

一阶随机系数整数值二项自回归模型的参数估计

陈志非, 卢飞龙

辽宁科技大学理学院, 辽宁 鞍山

收稿日期: 2026年1月10日; 录用日期: 2026年2月4日; 发布日期: 2026年2月10日

摘 要

本文系统研究了一阶随机系数整数值二项自回归模型, 在参数估计方面, 分别构建了模型参数的条件最小二乘估计量与拟似然估计量, 并基于最小二乘估计量进一步引入经验似然方法, 构造了具有辅助信息的最大经验似然估计量。最后, 通过仿真实验, 在不同样本量、参数设置与数据生成机制下, 综合比较了上述估计方法的有限样本性能, 为离散时间序列数据的建模提供了更丰富的估计工具与理论参考。

关键词

随机系数, 二项自回归模型, 参数估计, 最大经验似然方法

Parameter Estimation for the First-Order Random Coefficient Integer-Valued Binomial Autoregressive Model

Zhifei Chen, Feilong Lu

College of Science, University of Science and Technology Liaoning, Anshan Liaoning

Received: January 10, 2026; accepted: February 4, 2026; published: February 10, 2026

Abstract

This paper systematically studies the first-order random coefficient integer-valued binomial autoregressive model. In parameter estimation, it constructs the conditional least squares estimator and the quasi-likelihood estimator for the model parameters, and further introduces an empirical likelihood method based on the least squares estimator to construct the maximum empirical likeli-

hood estimator with auxiliary information. Finally, through simulation experiments under different sample sizes, parameter settings, and data generating mechanisms, it comprehensively compares the finite-sample performance of the above estimation methods, providing more estimation tools and theoretical references for modeling discrete time series data.

Keywords

Random Coefficients, Binomial Autoregressive Model, Parameter Estimation, Maximum Empirical Likelihood Method

Copyright © 2026 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

在计量经济学、生物统计学与社会学等领域,计数型时间序列的建模具有重要意义。此类数据通常取值为非负整数,且序列内部存在显著的自相关性。传统连续型时间序列模型因假定数据连续且服从正态分布,难以直接适用于此类离散数据。为此 McKenzie [1]、Al-Osh 和 Alzaid [2]提出了整数值自回归(Integer-Valued Autoregressive, INAR)模型。其核心在于引入了由 Steutel 和 van Harn [3]提出的二项稀疏算子作为传统自回归算子的离散化替代,详细内容参考 WEIB [4]。具体而言,算子“ $\circ$ ”定义为:

$$\alpha \circ X = \sum_{i=1}^X B_i, \quad (1)$$

其中, $\alpha \in [0, 1)$, X 为非负整数值随机变量, $\{B_i\}$ 是一列独立同分布(i.i.d.)的伯努利随机变量,满足 $P(B_i = 1) = \alpha$, 并且与 X 独立。然而,标准的一阶整数值二项自回归(First-order Integer-Valued Binomial Autoregressive, BAR(1))模型 α 是一个固定常数,该模型具体内容可以参考 Weiß [5],这隐含了过程的同质性假设。在实际应用中,许多计数序列的动态性往往受到未观测到的时变因素影响,导致其依赖强度 α 可能随时间发生随机波动。固定系数模型难以刻画这种内在的异质性与过度离散现象。为了增强模型的灵活性与现实描述能力,研究者将随机系数的思想引入 INAR 框架,即假设 α 本身是一个随机变量或随机过程,从而得到一阶随机系数整数值二项自回归(First-order Random Coefficient Integer-Valued Binomial Autoregressive, RCBAR(1))模型。该类模型由 Zheng 等人[6]提出,允许自回归系数随时间随机变动,能更灵活地捕捉计数序列的波动聚集与厚尾特征。虽然 RCBAR(1)模型更具一般性,但其统计推断也面临更多挑战。现有研究已初步探讨了模型性质并发展了基础估计方法,但仍存在以下不足:首先,不同估计方法在有限样本下的性能缺乏系统比较;其次,在更一般的模型设定下,如何构建更高效、更稳健的估计量仍不明确;最后,已有研究对估计量在不同数据生成机制下的稳健性关注不足。

