

# 基于贝叶斯推断的缓冲自回归模型参数估计

韦祖栋

华南农业大学数学与信息学院, 广东 广州

收稿日期: 2023年1月10日; 录用日期: 2023年1月30日; 发布日期: 2023年2月15日

## 摘要

本文基于马尔科夫链蒙特卡罗(MCMC)算法的贝叶斯推断下, 研究缓冲自回归模型的参数估计问题。通过缓冲自回归模型参数的联合后验分布得到各参数的条件后验分布, 再利用Gibbs抽样、随机游走Metropolis-Hastings算法抽取样本, 并以正态分布为建议分布, 对缓冲自回归模型进行参数估计。随机模拟结果显示用该方法估计各参数的效果较好。

## 关键词

缓冲自回归模型, 贝叶斯推断, MCMC算法

# Parameter Estimation of Buffered Autoregressive Model Based on Bayesian Inference

Zudong Wei

College of Mathematics and Informatics, South China Agricultural University, Guangzhou Guangdong

Received: Jan. 10<sup>th</sup>, 2023; accepted: Jan. 30<sup>th</sup>, 2023; published: Feb. 15<sup>th</sup>, 2023

## Abstract

Based on Bayesian inference of Markov chain Monte Carlo (MCMC) algorithm, this paper studies the parameter estimation of buffered autoregressive models. The conditional posterior distribution of each parameter is obtained through the joint posterior distribution of the parameters of the buffered autoregressive model. Then, the Gibbs sampling and random walk Metropolis Hastings algorithm are used to extract samples, and the normal distribution is used as the suggested distribution to estimate the parameters of the buffered autoregressive model. The results of stochastic simulation show that the method is effective in estimating the parameters.

## Keywords

Buffered Autoregression Model, Bayesian Inference, MCMC Algorithm

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

时间序列分析作为一个非常有效的研究实际问题的方法,在过去的 20 多年间,前人对时间序列模型进行了大量的理论与实证研究。但随着经济社会的快速发展, Clements 和 Hansen 等人发现,经典的线性自回归时间序列模型不再适用于经济和金融时间序列的研究[1] [2],难以解释金融时间序列中的非线性结构变化,非线性的时间序列模型也因此应运而生。门限自回归(TAR)模型作为非线性时间序列模型中成熟的代表,自 Tong 开创性的引入 TAR 模型以来[3],门限型时间序列模型在经济、环境等领域的研究发挥了重要作用。传统的门限自回归模型根据门限变量的状态变化将时间序列划分为两个甚至更多的不同区域,并在每个区域建立不同的线性自回归模型。当门限变量超过或低于某个值时,时间序列的概率结构突然发生变化。但是在一些领域中,例如经济学、地理学等,我们经常观察到状态转变的缓冲现象,即当门限变量超出或低于限制区域时,我们感兴趣的变量仍停留在前一状态保持不变。Li 等人将这种缓冲现象引入到门限自回归模型当中[4]。同时,考虑到在金融时间序列模型中,条件异方差是一个极其重要的部分,Lo 等人在汇率研究中提出了缓冲 GARCH 模型[5]。Zhu 等人随后也介绍了具有广义自回归条件异方差的缓冲自回归模型[6],并通过汇率分析,将其与 AR-GARCH 和 T-GARCH 模型进行了对比,进一步说明了缓冲自回归 GARCH 模型在刻画时间序列模型中的条件均值与方差的优越性。

Markov 链 Monte Carlo (MCMC)算法作为现代统计计算中重要的一种方法,主要利用马尔科夫链的细致平衡条件进行采样,再通过所采样的样本进行数值计算的一种方法。过去的许多研究员,为了简化 TAR、TARMA 等模型统计推断中的分析工作和数值多重积分问题,均使用 MCMC 方法来估计模型的未知参数,不仅相对快速地估计出模型参数,同时对比其他方法,提高了估计的精度。早在 Chen 和 Lee 中,马尔科夫链蒙特卡罗(MCMC)算法就被用于门限自回归模型的贝叶斯推断[7]。So 则利用可逆跳 MCMC 算法检验 GARCH 和 TGARCH 模型[8]。不管是对门限参数估计还是对其他参数估计,MCMC 算法都显示出了较好的效果。

