ARMA(1,1)** combines one autoregressive lag and one moving-average term, achieving high sensitivity to recent fluctuations but relatively wide credible bands immediately after the training/test split.
- **ARMA(2,2)** incorporates two lags in both **AR** and **MA** components, better capturing both short and medium-term oscillations and yielding somewhat tighter credible intervals.
- **ARMA(3,1)** uses three autoregressive lags with a single **MA** term, providing an intermediate profile that balances responsiveness and stability in the forecast trajectory.

Overall, the choice of lag orders governs the trade-off between rapid shock absorption and uncertainty compression: for our quarterly **CPI** sample, an intermediate specification such as **ARMA(2,2)** appears to deliver the most balanced performance.

## GDP-CPI Dynamics under VAR Modeling



## Assessment of Contemporaneous Correlation through Prior Specification

The objective of our analysis is to determine whether a certain degree of correlation exists between the respective stochastic components of the two time series. To formally investigate the residuals estimated under a **Wishart** prior for the **GDP** and **CPI** series, we compute the *Pearson* correlation coefficient, which allows us to quantify the strength and direction of the linear relationship between the errors of two macroeconomic variables.

**Case 1: Diagonal Prior (Uncorrelated Errors)** Our empirical exploration will initially focus on the deliberate choice of a diagonal prior in order to impose a zero-correlation structure between the residuals of **GDP** and **CPI**.

Consequently, given the absence of non-zero entries outside the diagonal in the precision matrix (i.e., non-zero values only along the diagonal), we enforce the assumption of no correlation between the two variables. In doing so, we model a dynamic in which shocks affect only the series to which they belong, thereby enabling a meaningful comparison with the alternative hypothesis of correlated errors, which will be addressed through the specification of a *Wishart* prior.

$$\tau = \begin{pmatrix} \tau_{11} & 0 \\ 0 & \tau_{22} \end{pmatrix}, \quad \text{with } \tau_{11}, \tau_{22} \sim \text{Gamma}(a, b).$$

**Case 2: Wishart Prior (Correlated Errors)** In contrast, by adopting a **Wishart** prior, we allow the covariance between the residuals to be freely estimated. Accordingly, we define a precision matrix with non-zero entries outside the diagonal, as long as it remains positive definite, i.e., invertible and allowing the density to be well-defined. In this way, the model is able to estimate the strength of the correlation between **GDP** and **CPI** directly from the data.

$$\tau \sim \text{Wishart}(R, 3), \quad \text{where } R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This enables estimation of the off-diagonal covariance terms directly from the data.

## Posterior Means

**Using Diagonal Prior:**

$$A = \begin{pmatrix} 0.9085 & -0.0350 \\ 0.0801 & 0.8854 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 2.6105 & 0 \\ 0 & 0.9590 \end{pmatrix}$$

**Using Wishart Prior:**

$$A = \begin{pmatrix} 0.9083 & -0.0364 \\ 0.0795 & 0.8854 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 2.5838 & 0.5505 \\ 0.5505 & 0.9538 \end{pmatrix}$$

**Empirical Results** The *Pearson* correlation coefficient between the residuals obtained from the model with a Wishart prior is:

$$\rho(\text{err}_{\text{GDP}}, \text{err}_{\text{CPI}}) = \frac{\text{Cov}(\text{err}_{\text{GDP}}, \text{err}_{\text{CPI}})}{\sqrt{\text{Var}(\text{err}_{\text{GDP}}) \cdot \text{Var}(\text{err}_{\text{CPI}})}} = \frac{0.5505}{\sqrt{2.5838 \cdot 0.9538}} \approx 0.3507.$$

Based on the result of the *Pearson* correlation coefficient which allows us to check whether it is reasonable to assume correlation between the residuals and to quantify its intensity we can conclude that a value of approximately 0.35 represents a moderate and positive correlation.

The correlation is not null, yet not dominant either: it improves, albeit slightly, the modeling of contemporaneous volatility, but it does not affect the lagged dynamics—indeed, the matrix  $A$  remains nearly unchanged.

We specified these priors for the **VAR** because, in sensitivity checks using alternative prior formulations, the estimated dynamic relationships did not change significantly.

# 4 | Metrics of Truth: Evaluating Models Beyond the Horizon

## 4.0.1. Performance Assessment of Univariate Time Series

Performance comparison across AR, MA, and ARMA models on evaluation metrics. Best results in each block are highlighted in bold.

We will now undertake a rigorous assessment of the validity and effectiveness of our analyses by means of a suite of evaluation metrics designed to guide the challenging task of identifying which specific model or entire model family and its parametrization best capture and forecast the dynamics of our **GDP** and **CPI** indicators.

We will then move on to a deeper examination of the ideal model choice, one that is supported not only by formal mathematical criteria but also by a thoroughly empirical investigation.

As shown in the two tables, each column corresponds to a distinct evaluation metric, while each row lists the candidate models grouped by model family along with their respective parametrizations.

**Important note:** we will principally rely on the **Bayesian Information Criterion** (BIC) as our primary selection criterion, since it most effectively balances in-sample predictive performance against model complexity.

Finally, our discussion will also include visual diagnostics such as trace plots, density plots, and interval plots.

The computed metrics indicate that **ARMA(2, 3)** and **ARMA(3, 3)** emerge as the top-performing models for both the **GDP** and **CPI**.

Particularly noteworthy is the performance of the **AR(9)** model on the **CPI** data: its out-of-sample forecast over the following 18 months, based solely on in-sample information, shows an almost perfect alignment with the observed values.

## GDP-POWER

Model	DIC	WAIC	BIC	MSE	MAE	MAPE	SMAPE	MASE
<b><i>AR(n)</i></b>								
AR(1)	745.9	705.9	694.3	7.955	<b>1.384</b>	<b>75.0</b>	26.8	<b>1.200</b>
AR(2)	689.4	666.5	665.6	11.903	1.810	141.1	31.7	1.570
AR(3)	689.2	666.8	670.3	11.406	1.745	136.3	30.7	1.513
AR(4)	679.0	663.3	673.7	11.815	1.738	135.7	29.7	1.507
AR(5)	<b>644.3</b>	630.8	<b>647.0</b>	8.339	1.534	122.7	26.8	1.330
AR(6)	650.4	634.1	652.9	8.429	1.547	126.0	27.4	1.341
AR(7)	649.0	634.2	660.4	8.381	1.548	123.4	27.2	1.342
AR(8)	653.6	634.9	659.2	8.473	1.614	125.7	28.7	1.400
AR(9)	647.6	<b>627.1</b>	655.3	<b>7.697</b>	1.475	126.0	<b>26.1</b>	1.279
<b><i>MA(n)</i></b>								
MA(1)	846.7	835.2	838.6	9.149	2.085	115.2	42.7	1.808
MA(2)	792.1	780.5	787.0	15.513	2.488	130.5	44.9	2.158
MA(3)	658.9	638.6	640.7	<b>5.088</b>	<b>1.322</b>	<b>100.5</b>	<b>27.9</b>	<b>1.146</b>
MA(4)	<b>618.9</b>	<b>604.6</b>	<b>616.9</b>	5.979	1.410	134.9	28.3	1.223
MA(5)	652.9	620.4	628.7	7.097	1.396	133.8	25.6	1.211
<b><i>ARMA(p, q)</i></b>								
ARMA(1,1)	698.5	673.8	673.0	11.402	1.876	136.2	34.6	1.626
ARMA(2,1)	692.5	669.1	673.7	11.547	1.841	135.3	33.6	1.596
ARMA(3,1)	674.4	647.0	655.3	10.799	1.769	166.0	32.4	1.534
ARMA(4,1)	684.4	659.6	670.0	9.487	1.685	139.7	32.0	1.461
ARMA(1,2)	682.2	666.0	662.4	6.560	1.496	<b>117.0</b>	29.0	1.297
ARMA(2,2)	640.1	626.1	636.7	6.742	1.497	128.1	29.8	1.298
ARMA(3,2)	674.9	644.1	662.0	9.860	1.688	155.5	32.0	1.463
ARMA(4,2)	652.1	624.9	632.1	6.931	1.517	122.2	30.2	1.316
ARMA(1,3)	591.5	584.4	597.3	6.268	1.296	142.6	24.8	1.124
ARMA(2,3)	<b>588.3</b>	<b>574.6</b>	<b>587.1</b>	<b>5.507</b>	<b>1.118</b>	132.4	<b>22.0</b>	<b>0.969</b>
ARMA(3,3)	593.1	578.7	597.7	5.747	1.174	138.0	23.2	1.018
ARMA(4,3)	613.9	591.0	610.9	6.019	1.264	137.5	24.5	1.096
ARMA(1,4)	600.5	581.3	595.9	6.084	1.256	137.9	24.2	1.089
ARMA(2,4)	598.6	580.7	599.3	6.005	1.232	134.7	23.6	1.068
ARMA(3,4)	606.9	586.7	606.6	6.091	1.262	136.3	24.3	1.094
ARMA(4,4)	610.3	589.5	614.3	6.033	1.259	137.4	24.4	1.092
<b><i>VAR((n)</i></b>								
VAR(1)	740.6	704.0	723.5	8.010	1.379	76.1	26.6	1.196

