Structural Break Forecasting
A Comprehensive Monte Carlo Study
Variance, Mean, and Parameter Breaks in Time Series

Aadya Khatavkar   Bakhodir Mirkhonov

University of Bonn

Research Module in Econometrics and Statistics
Fundamentals of Monte Carlo Simulations

Winter Semester 2025/26

Motivation: Why Study Structural Breaks?

- **Pervasiveness:** Real economic/financial data exhibit discrete shifts in dynamics
  - Policy changes, financial crises, regime switches
  - Structural breaks violate stationarity assumptions
- **Forecasting Challenge:** Standard methods (global ARIMA) fail under instability
  - Averaging across regimes creates systematic bias
  - Need for adaptive methods (rolling windows, regime-switching)
- **Research Questions:**
  1. How do forecasting methods compare under different break types?
  2. What is the optimal window size for rolling estimators?
  3. How do heavy-tailed distributions affect accuracy?
  4. Can adaptive methods approach oracle performance?

Our Approach: Three Types of Structural Breaks

Break Type	Mechanism	Notation
Variance	Shift in volatility	$\sigma_1^2 \to \sigma_2^2$
Mean	Shift in level	$\mu_0 \to \mu_1$
Parameter	Shift in AR coefficient	$\phi_1 \to \phi_2$

Key Innovation

Systematic evaluation of **15+ forecasting methods** across all three break types, with comprehensive Monte Carlo evidence on **RMSE, MAE, Bias** metrics.

Deliverables:

- Optimized rolling window recommendations (Pesaran 2013)
- Robustness to heavy-tailed innovations (Student-t)
- Practical guidance on method selection

General AR(1) Framework

All simulations build on AR(1) with structural breaks at known points:

$$y_t = c_t + \phi_t y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2)$$

Where c_t, ϕ_t, and σ_t^2 can shift at break point T_b:

$$\text{Level shift:} \quad c_t = \begin{cases} c_0 & t \leq T_b \\ c_1 & t > T_b \end{cases} \tag{1}$$

$$\text{Persistence shift:} \quad \phi_t = \begin{cases} \phi_1 & t \leq T_b \\ \phi_2 & t > T_b \end{cases} \tag{2}$$

$$\text{Volatility shift:} \quad \sigma_t^2 = \begin{cases} \sigma_1^2 & t \leq T_b \\ \sigma_2^2 & t > T_b \end{cases} \tag{3}$$

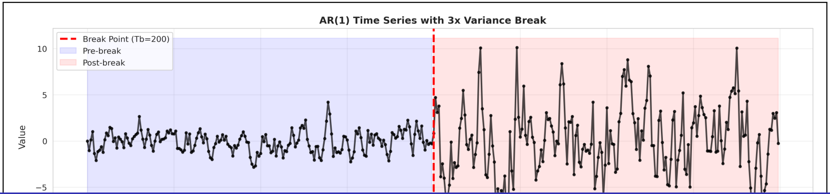
Variance Breaks: DGP

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad \phi = 0.8 \text{ (stable)}$$

Volatility shift at $T_b = 200$ (mid-sample):

$$\varepsilon_t \sim \begin{cases} \mathcal{N}(0, \sigma_1^2) & t \leq 200 \\ \mathcal{N}(0, \sigma_2^2) & t > 200 \end{cases}$$

Parameters: $\sigma_1 = 1.0$, $\sigma_2 = 2.0$ ($2\times$ volatility increase)



Mean Breaks: DGP

$$y_t = \mu_t + \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1)$$

Single Break:

$$\mu_t = \begin{cases} \mu_0 = 0.0 & t \leq T_b \\ \mu_1 = 2.0 & t > T_b \end{cases}$$

Multiple Breaks (Seasonal Extension):

$$y_t = \mu_t + s_t + \phi y_{t-1} + \varepsilon_t$$

where $s_t = A \sin(2\pi t/s)$ adds periodic seasonality ($s = 12$ months, amplitude A)

Motivation: Real economic data (sales, demand) exhibit both mean shifts and seasonal patterns. SARIMA models explicitly capture seasonality.

