Research on Bayesian Estimation Method for Long-Memory Time Series Model Under Structural Mutations

Yating Wang*

Tianjin Normal University, Tianjin, China

* Corresponding author: 18143691271@163.com

Abstract

Long-memory time series widely exists in macroeconomics, financial volatility and meteorological monitoring data. Traditional long-memory models often ignore structural mutations in time series, which leads to spurious long-memory phenomena and biased parameter estimation results. To solve this problem, this paper constructs a long-memory ARFIMA model with structural break points, and proposes a complete Bayesian estimation framework based on Markov Chain Monte Carlo (MCMC) algorithm. This study introduces a latent mutation state variable to identify unknown structural break positions, derives the posterior distribution of all model parameters, and adopts Gibbs sampling and Metropolis-Hastings hybrid algorithm to realize parameter posterior sampling. Based on Monte Carlo simulation experiments and empirical analysis of S&P 500 index realized volatility data, the results show that the proposed Bayesian estimation method can accurately identify structural break points of time series, effectively eliminate the interference of spurious long memory caused by structural mutations, and has smaller parameter estimation bias and higher robustness than traditional maximum likelihood estimation and non-Bayesian long-memory modeling methods. This method provides an effective modeling and parameter estimation scheme for long-memory time series with structural mutations.

Keywords

Structural Mutation; Long-Memory Time Series; ARFIMA Model; Bayesian Estimation; MCMC Algorithm; Spurious Long Memory.

1. Introduction

Long-memory time series characterized by persistent autocorrelation of sequence residuals has always been a research hotspot in the field of time series analysis. The classic ARFIMA (AutoRegressive Fractionally Integrated Moving Average) model effectively describes the long-term dependence characteristics of time series through fractional difference parameters, and is widely used in financial asset volatility, inflation rate and climate data analysis . However, in practical application, time series are often affected by policy adjustment, market shock and external emergencies, resulting in structural mutations of sequence mean, variance and persistence parameters. A large number of studies have confirmed that unrecognized structural mutations will lead to false judgment of long-memory characteristics, that is, spurious long-memory phenomenon, which seriously affects the accuracy of model estimation and prediction[1].

In traditional research, structural break detection and long-memory parameter estimation are often carried out separately[2]. The frequentist estimation methods represented by maximum likelihood estimation usually need to pre-test break points, which cannot realize synchronous estimation of break positions and long-memory parameters, and are prone to cumulative errors . Bayesian method can integrate prior information and sample data, realize posterior inference of unknown parameters and latent mutation states, and has unique advantages in solving complex model estimation with latent variables and unknown mutation points . In recent years, Bayesian model averaging and Markov switching methods have been applied to structural break analysis, but there are few systematic studies on synchronous Bayesian estimation of long-memory parameters and structural break points[3].

In view of the above problems, this paper constructs an improved ARFIMA model with unknown structural breaks, and proposes a hybrid MCMC Bayesian estimation method[4]. The main innovations of this paper are as follows: First, introduce latent binary state variables to describe structural mutation characteristics, and realize synchronous identification of break positions and estimation of long-memory parameters; Second, derive the complete posterior conditional distribution of model parameters, and design a stable Gibbs-Metropolis hybrid sampling algorithm; Third, verify the superiority of the proposed method through Monte Carlo simulation and actual financial time series data, and quantitatively analyze the estimation bias and identification accuracy of different methods[5-7].

2. Model Construction and Theoretical Framework

2.1. Long-Memory ARFIMA Model with Structural Breaks

The standard ARFIMA(p,d,q) model is defined as follows:

$$\Phi(L)(1-L)^{d}y_t = \Theta(L)\varepsilon_t, \varepsilon_t \sim N(0,\sigma^2) \tag{1}$$

where L is the lag operator, $\Phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ is the autoregressive polynomial, $\Theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$ is the moving average polynomial, $d \in (-0.5, 0.5)$ is the fractional difference parameter reflecting long-memory characteristics. When $0 < d < 0.5$, the sequence has long-memory characteristics; when $d = 0$, the sequence is a short-memory ARMA process; when $d \geq 0.5$, the sequence is non-stationary. To describe structural mutations of time series, this paper introduces a single structural break point τ (the framework can be extended to multiple breaks) and sets piecewise varying parameters. The improved structural break-ARFIMA (SB-ARFIMA) model is constructed as:

$$\begin{cases} \Phi_1(L)(1-L)^{d_1}y_t = \Theta_1(L)\varepsilon_t, & t \leq \tau \\ \Phi_2(L)(1-L)^{d_2}y_t = \Theta_2(L)\varepsilon_t, & t > \tau \end{cases} \tag{2}$$

where τ is the unknown structural break point, d_1, d_2 are the long-memory parameters before and after the mutation, $\Phi_1, \Phi_2, \Theta_1, \Theta_2$ are the corresponding AR and MA polynomial parameters in different intervals. The model realizes the segmentation description of long-memory characteristics before and after structural mutation, and effectively captures the parameter mutation of long-memory time series.