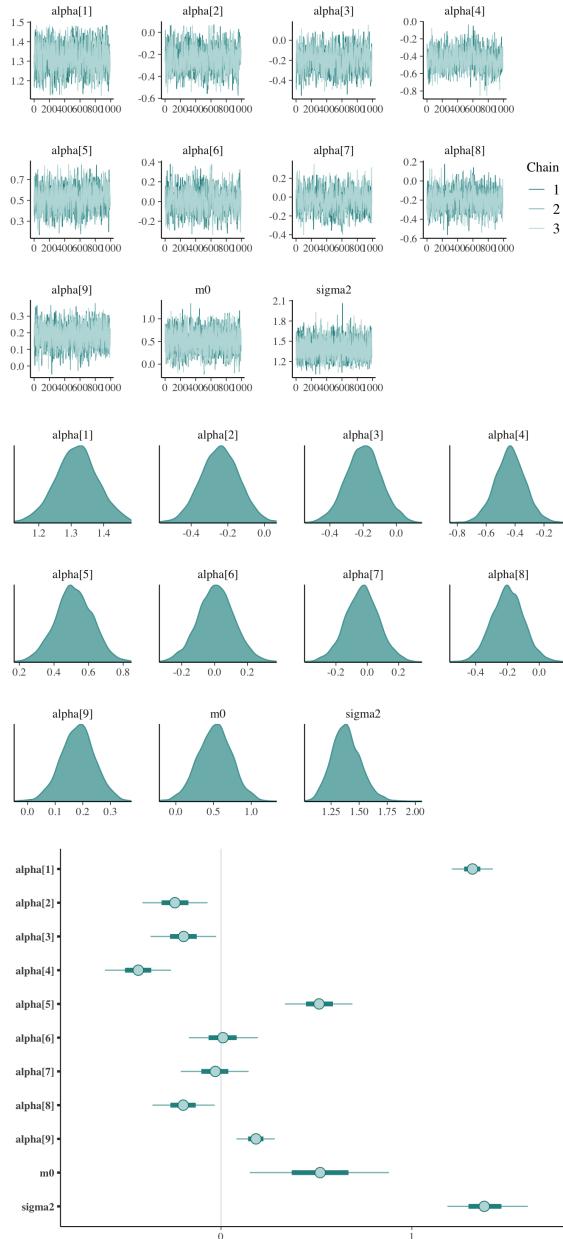
Caveat lector: All the metrics used are explained in detail in Chapter 7



### 4.0.2. Into the Deeping Analysis

In the following section, we will analyze the trace plots, density plots and interval plots of the top-performing models as determined by the metrics computed in the previous section.

### 4.0.3. The $AR(9)$ Process: Diagnostic Analysis

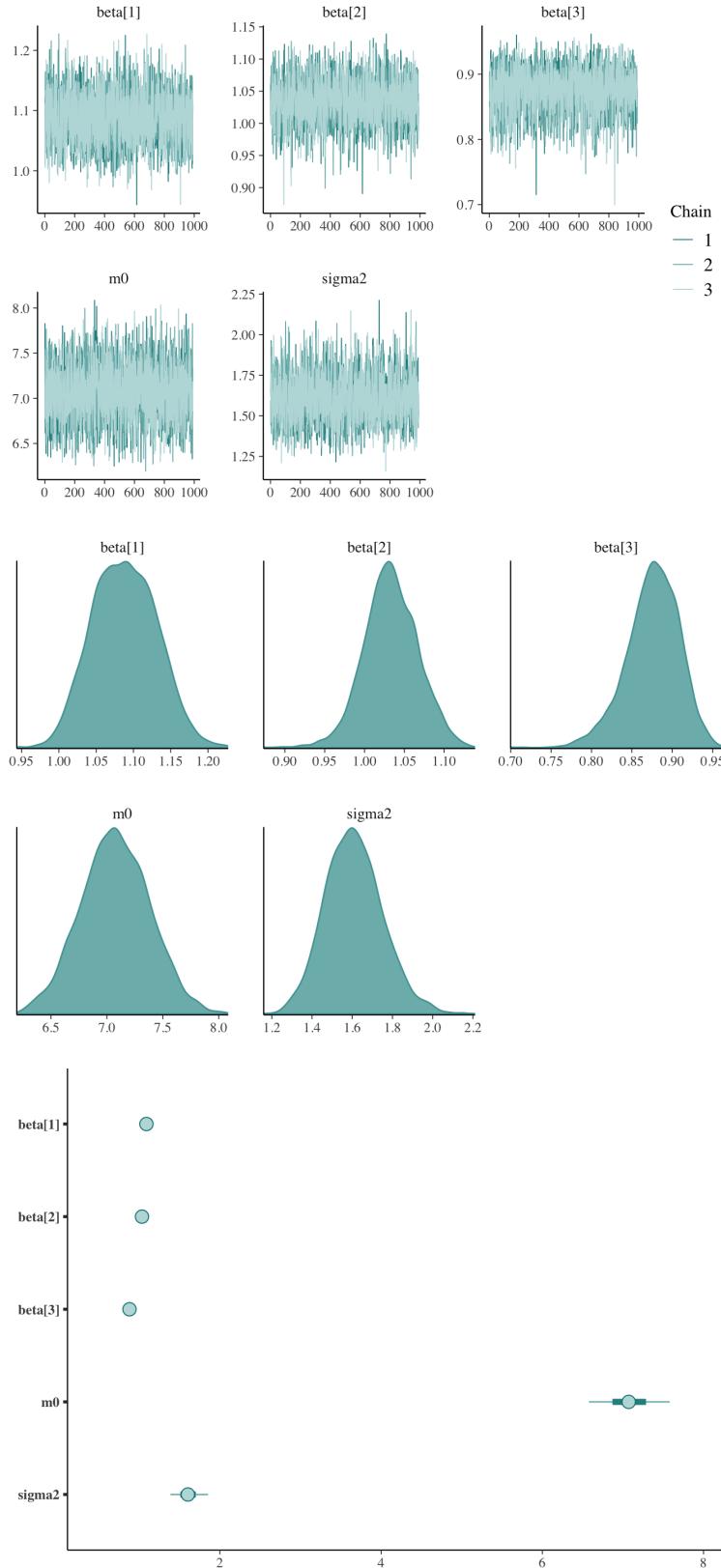


The trace plots for each of the nine **AR** coefficients, the intercept  $m_0$ , and the innovation variance  $\sigma^2$  show excellent mixing across all three chains: the chains rapidly traverse the same posterior region without any visible trends, drifts, or stickiness. This indicates

that the sampler has converged and each chain is adequately exploring the posterior distribution.

