

Forecasting Under Structural Breaks: A Monte Carlo Study

Research Module in Econometrics and Statistics
Fundamentals of Monte Carlo Simulations in Data Science

Aadya Khatavkar
University of Bonn
`s6aakhat@uni-bonn.de`

Winter Semester 2025/26

Abstract

This study evaluates the forecasting performance of various time series methods in the presence of structural breaks in the mean. Using Monte Carlo simulations, we compare global AR(1), rolling-window AR(1), oracle break-dummy specifications, estimated break detection, and Markov switching models. Our simulation design generates AR(1) processes with single and multiple deterministic breaks in the intercept, and we assess each method's one-step-ahead forecast accuracy using RMSE, MAE, and bias. Results indicate that oracle break-dummy models achieve the best performance, reflecting the value of correctly incorporating structural change. Among practical methods where break dates are unknown, rolling-window estimation offers improvements over global fitting by adapting to recent data. These findings have implications for practitioners choosing forecasting strategies under parameter instability.

Keywords: Structural breaks, Monte Carlo simulation, time series forecasting, ARIMA, rolling window estimation, regime switching

Contents

1	Introduction	3
2	Data-Generating Processes	3
2.1	Single-Break DGP	3
2.2	Multiple-Break DGP	4
3	Forecasting Methods	4
3.1	Global AR(1)	4
3.2	Rolling-Window AR(1)	4
3.3	AR(1) + Break Dummy (Oracle)	5
3.3.1	Single-Break Dummy Model	5
3.3.2	Multiple-Break Dummy Model	5
3.4	AR(1) + Estimated Break (Grid Search)	5
3.5	Markov Switching (Two Regimes)	6
4	Monte Carlo Design	6
5	Evaluation Metrics	7
5.1	Bias	7
5.2	Mean Absolute Error (MAE)	7
5.3	Root Mean Squared Error (RMSE)	8
6	Results and Interpretation	8
6.1	Single-Break Case	8
6.2	Multiple-Break Case	8
6.3	Markov Switching Results	9
7	Conclusion	9
7.1	Implications for Practice	9
7.2	Future Work	10

1 Introduction

Time series forecasting is central to economic and financial decision-making. A fundamental challenge arises when the underlying data-generating process (DGP) is subject to *structural breaks*—discrete changes in model parameters that occur at specific points in time. Ignoring such breaks can lead to severely biased forecasts, as models trained on pooled data fail to reflect the current regime.

This study investigates the forecasting performance of several methods under structural breaks in the mean of an AR(1) process. The research questions are:

1. How do different forecasting methods perform when the DGP exhibits structural breaks?
2. What is the cost of not knowing the break date(s) in terms of forecast accuracy?
3. Can adaptive methods (rolling windows, regime switching) provide a practical alternative to oracle specifications?

The remainder of this paper is organized as follows. Section 2 describes the data-generating processes. Section 3 presents the forecasting methods under comparison. Section 4 details the Monte Carlo design. Section 5 defines the evaluation metrics. Section 6 reports and interprets the simulation results. Section 7 concludes.

2 Data-Generating Processes

We consider AR(1) processes with deterministic breaks in the intercept (mean). The general model is:

$$y_t = \mu_t + \phi y_{t-1} + u_t, \quad u_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2), \quad (1)$$

where μ_t is a piecewise-constant intercept that changes at known break dates.

2.1 Single-Break DGP

Let T_b denote the break date. The intercept is:

$$\mu_t = \begin{cases} \mu_0 & \text{if } t \leq T_b, \\ \mu_1 & \text{if } t > T_b. \end{cases} \quad (2)$$

The simulation uses parameters $\phi = 0.6$, $\sigma = 1.0$, $\mu_0 = 0$, and $\mu_1 = 2$.

2.2 Multiple-Break DGP

For two breaks at b_1 and b_2 (with $b_1 < b_2$):

$$\mu_t = \begin{cases} \mu_0 & \text{if } t \leq b_1, \\ \mu_1 & \text{if } b_1 < t \leq b_2, \\ \mu_2 & \text{if } t > b_2. \end{cases} \quad (3)$$

The simulations use $\mu_0 = 0$, $\mu_1 = 2$, and $\mu_2 = -1$.

3 Forecasting Methods

We compare five forecasting methods, ranging from naive global fitting to oracle specifications that exploit knowledge of break dates.