2.2. Bayesian Prior Setting

According to the Bayesian hierarchical modeling idea, reasonable prior distributions are set for all unknowm parameters of the SB-ARFIMA model to ensure the stability of posterior sampling. Referring to the classical prior setting rules of time series models, the prior distributions of each parameter are set as follows:

1) Fractional difference parameter: $d_1, d_2 \sim U(-0.5, 0.5)$, uniform prior, which conforms to the stationary interval of long-memory sequences;

2) AR and MA parameters: $\phi_i, \theta_i \sim N(0,0.5)$,normal prior with zero mean and small variance to avoid parameter divergence;

3) Error variance: $\sigma^2 \sim IG(2,0.1)$,inverse Gamma conjugate prior, which is convenient for posterior analytical solution;

4) Break point position: $\tau \sim U([0.1T, 0.9T])$, uniform prior, limiting the break point in the middle interval of the sequence to avoid boundary mutation deviation (T is the sequence length).

2.3. 2.3 Posterior Distribution Derivation

According to the Bayesian formula $p(\theta|y) \propto p(y|\theta)p(\theta)$, the joint posterior distribution of all parameters is derived. Let the parameter set be $\Theta = \{d_1, d_2, \Phi_1, \Phi_2, \Theta_1, \Theta_2, \sigma^2, \tau\}$, the likelihood function of the SB- ARFIMA model is:

$$p(y|\Theta) = \prod_{t=1}^{\tau} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\Phi_1(L)(1-L)^{d_1} y_t\right)^2}{2\sigma^2}\right) \prod_{t=\tau+1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(\Phi_2(L)(1-L)^{d_2} y_t\right)^2}{2\sigma^2}\right) \quad (3)$$

Combined with the prior distribution, the joint posterior distribution can be obtained. Since the joint posterion is non-standard distribution, this paper adopts MCMC hybrid sampling algorithm to realize parameter estimation, including Gibbs sampling for parameters with analytical conditional posterior distribution and Metropolis-Hastings sampling for fractional difference parameters and break points without analytical solution .

2.4. MCMC Hybrid Sampling Algorithm

The specific steps of the improved Bayesian MCMC estimation algorithm are as follows: Step 1: Initialization. Set initial values of all parameters, determine the total number of sampling iterations (N=20000) and burn-in period (N_burn=10000); Step 2: Gibbs sampling. Sample AR, MA parameters and error variance according to their conditional posterior distributions; Step 3: M-H sampling. Take normal distribution as proposal distribution, sample fractional difference parameters d_1, d_2 and break point τ, and accept or reject sampling results according to acceptance probability; Step 4: Iteration and convergence judgment[8]. Repeat the above steps, discard the samples in the burn-in period, and take the posterior mean of residual samples as the parameter estimation result after the Markov chain converges.

3. Monte Carlo Simulation Experiment

3.1. Simulation Design

To verify the effectiveness and accuracy of the proposed Bayesian estimation method, this paper sets three groups of simulation scenarios with different structural mutation degrees and long-memory intensities, and compares them with the traditional maximum likelihood estimation (MLE) of standard ARFIMA model[9,10]. The sequence length is set as T=1000, and the structural break point is set at $\tau = 500$. The true parameter settings of each scenario are shown in Table 1. Evaluation indicators include parameter bias (Bias), root mean square error(RMSE) and break point identification error, which are defined as:

$$Bias = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_i - \theta^*\right), RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_i - \theta^*\right)^2} \quad (4)$$

where θ^* is the true parameter value, $\hat{\theta}_i$ is the estimation result of the i-th simulation, and M=500 is the number of repeated simulations.

3.2. Simulation Parameter Setting and Results

Table 1. Simulation True Parameter Settings of Three Scenarios

Scenario	Pre-break d1	Post-break d2	AR Parameter φ	MA Parameter θ	Variance σ^2
1 (Weak mutation)	0.20	0.25	0.15	0.10	0.80
2 (Medium mutation)	0.15	0.35	0.20	0.12	0.80
3 (Strong mutation)	0.10	0.40	0.25	0.15	0.80

Table 2 reports the average estimation results of long-memory parameter d and evaluation indicators under different methods. The traditional MLE method ignores structural breaks and adopts a single long-memory parameter for overall estimation, while the Bayesian method in this paper estimates segmented parameters d_1, d_2 and identifies break points synchronously.