过去的研究主要集中在缓冲自回归模型的应用方面,而对该模型的参数估计及检验研究较少,特别是结合 MCMC 算法与贝叶斯推断的方法寥寥无几。而基于过往研究的基础,我们对模型进行贝叶斯推断后,我们主要利用 Gibbs 抽样和随机游走的 Metropolis-Hastings (M-H)算法的 MCMC 方法对缓冲自回归模型做参数估计,得到参数估计结果落在 95%置信区间内,证明结果是可靠和有效的。

## 2. 模型描述

我们定义如下的缓冲自回归模型: 对于序列  $\{y_t, t = 1, 2, \dots, n\}$ ,

$$y_t = \begin{cases} \psi_0 + \sum_{i=1}^p \psi_i y_{t-i} + a_t, & R_t = 1, \\ \varphi_0 + \sum_{i=1}^p \varphi_i y_{t-i} + a_t, & R_t = 0, \end{cases}$$

其中，状态参数  $R_t$  定义如下：

$$R_t = \begin{cases} 1 & y_{t-d} \leq r_L \\ 0 & y_{t-d} \geq r_U \\ R_{t-1} & \text{其他} \end{cases}$$

其中， $a_t, t=1,2,\dots,n$  相互独立，其均值为 0，方差为  $\sigma^2$ ； $y_{t-d}$  为门限变量； $d$  是延迟参数； $r_L$  和  $r_U$  是缓冲区  $(r_L, r_U]$  的门限参数。对于现实中的时间序列  $\{y_t\}$ ，考虑到前几个观测值的状态可能无法很好地识别，我们很自然地假设， $(r_L, r_U]$  为  $[a, b]$  的子集， $a, b$  为门限变量  $y_{t-d}$  数据样本的上下  $\alpha$  经验分位数[9] [10]。

### 3. 贝叶斯推断

为了对参数进行抽样，我们需要求得各个参数的条件后验分布结果，同时，各参数条件后验分布形式决定了参数所使用的抽样方式。根据共轭先验分布的特点，我们进行如下设定：

设  $y_t$  由  $t=1$  时刻开始观察，记：

$$\begin{aligned} \psi_1 &= (\psi_0, \psi_1, \dots, \psi_p)^\top, \quad \psi_2 = (\varphi_0, \varphi_1, \dots, \varphi_p)^\top, \\ \gamma &= (r_L, r_U)^\top, \quad Y = \{y_1, y_2, \dots, y_n\} \end{aligned}$$

则  $y_t$  的条件均值可得：

$$\mu_t = E[y_t | Y_{(t-1)}] = \left[ \psi_0 + \sum_{i=1}^p \psi_i y_{t-i} \right] R_t + \left[ \varphi_0 + \sum_{i=1}^p \varphi_i y_{t-i} \right] (1 - R_t)$$

其中， $Y_{(t-1)} = (y_{t-1}, \dots, y_{t-p})^\top$ 。

记  $\Psi = (\psi_1, \psi_2, \gamma, \sigma, d)^\top$  为待估的参数向量。我们利用贝叶斯推断得到各参数的条件后验分布，再根据条件后验分布的情况，选择合适的 MCMC 方法进行参数估计。

对于未知参数，我们选择的先验分布如下：

- 1)  $\psi_1, \psi_2$  相互独立， $\psi_j \sim N(\psi_{j0}, V_j^{-1}), j=1,2$ ，其中  $\psi_{j0}, V_j^{-1}$  为正态分布的均值向量和协方差矩阵；
- 2)  $\sigma^2$  服从逆伽马分布  $IG(\alpha, \beta)$ ；
- 3) 门限参数  $r_L, r_U$  服从  $[a, b]$  上的均匀分布，且  $a \leq r_L < r_U \leq b$ ；
- 4) 延迟参数  $d$  服从  $1, 2, \dots, d_0$  上离散的均匀分布，且  $d_0$  数值较大。