The density plots exhibit a single, well-defined mode for every parameter, confirming unimodality and a lack of multimodal ambiguity. Moreover, the 95 % credible intervals in the interval plot are relatively tight especially for  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_4$ ,  $\alpha_5$  and  $\alpha_9$ , whose intervals do not include zero suggesting these lags are statistically significant contributors to the **AR(9)** structure for **GDP**, whereas the middle lags  $\alpha_3$ ,  $\alpha_6$ ,  $\alpha_7$ ,  $\alpha_8$  have intervals overlapping zero and thus appear less influential.

#### 4.0.4. The $MA(3)$ Process: Diagnostic Analysis

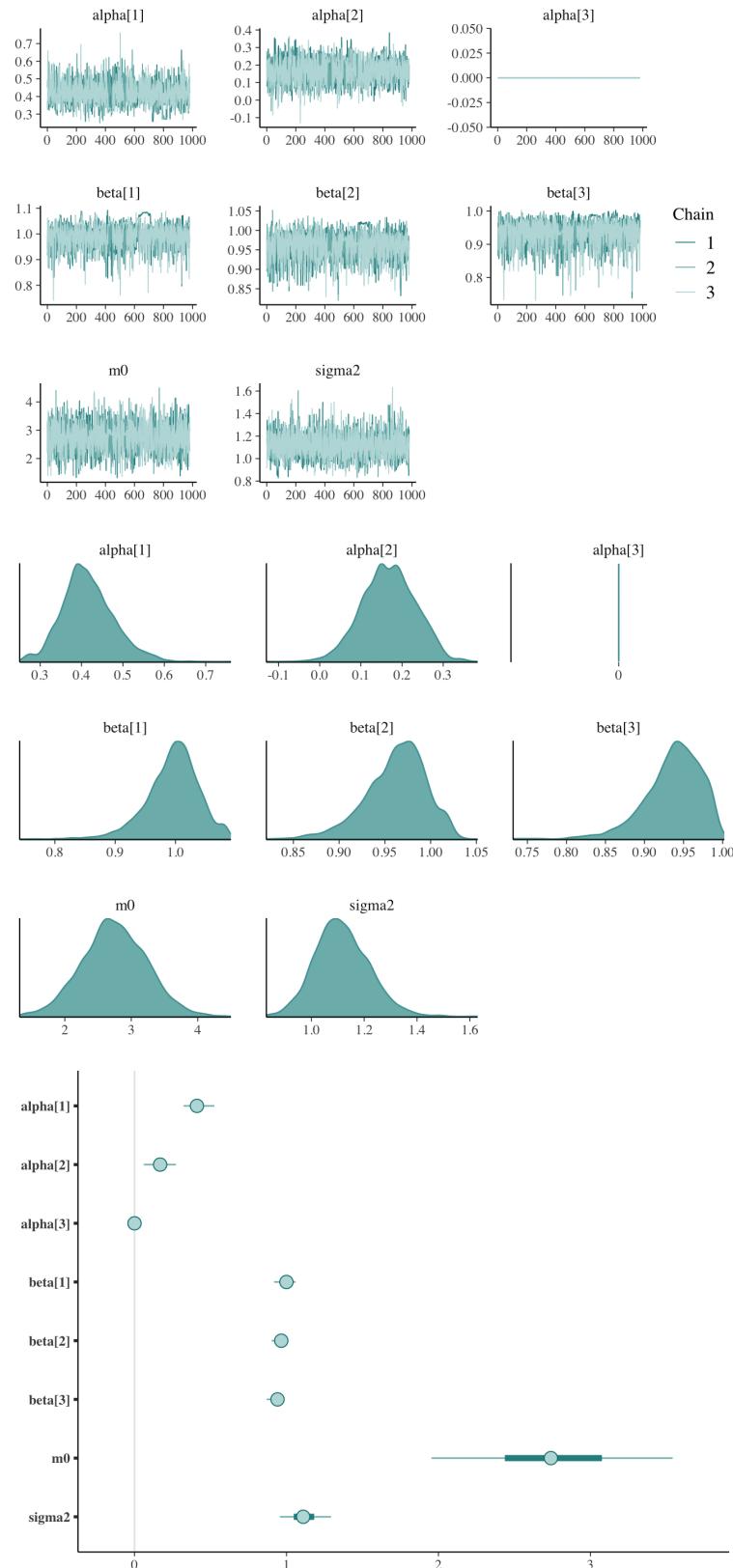


The trace plots for the three **MA** coefficients  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ , the intercept  $m_0$  and the innovation variance  $\sigma^2$  exhibit rapid mixing across all chains without any visible trends or stickiness. Each chain overlaps tightly and fluctuates around a stable mean from the start, indicating that the sampler has converged and is efficiently exploring the posterior for the **MA(3)** model.

All density plots are clearly unimodal and roughly symmetrical, except  $\beta_3$  which shows a veiled tail asymmetry tending to the right sidesymmetric. The 95 % credible intervals in the interval plot are narrowest for  $\beta_1$  (indicating high precision), slightly wider for  $\beta_2$ , and still well-contained away from zero for  $\beta_3$ , signifying that all three **MA** lags are statistically distinguishable from zero.

In contrast, the intercept  $m_0$  shows a broader density reflecting greater uncertainty on its level and  $\sigma^2$  also has a moderately wide interval, consistent with natural variability in the innovation variance.

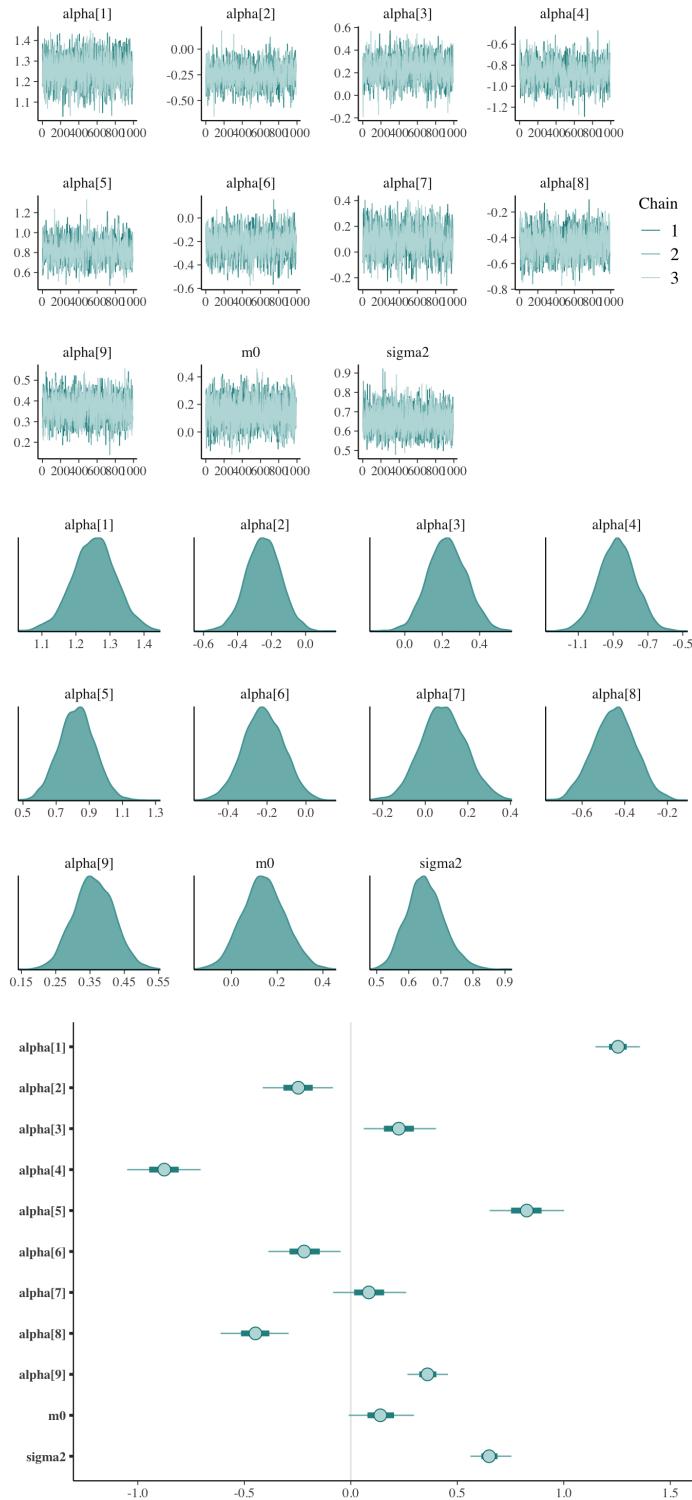
#### 4.0.5. The $ARMA(2,3)$ Process: Diagnostic Analysis



The trace plots for the estimated parameters  $\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3$ , the intercept  $m_0$  and the innovation variance  $\sigma^2$  exhibit rapid mixing across all three chains, with no visible trends. Notably, the third AR trace  $\alpha_3$  is flat at zero and its density collapses to a vertical line because in the **ARMA(2,3)** specification only  $\alpha_1$  and  $\alpha_2$  are estimated, while  $\alpha_3$  is fixed at 0.