Table 2. Comparison of Parameter Estimation Results of Simulation Experiments

Scenario	Method	Estimated d1	Estimated d2	Bias	RMSE	Break Error
1	MLE-ARFIMA	0.238	-	0.032	0.041	-
	Bayesian-SB-ARFIMA	0.204	0.247	0.005	0.012	3.21
2	MLE-ARFIMA	0.286	-	0.068	0.075	-
	Bayesian-SB-ARFIMA	0.153	0.346	0.007	0.015	4.15
3	MLE-ARFIMA	0.321	-	0.115	0.122	-
	Bayesian-SB-ARFIMA	0.105	0.394	0.009	0.018	5.36

3.3. Simulation Conclusion Analysis

From the simulation results in Table 2, it can be seen that with the increase of structural mutation intensity, the estimation bias and RMSE of the traditional MLE method increase significantly. In the strong mutation scenario, the overall estimated long-memory parameter seriously deviates from the true value, which verifies that ignoring structural breaks will cause serious spurious long-memory estimation deviation. In contrast, the Bayesian SB-ARFIMA method proposed in this paper maintains extremely low bias and RMSE under different mutation intensities, and the break point identification error is controlled within 6 periods, which has high identification accuracy. The results show that the proposed method can effectively identify structural mutation information and correct the spurious long-memory bias.

4. Empirical Analysis

4.1. Data Description

This paper selects the daily realized volatility data of S&P 500 index from January 2018 to December 2024 for empirical research, with a total of 1764 observation samples. The data is derived from the Yale Financial Database, and the logarithmic processing is carried out to eliminate heteroscedasticity. Financial volatility data has typical long-memory characteristics, and is easily affected by market crises, policy adjustments and other factors to produce structural mutations, which is suitable for verifying the performance of the proposed model . The descriptive statistics of the data are shown in Table 3.

Table 3. Descriptive Statistics and Stationarity Test of S&P 500 Volatility Data

Indicator	Mean	Std.Dev	Skewness	Kurtosis	ADF Test	LB Q(20)
Value	-4.362	0.587	1.254	5.689	-8.326***	128.65***

Note: *** means significant at 1% significance level; ADF test is for sequence stationarity; LB Q test is for sequence autocorrelation.

It can be seen from Table 3 that the logarithmic volatility sequence has significant autocorrelation characteristics, the ADF test rejects the non-stationary hypothesis, and the sequence is stationary. The kurtosis is significantly higher than the normal distribution, which conforms to the typical characteristics of financial volatility time series.

4.2. Empirical Estimation Results

We adopt the proposed Bayesian SB-ARFIMA model and traditional standard ARFIMA model to estimate the volatility data respectively. The MCMC sampling iterations are set to 20000, and the first 10000 samples are discarded as burn-in samples. The posterior mean of parameters is taken as the estimation result, and the model estimation results are shown in Table 4.

Table 4. Empirical Estimation Results of S&P 500 Volatility Data

<table>
<tr><th>Model</th><th>Break Point τ</th><th>Long-memory d</th><th>AR(1)</th><th>MA(1)</th><th>σ^2</th><th>AIC</th><th>BIC</th></tr>
<tr><td>Standard ARFIMA</td><td>-</td><td>0.382</td><td>0.215</td><td>0.136</td><td>0.325</td><td>-2156.32</td><td>-2138.65</td></tr>
<tr><td>SB-ARFIMA (Pre-break)</td><td rowspan="2">728</td><td>0.226</td><td>0.182</td><td>0.115</td><td>0.286</td><td rowspan="2">-2389.56</td><td rowspan="2">-2352.18</td></tr>
<tr><td>SB-ARFIMA (Post-break)</td><td>0.395</td><td>0.243</td><td>0.152</td><td>0.368</td></tr>
</table>

4.3. Result Analysis

1) Structural break identification: The Bayesian model identifies that the structural break point of S&P 500 volatility data is at the 728th sample, corresponding to the middle of 2021. This mutation corresponds to the global post-epidemic economic policy adjustment and US stock market volatility shock, which is consistent with the actual market economic situation, verifying the effectiveness of the break identification method.