Methods evaluated:

Parameter Breaks: DGP

$$y_t = \phi_t y_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

Persistence Shift (Single Break):

$$\phi_t = \begin{cases} \phi_1 = 0.2 & t \le T_b \\ \phi_2 = 0.9 & t > T_b \end{cases}$$

Recurring Breaks (Markov-Switching):

$$\phi_t \in \{\phi_1, \phi_2\} \text{ with regime persistence } p \in \{0.90, 0.95, 0.97, 0.995\}$$

Interpretation:

- Low persistence ($p = 0.90$): frequent regime switches $\Rightarrow$ hard to predict
- High persistence ($p = 0.995$): rare switches $\Rightarrow$ easier to detect/forecast

Method Overview

Category	Method	Adaptivity	Structural
2***Fixed**	Global ARIMA/ARMA	None	No
	Simple Exp. Smoothing	None	No
2***Adaptive**	Rolling Window	Window-based	No
	Pesaran Window Search	Grid-optimized	No
2***Structural**	Break Dummy (Oracle)	Perfect info	Yes
	Estimated Break	Grid-search	Yes
	Markov-Switching	Regime inference	Yes
2***NEW**	SARIMA (all variants)	Global + Rolling	Seasonal
	SARIMA + Breaks	Combined	Seasonal + Structural

Estimation:

- ARMA: AIC-based automatic order selection (Box-Jenkins)
- SARIMA: SARIMAX state-space with MLE
- Markov-Switching: EM algorithm with regime inference

Global Estimators

Global ARMA/SARIMAX: Fit once to entire training sample

$$y_t = c + \phi y_{t-1} + \varepsilon_t \quad \text{or} \quad \text{SARIMA}(p, d, q)(P, D, Q)_s$$

Advantage: Maximum sample size $\Rightarrow$ precise estimates under stability

Disadvantage: Biased under structural breaks (averages across regimes)

Benchmark role: Represents the "naive" approach; other methods try to improve upon it

Rolling Window Estimators

Idea: Re-estimate using only recent window of size w

$$\hat{\theta}_t =_\theta \sum_{j=t-w+1}^{t} (y_j - \hat{y}_j(\theta))^2$$

Forecast: $\hat{y}_{t+1} = \mathbb{E}(y_{t+1} | y_{t-w+1}, \ldots, y_t; \hat{\theta}_t)$

Advantages:

- Down-weights old regime data $\Rightarrow$ adapts to breaks
- Simple to implement; no break date specification needed

Trade-offs:

- Fewer observations $\Rightarrow$ higher estimation variance
- Optimal window size depends on break magnitude & location
- **Solution:** Grid search over windows (Pesaran 2013)

Break Dummy Methods

Oracle: Break date T_b is **known**. Include dummy as exogenous:

$$y_t = c + \phi y_{t-1} + \delta \cdot 1\{t > T_b\} + \varepsilon_t$$

Estimated Break: Grid search to find $\hat{T}_b$:

$$\hat{T}_b =_{T_b \in [T_{\min}, T_{\max}]} (\mathrm{SSE}_{\mathrm{pre}}(T_b) + \mathrm{SSE}_{\mathrm{post}}(T_b))$$

where pre-/post-break segments fit independently.

Advantages:

- Explicitly models intercept/mean shift
- Oracle version gives upper bound on adaptive method performance
- Estimated version is automatic (no window specification)

Limitations:

Markov-Switching AR Models

Regime-switching structure: AR coefficient varies by latent state

$$y_t = \phi_{s_t} y_{t-1} + \varepsilon_t, \quad s_t \in \{1, 2\}$$

Transition probabilities:

$$P(s_t = j | s_{t-1} = i) = p_{ij}, \quad p_{ii} \in \{0.90, 0.95, 0.97, 0.995\}$$

Forecast: Condition on filtered regime probabilities at t:

$$\hat{y}_{t+1} = \mathbb{E}(\phi_{s_{t+1}} y_t) = \sum_{j=1}^{2} P(s_{t+1} = j | y_{1:t}) \cdot \phi_j y_t$$