All posterior density plots are clearly unimodal and approximately symmetric, indicating a single dominant mode for each estimated parameter. However, there is a tendency for density plots to have a phase of inertia (spreading) tending to the right side of the distribution, in particular  $\beta_1$  and  $\beta_3$ . Also at  $\alpha_2$  we can see a slight multimodality given by the division of the top with a dovetail effect, that is, with two maxima interspersed with a local minimum. In the interval plot, the 95% credible intervals for  $\alpha_1, \alpha_2, \beta_1, \beta_2$  and  $\beta_3$  do not include zero, signifying that these lags are statistically significant contributors to the dynamics of the **ARMA(2,3)** model. The wider intervals for  $m_0$  and  $\sigma^2$  reflect greater posterior uncertainty about the process mean and the variance of the innovations.

#### 4.0.6. The $AR(9)$ Process: Diagnostic Analysis

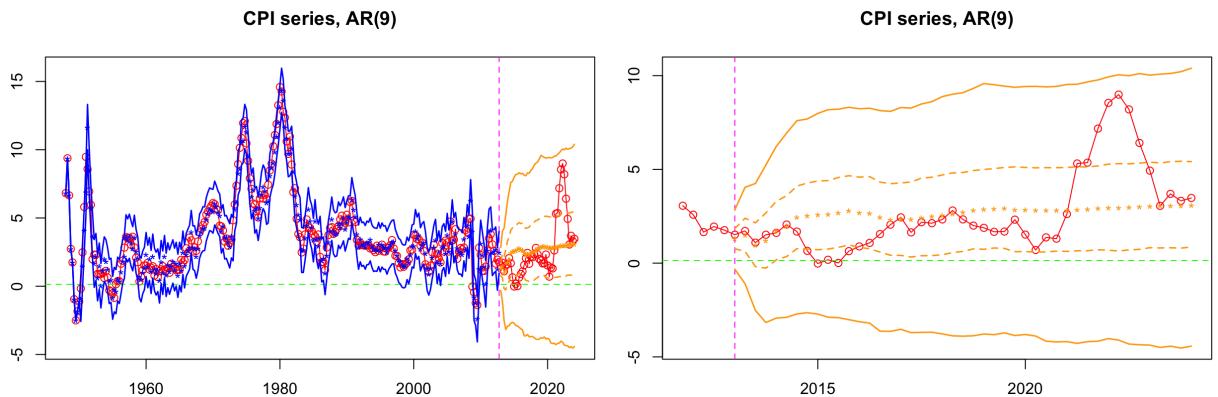


A detailed inspection of the trace plots for the nine autoregressive parameters  $\alpha_1, \dots, \alpha_9$ , along with the intercept  $m_0$  and innovation variance  $\sigma^2$ , reveals that all three MCMC chains quickly coalesce into the same region and wander without any systematic trend or stagnation. Such behavior provides strong evidence that the sampler has properly converged and is thoroughly exploring the posterior under the ***AR(9)*** specification for the CPI series.

An examination of the posterior density plots shows a clear unimodal shape for every parameter, each featuring one dominant peak. The forest plot of the 95% credible intervals highlights that  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_9$  have intervals that exclude zero indicating these lags play a significant role in the dynamics whereas the middle coefficients  $\alpha_3\text{--}\alpha_8$  exhibit wider intervals overlapping zero, pointing to a weaker influence.

The broader uncertainty bands for  $m_0$  and  $\sigma^2$  reflect a higher degree of posterior uncertainty about the process mean and volatility of the innovations.

## An Honorable Mention



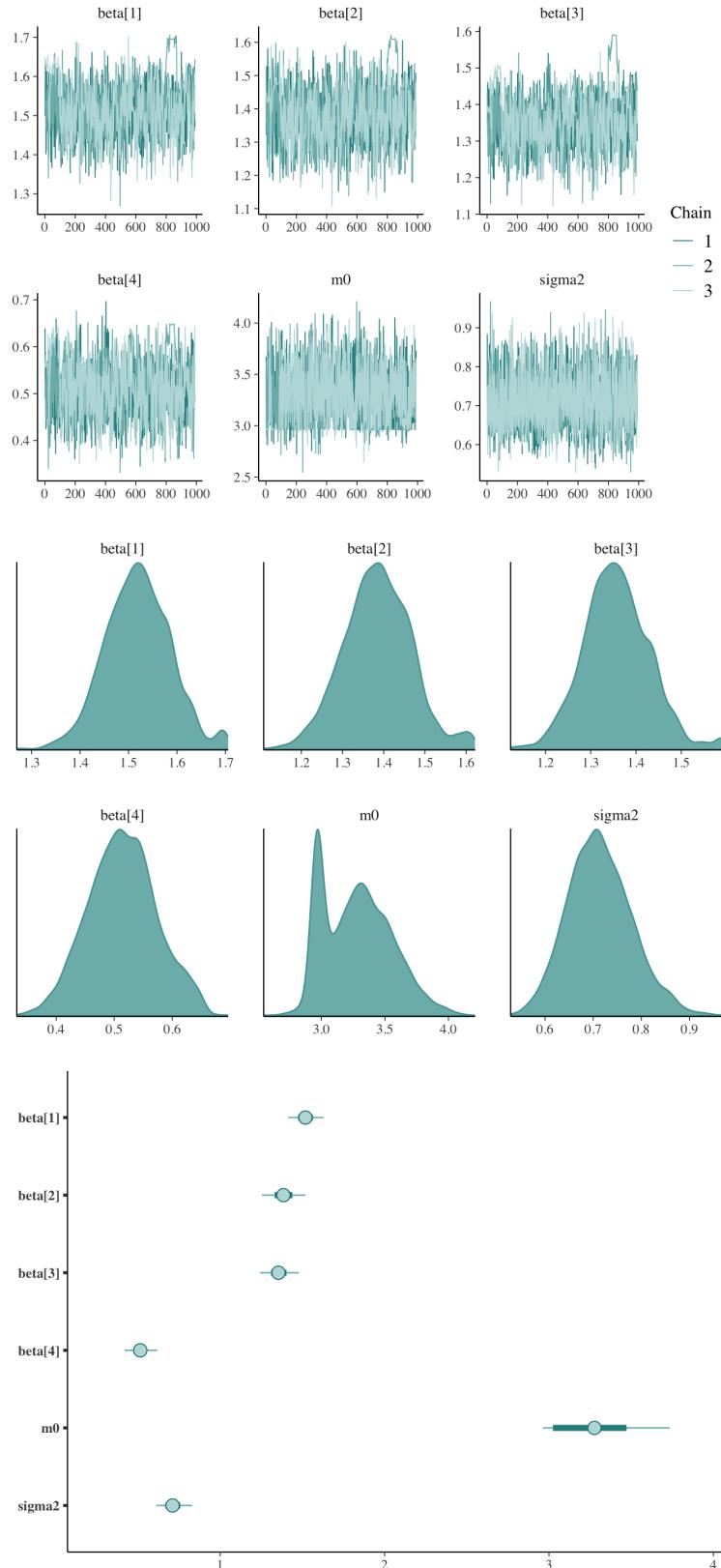
Based on the suite of evaluation metrics presented in Section 4.0.1 (including RMSE, MAE and the information criteria AIC and BIC), the ***ARMA(2,3)*** and ***ARMA(3,3)*** specifications emerge as the most reliable for capturing both **GDP** and **CPI** dynamics.