以上各先验分布得超参数已知，根据贝叶斯公式，可推出  $\Psi$  的先验分布满足：

$$p(\Psi) \propto \prod_{j=1}^2 p(\psi_j) p(\sigma^2) I(a < r_L \leq r_U \leq b)$$

可得  $\Psi$  的联合后验分布为：

$$p(\Psi | y) \propto p(y | \Psi) p(\Psi)$$

模型的似然函数为：

$$p(\Psi | y) \propto \prod_{t=d_0+1}^n \frac{1}{\sqrt{2\pi} a_t} \left\{ \left[ \frac{(y_t - u_{t1})^2}{2a_t^2} \right] \cdot R_t + \left[ \frac{(y_t - u_{t2})^2}{2a_t^2} \right] \cdot (1 - R_t) \right\}$$

其中，

$$u_{it} = E(y_{it} | y_{it-1}) = \begin{cases} \psi_0 + \sum_{i=1}^p \psi_i y_{t-i} + a_t, & i = 1 \\ \varphi_0 + \sum_{i=1}^p \varphi_i y_{t-i} + a_t, & i = 2 \end{cases}$$

根据贝叶斯定理，计算各参数后验分布如下：

1)  $\psi_j, j=1,2$  条件后验分布相互独立且服从正态分布：

$$p(\psi_j | y, \Psi - \psi_j) \sim N(\psi_{j0}^*, V_j^{*-1}), j=1,2$$

其中

$$\begin{aligned} \psi_{j0}^* &= \left( \frac{A_j}{\sigma^2} + V_j \right)^{-1} \left( \frac{A_j}{\sigma^2} \hat{\psi}_j + V_j \psi_{j0} \right), V_j^* = \left( \frac{A_j}{\sigma^2} + V_j \right) \\ x_{j,t} &= (1, y_{t-1}, \dots, y_{t-h})^T, X_1^* = (x_{1,h+1}, x_{1,h+2}, \dots, x_{1,\tau})^T, \\ X_2^* &= (x_{2,\tau+1}, x_{2,\tau+2}, \dots, x_{2,n})^T, Y_1^* = (y_{h+1}, y_{h+2}, \dots, y_\tau)^T, \\ Y_2^* &= (y_{\tau+1}, y_{\tau+2}, \dots, y_n)^T, A_i = X_i^{*T} X_i^*, \\ \hat{\psi}_j &= A_i^{-1} X_i^{*T} Y_i^*. \end{aligned}$$

2)  $\sigma^2$  条件后验分布为逆伽马分布：

$$p(\sigma^2 | y, \Psi - \sigma^2) \sim IG\left(\alpha + \frac{n-p}{2}, \beta + \frac{s^2}{2}\right),$$

其中，

$$s^2 = \sum_{i=1}^2 (Y_i^* - X_i^{*T} \psi_i)^T (Y_i^* - X_i^{*T} \psi_i).$$

3)  $\gamma$  的条件后验密度函数为：

$$p(\gamma | y, \Psi - \gamma) \propto \exp\left(-\frac{1}{2\sigma^2} \cdot s^2\right) \cdot I(a < r_L < r_U < b),$$

其中  $s^2$  是关于  $r$  的函数。

4)  $d$  条件后验分布为多项分布，密度函数为：

$$p(d | y, \Psi - d) = \frac{L(\psi_1, \psi_2, \sigma^2, \gamma, d | y)}{\sum_{d=1}^{d_0} L(\psi_1, \psi_2, \sigma^2, \gamma, d | y)},$$

其中， $L(\psi_1, \psi_2, \sigma^2, \gamma, d | y) = \sigma^{n-p} \exp\left(-\frac{s^2}{2\sigma^2}\right)$ 。