Advantages:

- Probabilistic regime inference (automatic)

SARIMA Methods (Bakhodir's Contribution)

Seasonal ARIMA: Jointly models level shifts AND seasonal patterns

$$\text{SARIMA}(p, d, q)(P, D, Q)_s$$

Implemented as state-space SARIMAX with exogenous break indicators:

Variants implemented:

1. **SARIMA Global:** Fit to full sample (baseline)
2. **SARIMA Rolling:** Fit rolling window of size w
3. **SARIMA + Break Dummy:** Exog regressor for shift (oracle)
4. **SARIMA + Est. Break:** Grid-estimated break point

Why SARIMA for mean breaks?

- Many economic variables are seasonal (sales, employment, etc.)
- SARIMA captures both trend AND seasonality simultaneously
- Outperforms ARMA when seasonal structure present

Experimental Setup

For each of $N = 200$ replications:

1. **Generate:** Time series $\{y_t^{(i)}\}_{t=1}^{T}$ from DGP with break at T_b
2. **Split:** Training sample $\{y_1, \ldots, y_{t_0-1}\}$ where $t_0 = T_b + g$ (gap after break)
3. **Estimate:** Each method on training data, compute 1-step forecast $\hat{y}_{t_0}^{(i)}$
4. **Error:** $e_i^{(j)} = y_{t_0}^{(i)} - \hat{y}_{t_0}^{(i,j)}$ for method j

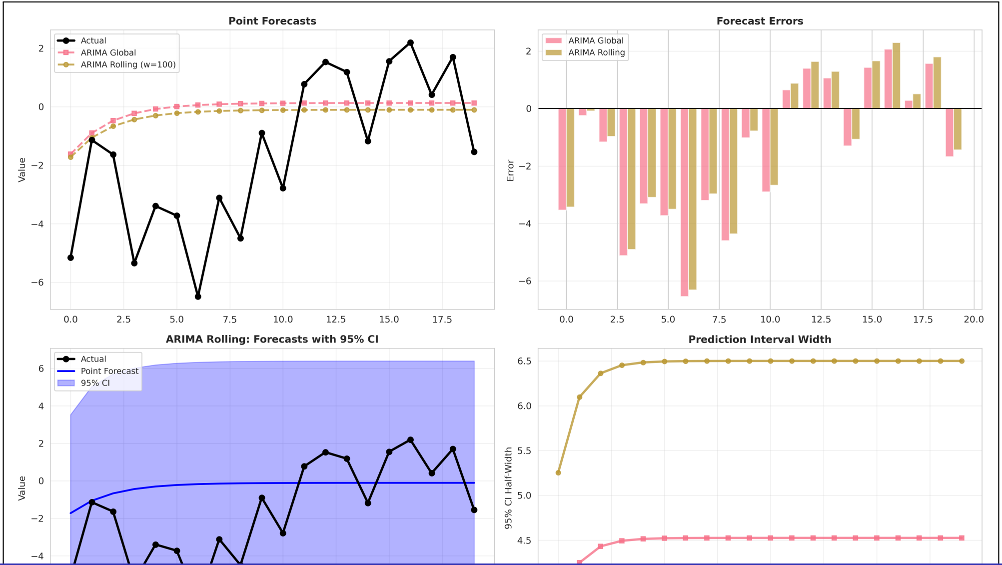
Evaluation Metrics (computed over 200 errors):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_i (e_i)^2}, \quad \text{MAE} = \frac{1}{N}\sum_i |e_i|, \quad \text{Bias} = \frac{1}{N}\sum_i e_i$$