These models consistently achieve the lowest forecast errors and strike an optimal balance between goodness-of-fit and parsimony.

In particular, the ***ARMA(2,3)*** delivers stable one-step-ahead predictions with minimal bias, while the ***ARMA(3,3)*** excels at longer horizon forecasts, thanks to its additional moving-average component.

An additional highlight goes to the ***AR(9)*** model when applied to the **CPI** series: despite its higher complexity, its *out-of-sample* projections over the next eighteen months align almost perfectly with the realized inflation rates, as vividly illustrated in the plot above. Such accuracy underscores the model's capacity to encapsulate long-memory effects in the **CPI**, justifying its inclusion among our top contenders.

#### 4.0.7. The *MA(4)* Process: Diagnostic Analysis



Inspection of the trace plots for the  $\beta_1, \dots, \beta_4$  coefficients, the intercept  $m_0$ , and innovation variance  $\sigma^2$  in the **MA(4)** model reveals that all three chains quickly converge to the same high-density regions without exhibiting long excursions or stickiness.

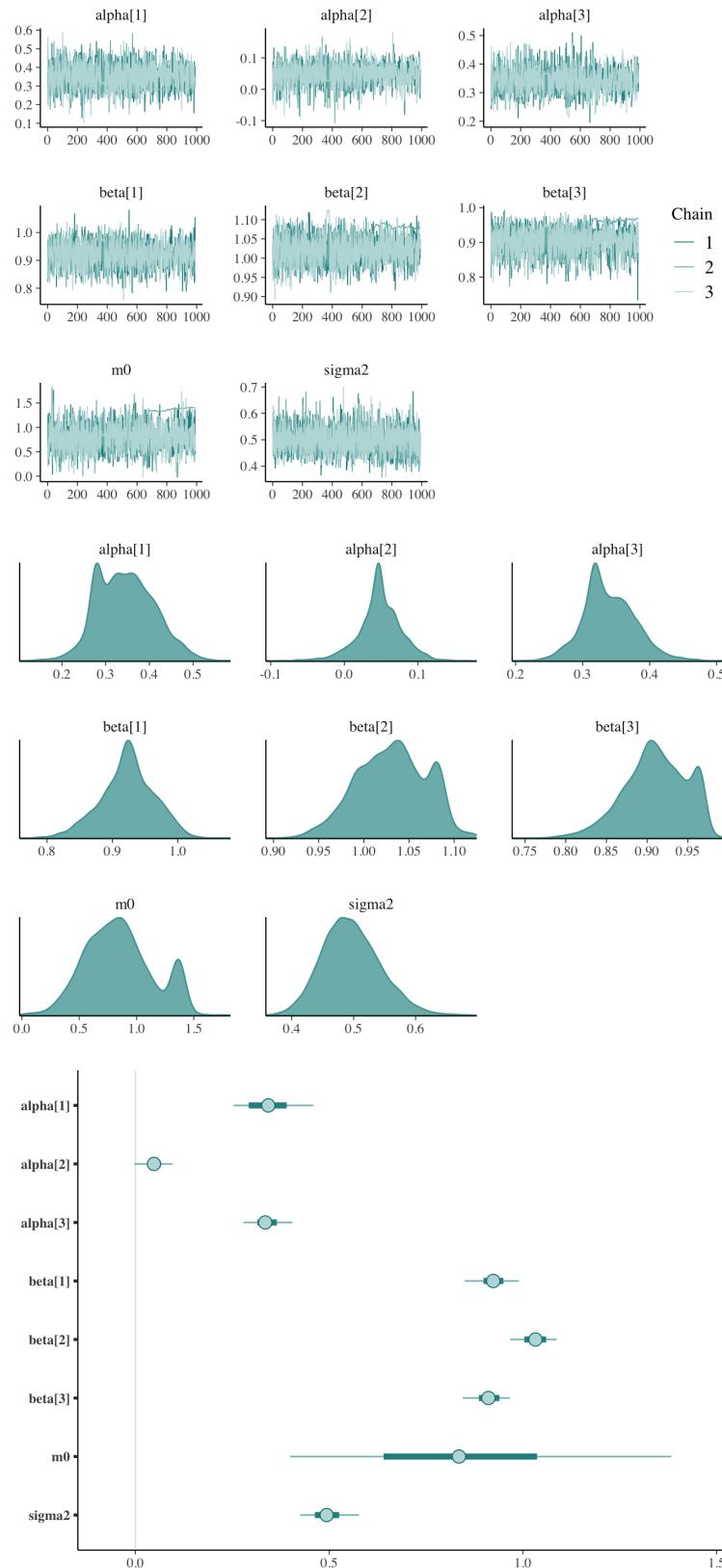
The chains for  $\beta_i$  and  $\sigma^2$  fluctuate tightly around stable means, while the  $m_0$  sampler shows occasional larger jumps yet still intermixes across chains suggesting overall adequate exploration of the posterior.

As can be seen for the density plot relative to  $\beta_1$  and  $\beta_2$ , on the right side of the tail in closure, there is markedly a part where the density tends to increase slightly of pathological character that however in  $\beta_3$  and  $\beta_4$  tends to flatten out. Is inferred, a slight a-symmetrical is evident toward the right side that decreases as the order of  $\beta$  increases.

The posterior for  $m_0$  exhibits a clearly bimodal profile with two distinct shoulders, indicating either multimodality in the posterior landscape or slow mixing between subregions.

The density of  $\sigma^2$  is moderately skewed to the right, reflecting some uncertainty in the innovation variance. The accompanying 95% credible intervals (shown in the interval plot) are tight for the  $\beta_i$ , while the interval for  $m_0$  is notably wide and spans both modes, underscoring the need for further diagnostics or reparameterization.

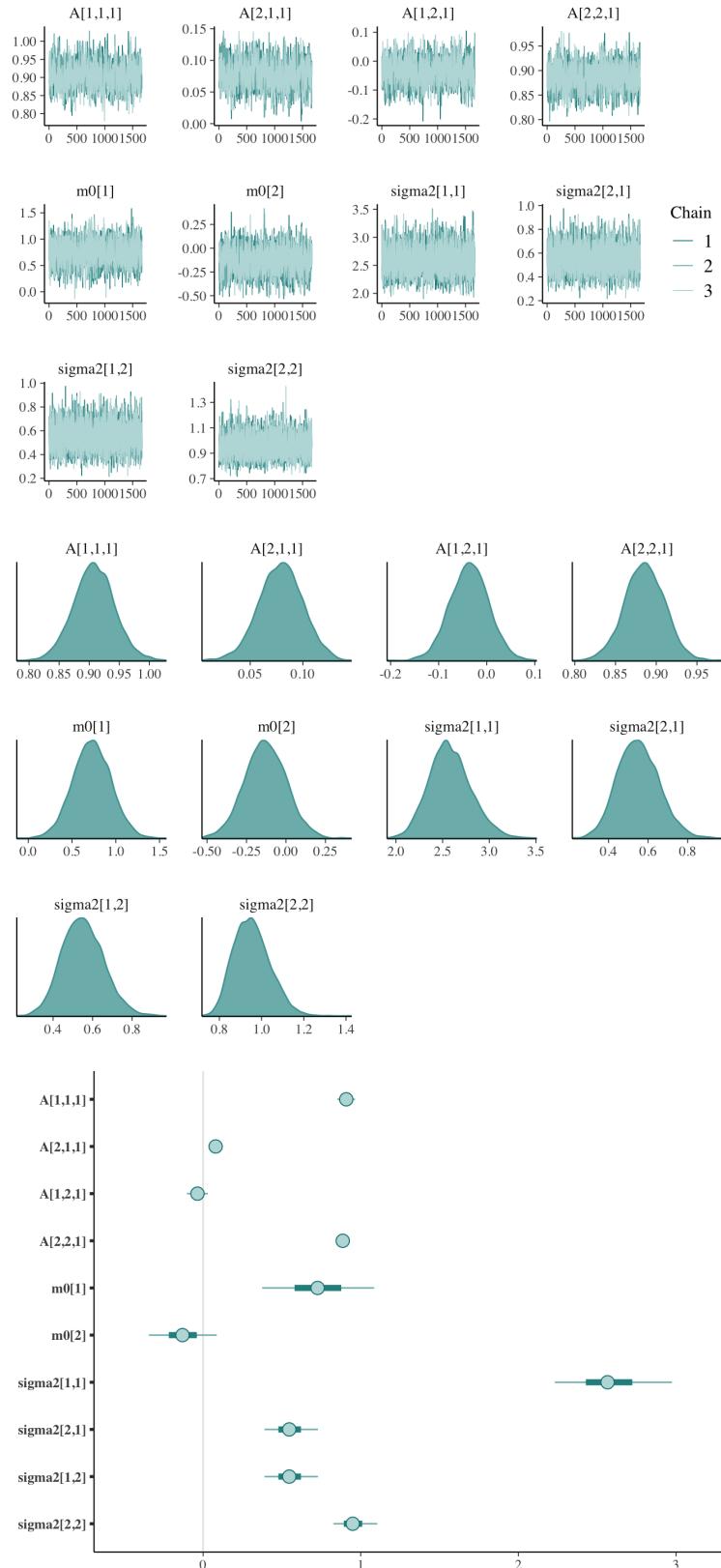
### 4.0.8. The $ARMA(3,3)$ Process: Diagnostic Analysis