综上，我们可看出，除了门限参数向量  $\gamma$  的条件后验分布不是标准分布，其余参数的条件后验分布均是常见的分布，因此可以利用 Gibbs 抽样。而对门限参数向量采用随机游走的 Metropolis-Hasting 抽样的 MCMC 方法[11]。

我们以  $\gamma$  为例，记  $f(\bullet)$  为其条件后验密度函数。M-H 算法的基本思想是通过在参数空间随机取值作为起始点，通过建议函数与转移概率构建一个细致平稳的马尔科夫链，使得  $f(\bullet)$  成为稳定分布，并通过

获得的随机点与起始点概率密度的比值大小决定是否保留随机点。则利用随机游走的 M-H 抽样步骤如下 [12]:

- 1) 在第  $l$  步迭代时, 通过建议分布产生一个  $\gamma$  候选值  $\gamma^*$ , 使其满足:

$$\gamma^* = \gamma^{[l-1]} + \varepsilon_\gamma, \varepsilon_\gamma \sim N(0, \sigma_\gamma^2)$$

$\gamma^{[l-1]}$  是  $l$  在  $l-1$  步迭代时接受的值。通常, 我们选择  $\sigma_\gamma^2$  为对角矩阵, 通过监测和调整对角线上的元素大小来保证接受率处在 0.25~0.60 之间。

- 2) 根据接受概率:

$$p = \min \left\{ 1, \frac{f(\gamma)}{f(\gamma^{[l-1]})} \right\} > u, u \sim U(0,1)$$

决定是否接受候选值  $\gamma^*$ 。否则, 在第  $l$  步迭代时,  $\gamma^{[l]} = \gamma^{[l-1]}$ 。其中,  $U(0,1)$  表示 0,1 之间的均匀分布。

#### 4. 随机模拟

我们利用 R 软件进行随机模拟, 基于 Gibbs 抽样和随机游走的 M-H 算法讨论参数估计效果, 我们设定如下缓冲自回归模型:

$$y_t = \begin{cases} 0.3 + 0.8y_{t-1} + a_t, & R_t = 1 \\ 0.4 + 0.7y_{t-1} + a_t, & R_t = 0 \end{cases}$$