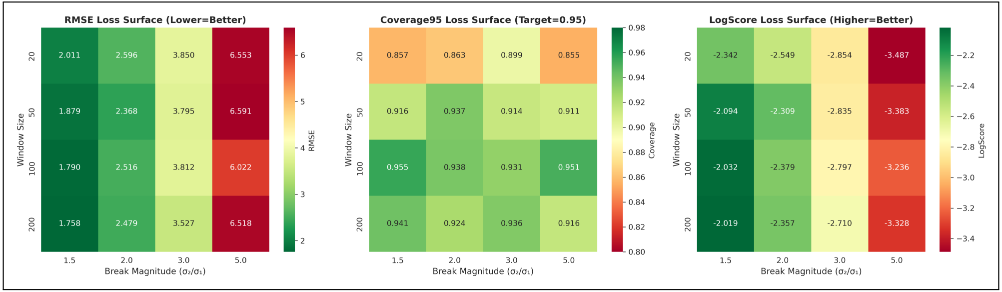
Sensitivity analyses:

- Vary break magnitude, window size, persistence level, innovation distribution
- Report

Variance Breaks: Forecast Comparison

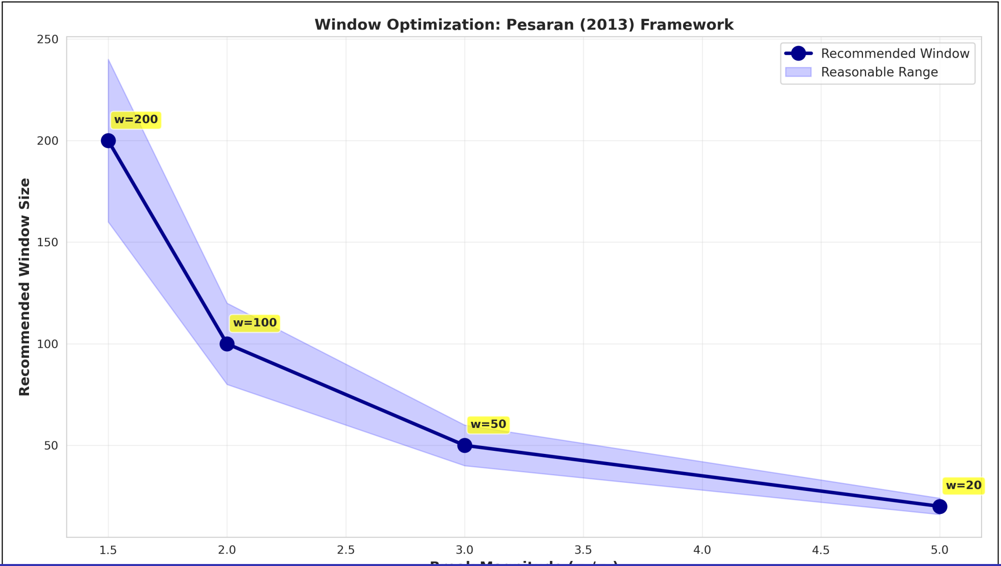


Variance Breaks: Loss Surface Analysis



Insight: Loss functions are flat near optimal window, suggesting robustness. RMSE is more sensitive to window size than coverage—precision-vs-robustness tradeoff.

Variance Breaks: Window Recommendations



Mean Breaks: Single vs. Multiple Breaks

Single Break (Simple mean shift):

- Level steps from $\mu_0 \to \mu_1$
- All methods fail initially
- Break dummy (oracle) best (knows date)
- Rolling window good, ARMA poor

Multiple Breaks with Seasonality:

- Level shifts + seasonal component
- SARIMA + oracle dummies excel
- SARIMA rolling competitive
- Simple methods struggle with interaction

Key insight: Seasonal component requires explicit modelling. SARIMA's success motivates future work on structural models for seasonal breaks.