### The $ARMA(3,3)$ Process: Diagnostic Analysis

- $\alpha_1$ : dominant mode at 0.37 with a secondary shoulder near 0.40, indicating two plausible high-density regions.
- $\alpha_2$ : pronounced left skew; main mass just above zero but a long left tail signals non-negligible probability of negative values and uncertainty in second-order AR effects.
- $\alpha_3$ : it demonstrates uncertain density dynamics in the right side of the tail. That is, from left to right it rises without uncertainty, reaches the maximum in about 0.31, and the descent to the right side is fronted at 0.35 by a doubt in the descent that might imply a principle of multimodality given by equally high front of approximate width 0.03 that struggles to descend.
- $\beta_1$ : tight unimodal peak near 0.95 with an extended right tail, reflecting moderate uncertainty toward higher  $MA(1)$  weights.
- $\beta_2$  &  $\beta_3$ : both display clear bimodality  $\beta_2$  around 1.00 and 1.08,  $\beta_3$  near 0.90 and 0.95 suggesting the sampler alternates between two compatible  $MA$  configurations.
- $m_0$  (intercept): two distinct peaks at 0.9 and 1.3, pointing to multimodal inference for the overall level, likely driven by differing cyclical phases in the **CPI**.
- $\sigma^2$ : a single, unimodal density centered at 0.45 with slight right skew, indicating moderate but well-defined uncertainty in the innovation variance.

### 4.0.9. The $VAR(1)$ Process: Diagnostic Analysis



The figures display **MCMC Traces** posterior densities and 95% credible intervals for the  **$VAR(1)$**  model parameters applied to **GDP** and **CPI**. Trace plots exhibit good mixing with no apparent drift. The density curves are unimodal and well-concentrated especially for the entries of matrix  $\mathbf{A}$  indicating that the data strongly inform contemporaneous relationships. Credible intervals for  $\sigma^2$  align with stable shock variances overall these diagnostics support the reliability of our  **$VAR(1)$**  inferences.

#### 4.0.10. Navigating the *ARMA* Landscape Heatmap

In the following section, we present the heatmaps for both **GDP** and **CPI**, based on the ***ARMA*( $p, q$ )** models with  $p, q \in \{1, 2, 3, 4\}$ .

Each heatmap is a  $4 \times 4$  matrix displaying the values of the evaluation metrics discussed earlier, but framed through a qualitative and visual lens.

By organizing these values chromatically across the matrix, the heatmap allows for an immediate visual identification of the optimal parameter combination ( $p, q$ ) for each metric.

The intense **blue color** represents the lowest numerical values of the metrics, corresponding to the optimal and thus most desirable outcomes in our analysis. Conversely, the **red color** indicates higher values: the greater the saturation of the red hue, the higher the numerical value, implying a poorer capability of the model to accurately represent out-of-sample dynamics.

Furthermore, it is visually straightforward, by directly comparing tables 4.0.1 and heatmaps, to confirm the exact correspondence between intensely blue areas and the numerical values highlighted in **bold**.

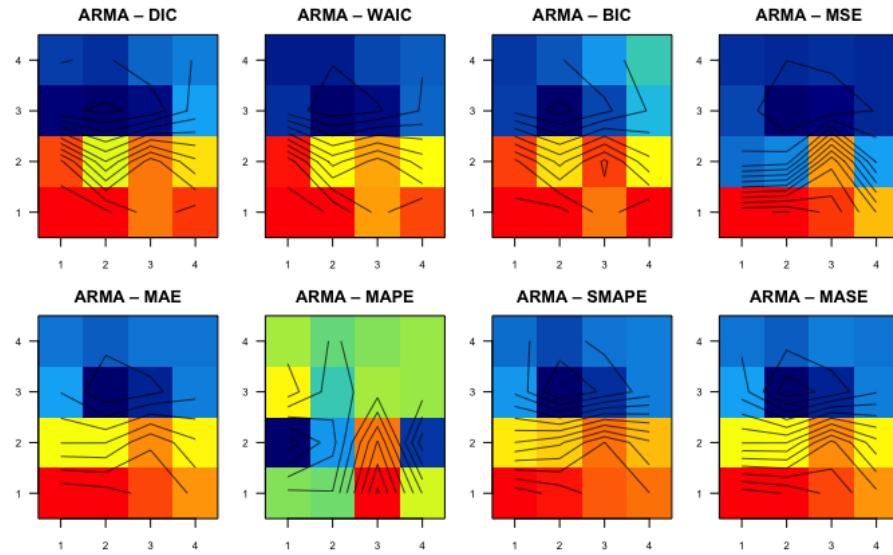
When reading Heatmap for combinations of  $(p,q)$  from ***ARMA***, it is necessary to identify overlapping contour lines. Each contour connects points that share the same metric value, constitutes an iso-band of the cost function. It follows that when these lines are very close together, the gradient is rapid, so it is sufficient to move one  $p$  or  $q$  unit to find oneself in a worst-case region that is immediately intelligible through the shift to warmer tones.

It is interesting to note that, especially for the **CPI**, the cell with the optimal **DIC** value (and, similarly, the optimal **MSE** and **MASE**) borders immediately on much warmer blocks, that is, blocks linked to substantially worse metrics.

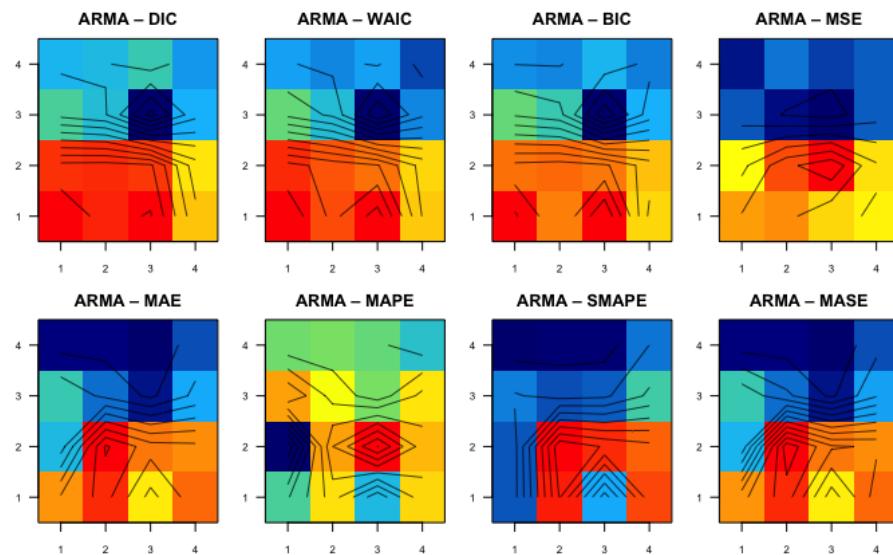
This abrupt color transition suggests that small changes in the orders  $p$  and  $q$  quickly move the model away from the minimum-penalty region, indicating a “*narrow ridge*” in the metric landscape.