3.1 Global AR(1)

The global AR(1) model estimates:

$$y_t = c + \phi y_{t-1} + u_t \quad (4)$$

using all available training observations $\{y_1, \dots, y_{t_0-1}\}$. The one-step forecast is:

$$\hat{y}_{t_0} = \hat{c} + \hat{\phi} y_{t_0-1}. \quad (5)$$

Limitation: After a break, the estimates \hat{c} and $\hat{\phi}$ are contaminated by pre-break observations, leading to biased forecasts.

3.2 Rolling-Window AR(1)

The rolling-window approach estimates the AR(1) model using only the most recent w observations:

$$\{y_{t_0-w}, \dots, y_{t_0-1}\}. \quad (6)$$

If the window lies entirely within the post-break regime, the estimator is consistent for the current parameters. The forecast is:

$$\hat{y}_{t_0} = \hat{c}_{\text{roll}} + \hat{\phi}_{\text{roll}} y_{t_0-1}. \quad (7)$$

Trade-off: Smaller windows adapt faster but increase estimation variance; larger windows reduce variance but may include pre-break data.

3.3 AR(1) + Break Dummy (Oracle)

This method explicitly models breaks via dummy variables. It is called *oracle* because it assumes the break date(s) are known (true in simulation, usually unknown in real data).

3.3.1 Single-Break Dummy Model

Define:

$$d_t = \mathbf{1}(t > T_b), \quad (8)$$

and estimate:

$$y_t = c + \phi y_{t-1} + \delta d_t + u_t. \quad (9)$$

The one-step forecast becomes:

$$\hat{y}_{t_0} = \hat{c} + \hat{\phi} y_{t_0-1} + \hat{\delta} d_{t_0}. \quad (10)$$

Interpretation: The intercept shifts by δ after the break, so the model can match the mean change.

3.3.2 Multiple-Break Dummy Model

Define two dummies:

$$d_{1,t} = \mathbf{1}(t > b_1), \quad d_{2,t} = \mathbf{1}(t > b_2), \quad (11)$$

and estimate:

$$y_t = c + \phi y_{t-1} + \delta_1 d_{1,t} + \delta_2 d_{2,t} + u_t. \quad (12)$$

Forecast:

$$\hat{y}_{t_0} = \hat{c} + \hat{\phi} y_{t_0-1} + \hat{\delta}_1 d_{1,t_0} + \hat{\delta}_2 d_{2,t_0}. \quad (13)$$

Why it is strong: The model adjusts the mean at each break, matching the DGP structure.

3.4 AR(1) + Estimated Break (Grid Search)

In practice, break dates are often unknown. This method estimates a single break date \hat{T}_b by searching over candidate break points and choosing the one that minimizes the sum of squared errors (SSE) from two segment-specific AR(1) fits:

$$\hat{T}_b = \arg \min_{T_b} (\text{SSE}_1(T_b) + \text{SSE}_2(T_b)). \quad (14)$$

After estimating \hat{T}_b , forecasting is performed using the most recent regime (post-break segment).

Why it can be worse than oracle dummies: Estimation error in \hat{T}_b can lead to incorrect segmentation and larger forecast errors.

3.5 Markov Switching (Two Regimes)

Markov switching models allow regime changes without specifying break dates. A simplified interpretation is:

$$y_t = c_{s_t} + \phi y_{t-1} + u_t, \quad (15)$$

where $s_t \in \{1, 2\}$ is an unobserved regime evolving according to Markov transition probabilities:

$$P(s_t = j \mid s_{t-1} = i) = p_{ij}. \quad (16)$$

Practical issue: Estimation relies on non-linear maximum likelihood and can be numerically fragile; convergence failures may occur frequently in automated Monte Carlo loops.

4 Monte Carlo Design

For each scenario (single or multiple breaks), the simulation proceeds as follows:

1. Generate a time series $\{y_t\}_{t=1}^T$ from the relevant DGP.
2. Choose the forecast time after the break(s):
 - Single break: $t_0 = T_b + 20$.
 - Multiple breaks: $t_0 = b_2 + 20$.
3. Fit each forecasting method using training data $\{y_1, \dots, y_{t_0-1}\}$ and compute a one-step forecast \hat{y}_{t_0} .
4. Compute forecast error $e = y_{t_0} - \hat{y}_{t_0}$.
5. Repeat steps (1)–(4) for N replications (e.g., $N = 200$) to obtain an empirical distribution of errors for each method.