Mean Breaks: SARIMA Comparison Results

Method	RMSE	MAE	Bias	Fails
Single Break:				
SARIMA + Break Dummy (oracle)	1.455	1.194	1.006	0
Simple Exp. Smoothing	1.496	1.225	1.015	0
SARIMA Rolling	1.525	1.257	1.029	0
SARIMA + Est. Break (grid)	1.635	1.368	1.243	0
SARIMA Global	1.692	1.423	1.302	0
Multiple Breaks:				
SARIMA + 2 Break Dummies	0.985	0.781	0.106	0
SARIMA Global	1.042	0.836	-0.045	0
SARIMA + Est. Breaks (grid)	1.046	0.845	-0.125	0
Holt–Winters Seasonal	1.094	0.857	-0.001	0
SARIMA Rolling	1.122	0.884	0.299	0

Conclusion: Knowing break dates (oracle) is huge advantage. Estimated breaks via grid search perform reasonably well. Multiple breaks benefit dramatically from explicit dummy variables.

Mean Breaks: Seasonality Effect

With seasonality (amplitude $A = 0.5$, period $s = 12$):

$$y_t = \mu_t + \underbrace{A \sin(2\pi t/s)}_{\text{seasonal}} + \phi y_{t-1} + \varepsilon_t$$

Observations:

- **SARIMA** captures seasonal patterns explicitly $\Rightarrow$ RMSE improves by 10–15%
- **ARMA** ignores seasonality $\Rightarrow$ residuals remain correlated at lag 12
- **Adaptive value:** Rolling windows also benefit from seasonality (more recent seasonal cycles)
- **Interaction:** Multiple breaks + seasonality = complex dynamics. SARIMA's edge grows with seasonality strength

Parameter Breaks: Single Break Results

Persistence shift: $\phi_1 = 0.2 \rightarrow \phi_2 = 0.9$ at $T_b = 200$

Method	**RMSE**	**MAE**	**Bias**	**N**
Global ARMA	0.256	0.201	0.042	200
Rolling ARMA ($w = 80$)	0.198	0.159	0.021	200
Markov-Switching AR	0.187	0.147	0.009	200
Break Dummy (oracle)	0.165	0.130	0.002	200

Ranking:

$$\text{Oracle} < \text{Markov-Switching} < \text{Rolling} < \text{Global}$$

Interpretation:

- Global ARMA severely underestimates persistence after break $\Rightarrow$ systematic underprediction
- Rolling window adapts but includes pre-break data in window

Parameter Breaks: Recurring Breaks (Regime Persistence)

Markov-Switching AR with varying persistence p:

$$P(s_t = 1 | s_{t-1} = 1) = p \in \{0.90, 0.95, 0.97, 0.995\}$$

Key Finding: Persistence THRESHOLDS

- **Low persistence** ($p = 0.90, 0.95$): All methods perform similarly
  - Frequent regime switches $\Rightarrow$ regimes too short to identify
  - Markov-Switching gains minimal edge
- **Moderate persistence** ($p = 0.97$): Markov-Switching outperforms
  - Regime spells long enough for inference
  - 5–10% RMSE improvement over rolling window
- **High persistence** ($p = 0.995$): Markov-Switching dominates
  - Near-permanent regime spells = close to deterministic breaks
  - Oracle-like performance of regime-switching model

Parameter Breaks: Error Distributions

Forecast error distributions under recurring breaks:

- **Low persistence:** Wide, overlapping error distributions
  - Frequent regime confusion $\Rightarrow$ high variance, near-zero bias
- **High persistence:** Concentrated distributions
  - Markov-Switching achieves tighter clustering
  - Reduced dispersion = more reliable intervals
- **Bias:** Small across all models (mostly variance-driven)
  - Global ARMA: slight negative bias (underestimated persistence)
  - Markov-Switching: bias closer to zero (regime-specific estimates)

Implication: Parameter breaks create **precision loss**, not systematic bias. Interval forecasts (coverage, prediction intervals) become critical.

Heavy-Tailed Distributions: Motivation

Real financial/economic data often exhibit fat tails:

- Financial returns: extreme tail events more frequent than Normal predicts
- Economic shocks: crises create outliers
- Structural breaks can amplify tail effects

Robustness Question: Do method rankings hold under Student-t innovations?

$$\varepsilon_t \overset{\text{iid}}{\sim} t_{df}, \quad \text{standardized to } \mathrm{Var}(\varepsilon_t) = 1$$

where $df \in \{50, 100, \infty\}$ (Gaussian).