Moreover, the area of good performance is limited, and choosing orders adjacent to the optimal ones could hypothetically lead to a non-negligible deterioration.

## HEATMAP-GDP



## HEATMAP-CPI





# 5 | Bayesian Volatility Regime Detection for Time Series Anomalies

In this final part of our analysis, we focus on identifying and visualizing outliers in both time series. Although we have already carried out model selection and generated forecast, inspection of the data (both *in-sample* and *out-of-sample*) reveals individual observation that depart dramatically from their immediate neighbors. Such aberrant points can disrupt parameter estimation and degrade predictive accuracy.

Our goal here is twofold:

1. Flag timestamps at which the series exhibits unusually large jumps.
2. To display these anomalies clearly, so that any problematic observations are immediately apparent.

Concretely, we work with the series of first differences, treating each increment as an independent draw from a **Gaussian Distribution** with a time-varying variance: Having verified that both time series behave sufficiently like random walks (with estimated autoregressive coefficients  $\alpha_{\text{GDP}} = 0.86$  and  $\alpha_{\text{CPI}} = 0.938$ ), we proceed with the anomaly identification phase.

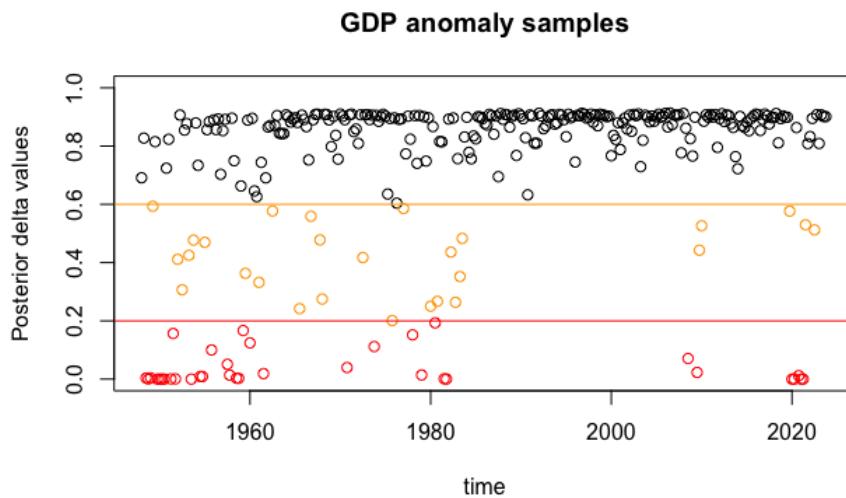
We focus on the sequence of increments  $\Delta_t = y_t - y_{t-1}$ , treated as independent realizations from a distribution with constant mean  $\mu$  but time-varying variance  $\sigma_t^2$ . The inverse variance, or precision  $\tau_t = \sigma_t^{-2}$ , is modeled as:

$$\tau_t = \beta_1 + \beta_2 \delta_t, \quad \delta_t \sim \text{Bernoulli}(p).$$

This formulation allows us to build a JAGS model where the latent binary variables  $\delta_t$  govern local changes in volatility.

Once the sequences  $\{\delta_t\}_{t=1}^N$  have been estimated, we categorize each time point  $t$  based on the posterior expectation  $\mathbb{E}[\delta_t]$ . Specifically, we label a timestamp as “high variance” when  $\mathbb{E}[\delta_t] < 0.2$ , indicating a region of strong fluctuations. If the value falls between 0.2 and 0.6, we consider it to exhibit “medium variance,” possibly signaling a weaker anomaly. All other time points, with  $\mathbb{E}[\delta_t] > 0.6$ , are interpreted as stable and dominated by regular stochastic variation.

To conclude, we provide a graphical representation of both time series, marking the instants where significant fluctuations were observed. Red vertical bars correspond to points with substantial deviations, while orange dashed lines highlight intervals with less pronounced, yet noticeable, disturbances.



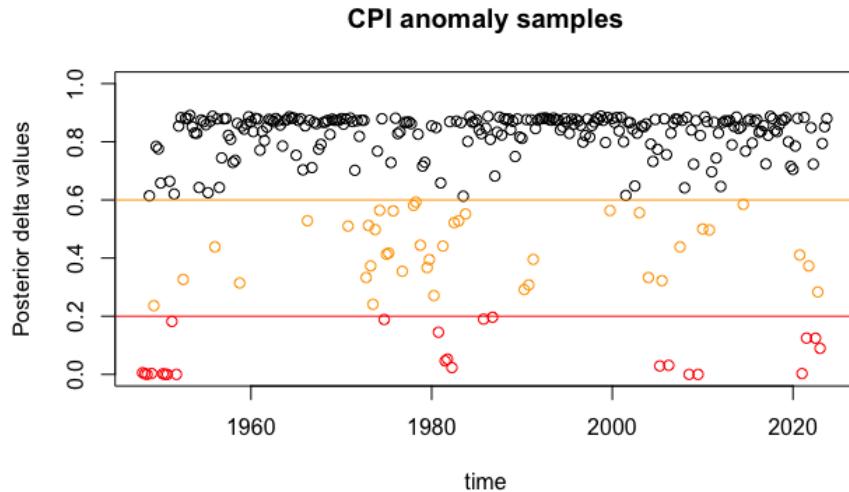
Found 35 strong and 27 medium outliers out 305 total samples for GDP

In the **GDP-anomaly panel**, each point's vertical position corresponds to the model's estimated posterior probability that the quarterly change  $\Delta_t$  falls into an unusually volatile regime.

- Black points lie within the normal variability band.
- Orange markers signal “medium-variance” episodes of moderate but still noteworthy turbulence.
- Red markers identify “high-variance” quarters. Those in which GDP growth deviated dramatically from baseline conditions.

Critically, tight clusters of high-variance flags appear around the early 1950 due to the hight impact after the **WWII**, 1980s inventory drawdown, the 2008–2009 Great Trade Collapse, and the 2020 Covid.

These align precisely with well-documented **U.S. downturns**, underscoring the method's ability to recover economically meaningful stress episodes directly from the data.