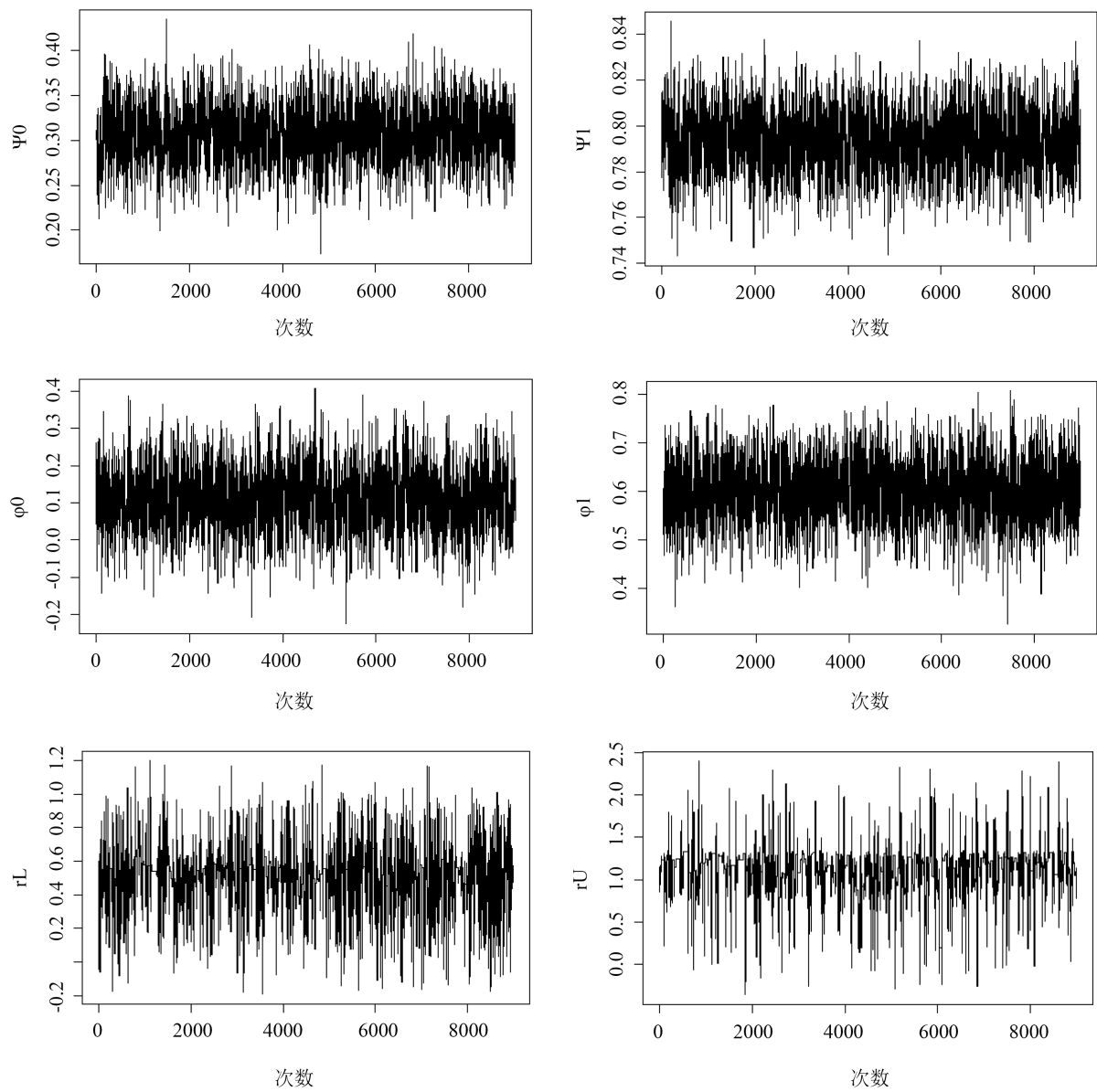
状态参数设定为:

$$R_t = \begin{cases} 1 & y_{t-d} \leq 0.5 \\ 0 & y_{t-d} \geq 1.0 \\ R_{t-1} & \text{其他} \end{cases}$$

在模拟实验中, 我们生成 1500 个观测值, 并选择最后 1000 个观测值作为样本。我们在实验中运行 MCMC 迭代过程 18,000 次, 根据每个参数的采样方案舍去样本产生的前 9000 个样本, 取后面 9000 个作为有效样本。为了验证方法的稳定性, 我们进行了两次模拟实验。最后缓冲自回归模型的随机模拟结果如下。

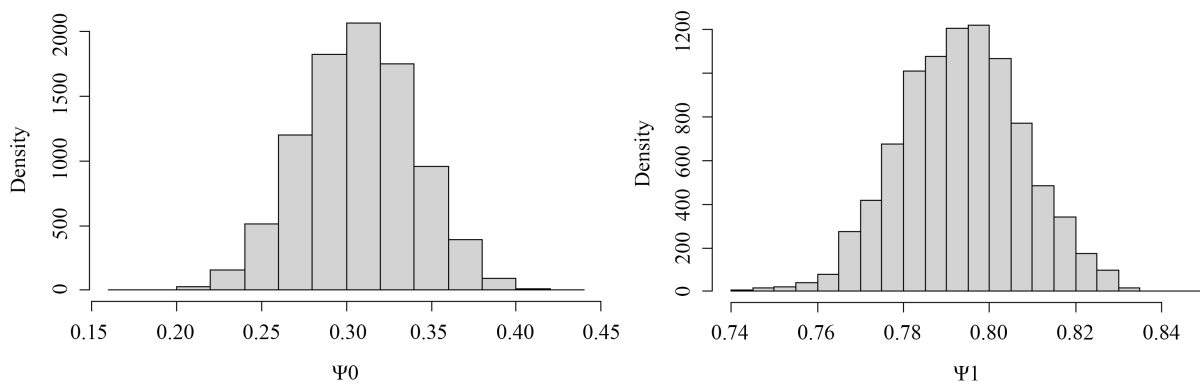
**Table 1.** Results of random simulation experiment  
**表 1.** 随机模拟实验结果