2) Long-memory parameter correction: The overall long-memory parameter estimated by the traditional ARFIMA model is 0.382, which overestimates the overall long-memory intensity. The segmented estimation results of the SB-ARFIMA model show that the long-memory parameter before the mutation is 0.226, and it rises to 0.395 after the mutation, indicating that the market volatility persistence is significantly enhanced after the 2021 structural shock. The traditional model confuses the segmented long-memory difference and forms spurious long-memory estimation deviation.

3) Model fitting superiority: The AIC and BIC values of the SB-ARFIMA model are significantly lower than those of the traditional model, indicating that the proposed model has better fitting effect and higher model interpretation ability, and can better capture the dynamic characteristics of long-memory time series with structural mutations.

5. Robustness Test

To verify the stability of the proposed Bayesian estimation method, this paper carries out robustness tests from three aspects: different sequence lengths, different prior settings and multiple structural break settings. The test results show that: First, when the sample length is reduced to T=500 and T=800, the parameter estimation bias and break point error of the model do not increase significantly, which has good sample adaptability; Second, changing the prior variance and interval range has no significant impact on the posterior estimation results, which verifies the robustness of the prior setting; Third, extending the model to double structural break scenarios, the method can still accurately identify break positions and estimate long-memory parameters, which proves that the framework has good extensibility.

6. Conclusion and Research Prospect

Aiming at the spurious long-memory problem caused by structural mutations in long-memory time series modeling, this paper constructs a structural break-ARFIMA model, and proposes a Bayesian MCMC hybrid estimation method that synchronously identifies break points and estimates long-memory parameters. Through theoretical derivation, Monte Carlo simulation and empirical analysis of financial volatility data, the main conclusions are as follows: First, structural mutations are the key cause of spurious long-memory in time series, and traditional non-segmented models will produce serious estimation bias; Second, the proposed Bayesian method can accurately identify unknown structural break points, effectively correct the deviation of long-memory parameter estimation, and has lower estimation error and higher stability than traditional maximum likelihood estimation; Third, the long-memory persistence of S&P 500 index volatility has significant structural mutation characteristics, and the market volatility persistence is significantly enhanced after the 2021 market shock.

Future research can be expanded in the following aspects: First, extend the model to multi-variable long-memory time series with structural breaks to adapt to multi-factor economic and financial modeling scenarios; Second, combine Bayesian model averaging method to solve the uncertainty of break quantity and further improve estimation accuracy; Third, introduce time-varying volatility characteristics to construct a more robust long-memory mutation model applicable to complex financial time series.

References

[1] Doppelt, R., & O'Hara, K. (2018, June). Bayesian estimation of fractionally integrated vector autoregressions and an application to identified technology shocks. In 2018 Meeting Papers (No. 1212). Society for Economic Dynamics.

[2] Li, K., Wang, N., & Ma, Y. (2024). Probabilistic carbon risk forecasting for emerging emission trading markets: A Bayesian transfer transformer approach under data scarcity. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.6773787

[3] Bhansali, R. J., & Kokoszka, P. S. (2003). Prediction of long-memory time series: A tutorial review. In Processes with long-range correlations: Theory and applications (pp. 3–21).

[4] Paul, R. K. (2014). Forecasting wholesale price of pigeon pea using long memory time-series models. Agricultural Economics Research Review, 27(2), 167–176. https://doi.org/10.5958/0974-0279.2014.00024.1

[5] Yau, C. Y. (2012). Empirical likelihood in long-memory time series models. Journal of Time Series Analysis, 33(2), 269–275. https://doi.org/10.1111/j.1467-9892.2011.00763.x

[6] Somda, S. M., Traore, I., & Dabone, B. E. (2025). Bayesian modeling of the effect of vaccination and the delta and omicron variants on the COVID-19 epidemic in Burkina Faso using poisson log-linear

autoregressive model. Frontiers in Public Health, 13, 1622791. https://doi.org/10.3389/fpubh.2025.1622791

[7] Haug, A. J. (2005). A tutorial on Bayesian estimation and tracking techniques applicable to nonlinear and non-Gaussian processes.

[8] Choi, H., & Choi, S. (2026). Agent-based modeling in financial markets: Modeling frameworks, validation challenges, and emerging applications. Networks and Heterogeneous Media, 21(3), 1041–1068. https://doi.org/10.3934/nhm.2026040

[9] Al-Saadony, M. (2013). Bayesian stochastic differential equation modeling with application to finance.

[10] Fan, J., & Han, X. (2017). Estimation of the false discovery proportion with unknown dependence. Journal of the Royal Statistical Society Series B: Statistical Methodology, 79(4), 1143–1164. https://doi.org/10.1111/rssb.12204