Algorithm 1: Monte Carlo Simulation for Structural Break Forecasting

1. **For** $n = 1, \dots, N$ **do**
2. Generate $\{y_t\}_{t=1}^T$ from DGP with breaks
3. Set forecast origin t_0 (post-break)
4. **For** each method m **do**
5. Fit method m on $\{y_1, \dots, y_{t_0-1}\}$
6. Compute forecast $\hat{y}_{t_0}^{(m)}$
7. Store error $e_n^{(m)} = y_{t_0} - \hat{y}_{t_0}^{(m)}$
8. **End For**
9. **End For**
10. Compute metrics (RMSE, MAE, Bias) for each method

5 Evaluation Metrics

Let $\{e_i\}_{i=1}^N$ be the forecast errors for a method across Monte Carlo replications. We compute:

5.1 Bias

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N e_i. \quad (17)$$

Bias indicates systematic under- or over-forecasting. Values close to zero imply the method is not systematically wrong.

5.2 Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |e_i|. \quad (18)$$

MAE measures the typical magnitude of forecast errors.

5.3 Root Mean Squared Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2}. \quad (19)$$

RMSE penalizes large errors more heavily than MAE. In this study, we rank methods primarily by RMSE (lower is better).

6 Results and Interpretation

The comparison table reports RMSE, MAE, Bias, and the number of successful replications N (and failure counts where relevant).

6.1 Single-Break Case

In the single-break scenario, the ranking by RMSE typically shows:

- **AR(1) + Break Dummy (oracle)** performs best (lowest RMSE/MAE, bias near zero).
- **AR(1) + Estimated Break (grid)** usually performs worse than oracle dummies due to break date estimation error.
- **Rolling AR(1)** improves over global AR(1) by adapting to recent data, but remains inferior to explicit break modeling.
- **Global AR(1)** performs worst because it mixes pre- and post-break regimes, generating biased forecasts after the break.

Table 1: Single-Break Scenario: Forecast Performance Comparison

Method	RMSE	MAE	Bias
AR(1) + Break Dummy (Oracle)	—	—	—
AR(1) + Estimated Break	—	—	—
Rolling AR(1)	—	—	—
Global AR(1)	—	—	—

6.2 Multiple-Break Case

In the multiple-break scenario:

- **AR(1) + Break Dummy (oracle)** again performs best, because multiple dummies adjust the mean at each break.
- **Rolling AR(1)** typically performs better than global AR(1) because it focuses on recent data dominated by the final regime.
- **Global AR(1)** remains weaker as it pools observations from all regimes, increasing bias and error.

6.3 Markov Switching Results

In the reported output, Markov switching produced NaN metrics with $N = 0$ and high failure counts. This indicates that the model did not successfully converge or produce forecasts during the Monte Carlo loop in the implemented settings. This is consistent with the known numerical sensitivity of regime-switching maximum likelihood estimation.

Therefore, Markov switching performance is not reported in those runs.

7 Conclusion

Across both single and multiple break designs, the strongest forecasting performance is achieved by the **AR(1) + Break Dummy (oracle)** specification, as it explicitly incorporates the structural changes in the mean and therefore produces the lowest RMSE and MAE and bias close to zero.

However, it is important to emphasize that the break dummy approach is an *oracle benchmark* because it assumes the break date(s) are known. In real applications where break dates are unknown, adaptive approaches such as rolling estimation or break detection procedures become necessary, though they may produce slightly higher forecast errors than the oracle benchmark.

7.1 Implications for Practice

1. When break dates can be reliably identified (e.g., policy changes, known events), explicit break modeling via dummies is preferred.
2. When breaks are unknown, rolling-window methods offer a robust alternative that adapts to recent data.
3. Break detection methods (grid search) can help but introduce estimation uncertainty.
4. Markov switching models, while theoretically appealing, may face numerical challenges in practice.

7.2 Future Work

Extensions of this research could include:

- Variance breaks and heteroskedasticity modeling (GARCH).
- Multi-step ahead forecasting.
- Real data applications using S&P 500 realized volatility.
- Confidence interval coverage and probabilistic forecast evaluation.

References

- Box, G. E. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- Chernozhukov, V., Hansen, C., Kallus, N., Spindler, M., and Syrgkanis, V. (2024). *Applied Causal Inference Powered by ML and AI*.
- Dormann, C. F. and Ellison, A. M. (2025). *Statistics by Simulation: A Synthetic Data Approach*. Princeton University Press.
- Francq, C. and Zakoïan, J. M. (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.
- Gaillac, C. and L'Hour, J. (2025). *Machine Learning for Econometrics*. Oxford University Press.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.
- Pesaran, M. H. (2013). The role of structural breaks in forecasting. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2B, pages 1159–1191. Elsevier.
- Wager, S. (2024). *Causal Inference: A Statistical Learning Approach*.