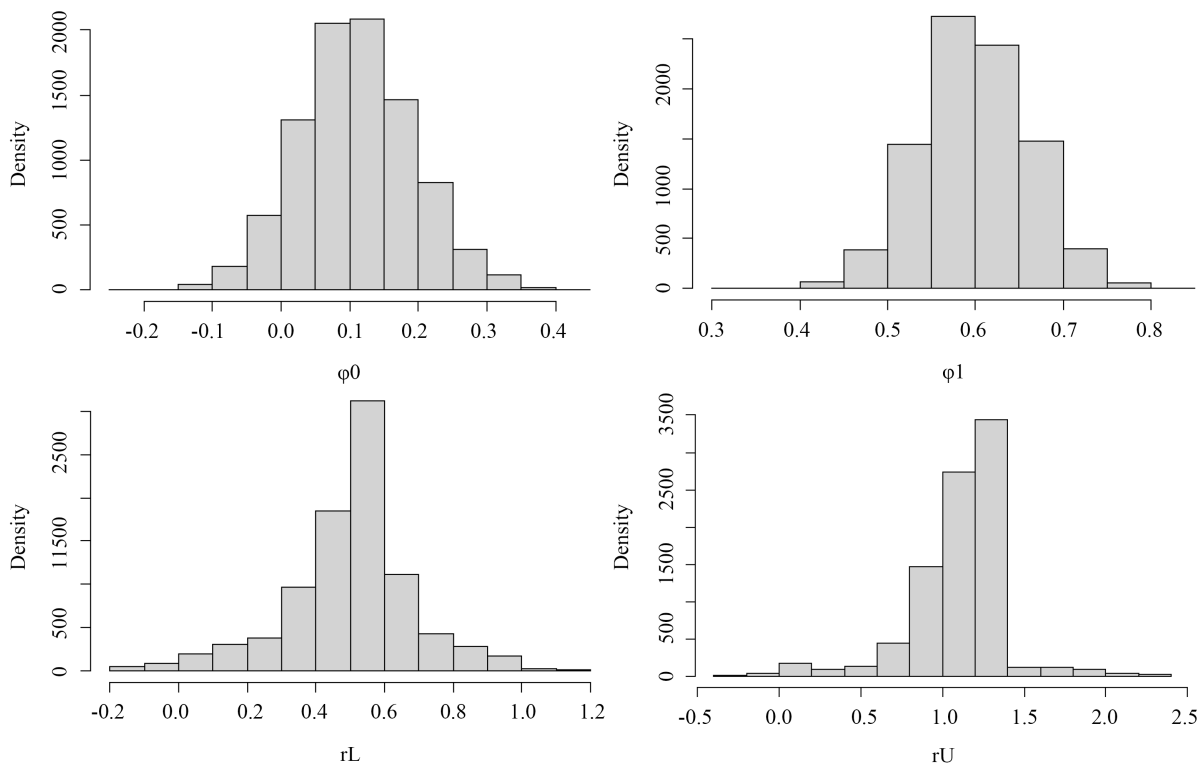
参数	真值	后验均值	后验标准差	置信区间
$\psi_0$	0.3	0.3072	0.0326	[0.3035,0.3109]
$\psi_1$	0.8	0.7936	0.0139	[0.7911,0.8098]
$\varphi_0$	0.1	0.1101	0.0843	[0.1040,0.1161]
$\varphi_1$	0.6	0.5992	0.0623	[0.5941,0.6044]
$r_L$	0.5	0.5057	0.1923	[0.4966,0.5148]
$r_U$	1.0	1.0946	0.3081	[1.0832,1.1061]
$d$	2	2	0	/