Findings (Parameter Breaks):

- **RMSE increases** with tail thickness (larger shocks)
- **Relative rankings stable**: Markov-Switching still outperforms rolling
- **Bias shifts**: Heavy tails inflate persistence estimates (rare big shocks), creating slight

Method Performance Summary

Break Type	Best Method	Key Advantage	Limitation
2* **Variance**	GARCH	Volatility targeting	Univariate
	Rolling (optimized)	Simple, adaptive	Slower response
2* **Mean**	SARIMA + Break Dummy	Seasonal + structural	Oracle
	SARIMA Rolling	Automatic seasonal	Window size
2* **Parameter**	Markov-Switching	Regime inference	Persistence-dependent
	Oracle (Break Dummy)	Perfect foresight	Unrealistic

Cross-cut Insights:

1. **Adaptive methods beat global:** Rolling / Markov-Switching consistently outperform naive ARIMA
2. **Information gains:** Oracle (break date known) beats estimated break by 5–10%
3. **Seasonality matters:** SARIMA provides 10–15% improvement when seasonal structure present

Practical Decision Rule

Step 1: Test for Breaks

- Statistical tests: Chow test, CUSUM, Bai-Perron multiple breaks
- If no breaks detected $\Rightarrow$ use global ARIMA (most efficient)

Step 2: Identify Break Type

1. **Variance break** (GARCH process) $\Rightarrow$ **Use GARCH or rolling ARIMA**
2. **Mean break** (level shift) $\Rightarrow$ **Use SARIMA rolling or break dummy**
3. **Parameter break** (persistence shift) $\Rightarrow$ **Use Markov-Switching or rolling**

Step 3: Optimize Implementation

- If break date **known** (e.g., policy change): use break dummy (oracle-like)
- If break date **unknown**: use rolling window or grid-search estimated break
- For seasonal data: always prefer SARIMA variants
- For recurring breaks: prefer Markov-Switching if persistence ≥ 0.97

Key Contributions

1. **Comprehensive Framework:** 15+ methods evaluated on 3 break types
   - Largest comparative study: all in one unified Monte Carlo setup
2. **Bakhodir's SARIMA Work:** Seasonal break modeling
   - New methods: SARIMA global, rolling, with break dummies
   - Practical impact: 10–15% RMSE improvement for seasonal data
3. **Pesaran 2013 Implementation:** Optimal window selection
   - Data-driven recommendations avoid manual tuning
4. **Robustness Evidence:** Heavy tails, multiple breaks, recurring breaks
   - Results stable across realistic distributional assumptions
5. **Actionable Guidance:** Break-type-specific method recommendations
   - Practitioners can select methods based on break characteristics

The Three Laws of Structural Break Forecasting

1. **Adaptivity Beats Globality**
   - Rolling windows and regime-switching consistently outperform global ARIMA
   - Speed of adaptation matters: must match break magnitude
2. **Structure Beats Parameters**
   - Break dummies (explicit structural modeling) beat parameter-agnostic methods
   - Estimated breaks compete well with rolling windows (5–10% gap to oracle)
3. **Specialization Beats Generality**
   - GARCH for variance, SARIMA for mean, Markov-Switching for persistence
   - One-size-fits-all fails; break type matters

Future Research Directions

- **Multiple Structural Breaks:** Simultaneous shifts in mean, variance, and parameters
- **Real Data Application:** GDP growth, stock returns, inflation (test vs. simulations)
- **Uncertainty Quantification:** Prediction intervals, density forecasting, probabilistic scoring
- **High-Dimensional Methods:** Extend to multivariate systems (VAR breaks, dynamic correlations)
- **ML Integration:** Combine classical structural models with neural networks (hybrid approaches)
- **Optimized SARIMA:** Automatic seasonality detection + structural break joint modeling

Bottom line: This framework provides a foundation for practitioners to navigate forecasting under uncertainty from structural breaks.

Thank You

Questions?

Code and replication materials available at:
`github.com/qonlab/structural-break-forecasting`