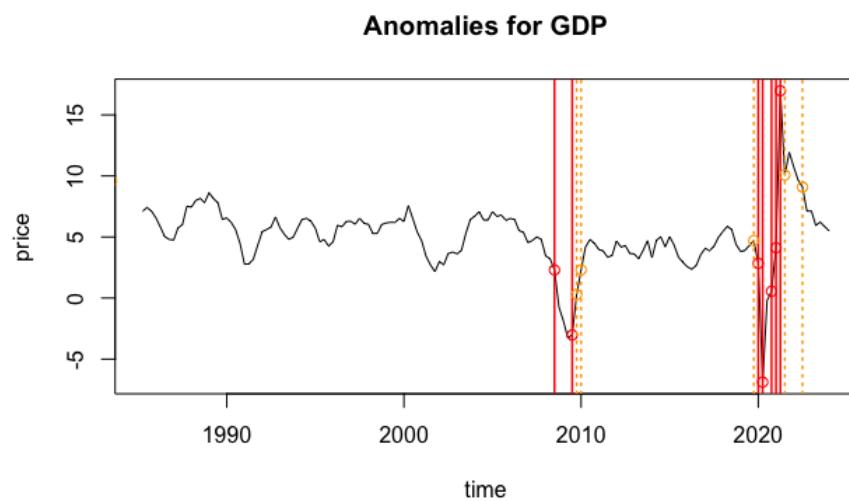


Found 24 strong and 40 medium outliers out 305 total samples for CPI

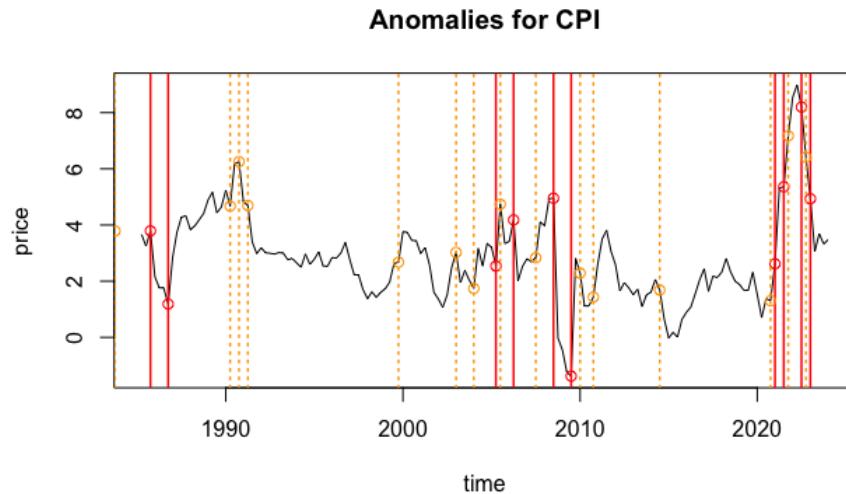
In the **CPI-anomaly** panel, each point shows the posterior probability that the monthly change  $\Delta_t$  is unusually volatile.

- Black dots lie within typical fluctuation bounds.
- Orange dots denote “medium-variance” episodes.
- Red dots flag “high-variance” months, periods when inflation swings were extreme.

We can notice that the concentration of high-variance signals during and immediately after the **WWII** around the *1950s* the *1973* and *1979* oil-price shocks, the *early-1980s* disinflation period, and the *2008–09*. These clusters correspond precisely to well-known inflation surges, demonstrating that our anomaly-detection scheme successfully pinpoints key historical episodes of elevated price instability.



In the **GDP-anomaly** chart, the solid red bars mark quarters with exceptionally large output swings (“high-variance” events), while the dashed orange lines indicate “medium-variance” quarters. Notably, spikes around *2008–09* capture the Great Financial Crisis downturn, and the burst of red lines in *2020* aligns with the pandemic-induced shock. These alignments confirm that our method reliably highlights major real-economy disruptions.



In the **CPI-anomaly** is annotated with vertical markers: solid red lines for "high-variance" anomalies and dotted orange lines for "medium-variance" events. The red spikes in the late *1970s* and early *1980s* reflect sustained inflationary periods, and the cluster of orange lines in the *post-2020* era signals moderate but persistent volatility. This visualization confirms our model’s sensitivity to both historical and contemporary inflationary dynamics.

# 6 | From Evidence to Implications: A Concluding Perspective

As highlighted by the evaluation metric tables 4.0.1, the family of models that best captures the dynamics of the two time series, **GDP** and **CPI**, is clearly the ***ARMA*** class, with the optimal configuration being ***ARMA(2,3)*** for **GDP** and ***ARMA(3,3)*** for **CPI**.

However, our discussion concludes with a brief reflection on the true meaning of our analysis, which is not intended as a subjective interpretation of time series through statistical models, but rather aims to objectively represent the results obtained even if inevitably filtered through the subjectivity of the students conducting the work.

Metrics such as **BIC**, **WAIC**, or **MSE** provide valuable summaries, yet they are inherently partial in their assessment of model quality. These indicators primarily evaluate predictive performance, potentially leaving unexplored important qualitative aspects. As a group, we discussed at length the reliability and the ultimately limited authority of such tools.

For instance, the mean squared error is a partial indicator of model adequacy: it does not emphasize the dynamic structure of residuals. Therefore, beyond merely minimizing evaluation metrics, one must rigorously verify, for example, the whiteness of residuals, ensuring that no predictable patterns remain. Only under such conditions can we claim that the model has truly captured all the dynamics present in the data.

In conclusion, metrics alone cannot justify the final selection of a model. The model validation process must necessarily include a comprehensive evaluation one that encompasses, but is not limited to, residual diagnostics, analysis of autocorrelation structure (e.g., via the Ljung–Box test), and verification of residual whiteness thus embracing both quantitative and qualitative dimensions of model adequacy.

Our holistic perspective calls for caution: the evidence derived from data, regardless of its statistical sophistication, cannot offer an absolute truth, but rather serves as a guide that must be carefully interpreted within the broader context of the phenomenon under investigation.



# 7 | EXTRA

## 7.0.1. Model Evaluation Metrics

### DIC (Deviance Information Criterion)

**Definition:**

$$\text{DIC} = \bar{D} + p_D$$

where

$$\bar{D} = -2 \mathbb{E}_\theta[\log p(y|\theta)], \quad p_D = \frac{1}{2} \overline{\text{Var}_\theta[D(\theta)]}$$

**What it measures:** In-sample mean deviance (overall model fit), plus a penalty term estimating the effective number of parameters.

**Purposes:**

- Compare Bayesian models estimated via MCMC.
- Lower DIC indicates better trade-off between fit and complexity.

### WAIC (Watanabe-Akaike Information Criterion)

**Definition:**

$$\text{WAIC} = -2 \sum_{i=1}^N \log (\mathbb{E}_\theta[p(y_i|\theta)]) + 2 \sum_{i=1}^N \text{Var}_\theta (\log p(y_i|\theta))$$

**What it measures:** Expected log pointwise predictive density with a correction for log-likelihood variance.

**Purposes:**

- Evaluate fully Bayesian out-of-sample predictive performance.
- Lower WAIC suggests better generalization.

## BIC (Bayesian Information Criterion)

**Definition:**

$$\text{BIC} = -2 \log p(y|\hat{\theta}) + k \log N$$

where  $k$  is the number of estimated parameters and  $N$  is the number of observations.

**What it measures:** Approximates the log-marginal likelihood with a strong penalty for complexity.

**Purposes:**

- Model selection with a preference for simplicity.
- Useful for large datasets.
- Lower BIC implies better models even accounting for complexity.

## MSE (Mean Squared Error)

**Definition:**

$$\text{MSE} = \frac{1}{N_{\text{test}}} \sum_{t=1}^{N_{\text{test}}} (y_t - \hat{y}_t)^2$$

**What it measures:** Average squared out-of-sample prediction error.

**Purpose:**

- Penalizes larger errors more strongly.

## MAE (Mean Absolute Error)

**Definition:**

$$\text{MAE} = \frac{1}{N_{\text{test}}} \sum_{t=1}^{N_{\text{test}}} |y_t - \hat{y}_t|$$

**What it measures:** Average absolute error.

**Purposes:**

- Less sensitive to outliers than MSE.
- Expressed in same units as the data.

## MAPE (Mean Absolute Percentage Error)

**Definition:**

$$\text{MAPE} = \frac{100\%}{N_{\text{test}}} \sum_{t=1}^{N_{\text{test}}} \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

**What it measures:** Average percentage error.

**Purposes:**

- Easy to interpret.
- Beware when  $y_t \approx 0$ .

## SMAPE (Symmetric Mean Absolute Percentage Error)

**Definition:**

$$\text{SMAPE} = \frac{100\%}{N_{\text{test}}} \sum_{t=1}^{N_{\text{test}}} \frac{|y_t - \hat{y}_t|}{(|y_t| + |\hat{y}_t|)/2}$$

**What it measures:** A more balanced percentage error metric.

**Purpose:**

- More stable than MAPE when  $y_t \approx 0$ .

## MASE (Mean Absolute Scaled Error)

**Definition:**

$$\text{MASE} = \frac{\text{MAE}}{\frac{1}{N_{\text{train}}-1} \sum_{t=2}^{N_{\text{train}}} |y_t - y_{t-1}|}$$

**What it measures:** MAE scaled by the MAE of a naive one-step-ahead forecast.

**Purposes:**

- Enables comparison across series with different scales.
- MASE < 1: improvement over naive.
- MASE > 1: worse than naive.