**Figure 1.** Trace plots of 9000 MCMC iterations of estimated parameters

**图 1.** 9000 次 MCMC 迭代估计参数的轨迹图





**Figure 2.** The histograms of the estimated parameters  
**图 2.** 参数频率分布直方图

当模型与真实情况下拟合度较高时，后验均值应与真实值十分接近。从表 1 的结果可以看出，缓冲自回归模型各个参数的后验均值与真值十分接近，并且每个参数的后验标准差也较小。同时，在 95% 的置信度下，参数的后验均值落在置信区间之中，可以看出结果是可靠的。在图 1 参数的迭代轨迹图与图 2 参数频率分布直方图也表明各参数迭代过程稳定、收敛的。依据以上的结果与分析，我们认为对缓冲自回归模型用 MCMC 抽样方案是有效的，随机模拟实验的结果较好。

### 5. 总结

本文在基于贝叶斯推断的基础上，得到缓冲自回归模型的条件后验分布，并结合 MCMC 抽样的方法对模型的参数进行估计。通过模拟实验表明，不仅模型的参数估计效果较为理想，还说明本文模型选择的后验概率估计是有效的。有了以上的实验结果依据，进一步，我们可以对模型的缓冲区进行非线性检验等。此外，在模拟实验的过程中，我们考虑到缓冲模型在预测均值、方差时的表现出来的优越性，因此，我们也可以将其应用在计算风险值方面。

### 参考文献

- [1] Clements, M.P., Franses, P.H. and Swanson, N.R. (2004) Forecasting Economic and Financial Time-Series with Non-linear Models. *International Journal of Forecasting*, **20**, 169-183. <https://doi.org/10.1016/j.ijforecast.2003.10.004>
- [2] Hansen, B.E. (2011) Threshold Autoregression in Economics. *Statistics and Its Interface*, **4**, 123-127. <https://doi.org/10.4310/SII.2011.v4.n2.a4>
- [3] Tong, H. (1978) On a Threshold Model. In: Chen, C.H., Ed., *Pattern Recognition and Signal Processing*. Sijthoff & Noordhoff, Amsterdam.
- [4] Li, G., Guan, B., Li, W.K. and Yu, P.L.H. (2015) Hysteretic Autoregressive Time Series Models. *Biometrika*, **102**,

- 
- 717-723. <https://doi.org/10.1093/biomet/asv017>
- [5] Lo, P., Li, W. and Li, G. (2016) On Buffered Threshold GARCH Model. *Statistica Sinica*, **26**, 1555-1567. <https://doi.org/10.5705/ss.2014.098t>
- [6] Zhu, K., Yu, P.L.H. and Li, W.K. (2017) Buffered Autoregressive Models with Conditional Heteroscedasticity: An Application to Exchange Rates. *Journal of Business and Economic Statistics*, **35**, 528-542. <https://doi.org/10.1080/07350015.2015.1123634>
- [7] Chen, C.W.S. and Lee, J.C. (1995) Bayesian Inference of Threshold Autoregressive Models. *Journal of Time Series Analysis*, **16**, 483-492. <https://doi.org/10.1111/j.1467-9892.1995.tb00248.x>
- [8] So, M.K.P. and Chen, C.W.S. (2003) Subset Threshold Autoregression. *Journal of Forecasting*, **22**, 49-66. <https://doi.org/10.1002/for.859>
- [9] Chan, K.S. (1991) Percentage Points of Likelihood Ratio Tests for Threshold Autoregression. *Journal of the Royal Statistical Society: Series B*, **53**, 691-696. <https://doi.org/10.1111/j.2517-6161.1991.tb01858.x>
- [10] Andrews, D.W.K. (1993) Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*, **61**, 821-856. <https://doi.org/10.2307/2951764>
- [11] 夏强, 刘金山. 基于贝叶斯推断的 TAR 模型的门限非线性检验[J]. 应用概率统计, 2011, 27(3): 276-282.
- [12] 茆诗松, 王静龙, 濮晓龙. 高等梳理统计[M]. 北京: 高等教育出版社, 1998.