本文主要研究 RCBAR(1)模型,首先给出了模型的基本性质,在参数估计方面,提出了条件最小二乘与拟似然估计量,并引入经验似然(Empirical Likelihood, EL)方法,具体可参考 Owen [7]-[9]、Qin 和 Lawless [10],而 Chuang 和 Chan [11],以及 Chan 和 Ling [12]将经验似然方法应用于自回归模型。本文构建了可融合辅助信息的最大经验似然估计量,该方法不依赖扰动项分布且具有更高理论效率。最后,参考 Zhang [13]和 Yang [14]通过模拟研究比较了各估计量在不同场景下的有限样本表现。

本文余下部分的结构安排如下:第二节正式给出模型设定并详述其基本性质;第三节逐一阐述三种估计量的构造方法;第四节展示并分析模拟实验结果;第五节总结全文并展望未来研究方向。

2. RCBAR(1)模型定义及基本性质

RCBAR(1)模型由以下方程定义:

$$X_t = \alpha_t \circ X_{t-1} + \beta_t \circ (n - X_{t-1}), \quad t \geq 1 \quad (2)$$

其中, $\{(\alpha_t, \beta_t)\}$ 是一列取值于 $(0,1) \times (0,1)$ 上的独立同分布二元随机变量序列。且 (α_t, β_t) 和 t 时刻所有的稀疏算子及它们的计数序列包括初始值 X_0 都是相互独立的, X_0 满足 $E(X_0^2) < \infty$, 同时 (α_t, β_t) 也与所有稀疏算子及历史状态序列 $\{X_s, s < t\}$ 相互独立。

由式(2)定义的 RCBAR(1)是一个有限状态空间上的齐次马尔可夫链。类似于 Zheng [6]中的式(4), 其一步转移概率具有如下形式:

$$\begin{aligned} p_{kl} &:= P(X_t = k | X_{t-1} = l) \\ &= \sum_{j=\max(0, k+l-n)}^{\min(k, l)} \binom{l}{j} \binom{n-l}{k-j} \times \int_0^1 \int_0^1 \alpha_t^j (1-\alpha_t)^{l-j} \beta_t^{k-j} (1-\beta_t)^{n-l-k+j} dP_{\alpha, \beta} > 0, \end{aligned}$$

其中, $P_{\alpha, \beta}$ 是二元随机变量 (α_t, β_t) 的累积分布函数(CDF)。因此 $P = (p_{kl})_{k, l=0, \dots, n}$ 是一个本原矩阵(不可约且非周期的), 这意味着过程 $\{X_t\}$ 是遍历的且存在唯一平稳解。

参考 Weiß [15]、Ma [16]和 RCBAR(1)模型的定义可得, 自相关函数为 $\rho(k) = (\alpha - \beta)^k$ (当 $k \geq 0$ 时)。最重要的回归性质是:

$$\begin{aligned} E(X_t | X_{t-1}) &= (\alpha - \beta) X_{t-1} + n\beta; \\ \text{Var}(X_t | X_{t-1}) &= (\sigma_\alpha^2 + \sigma_\beta^2) X_{t-1}^2 - 2n\sigma_\beta^2 X_{t-1} + n^2 \sigma_\beta^2. \end{aligned}$$

其中

$$\alpha = E(\alpha_t), \quad \sigma_\alpha^2 = \text{Var}(\alpha_t), \quad \beta = E(\beta_t), \quad \sigma_\beta^2 = \text{Var}(\beta_t).$$

高阶联合矩和累积量的闭式表达式由 Weiß 和 Kim [17]给出。

3. 参数估计

设 $(X_1, X_2, \dots, X_N)$ 是从模型(2)中获得的观测值, 其中参数向量 $\theta = (\alpha, \beta)^T$, $\gamma = (\sigma_\alpha^2, \sigma_\beta^2)^T$ 。

3.1. 条件最小二乘(CLS)估计

首先给出参数向量 θ 的 CLS 估计, CLS 准则函数定义如下:

$$S(\theta) = \sum_{t=2}^N (X_t - E(X_t | X_{t-1}))^2 \quad (3)$$

CLS 估计量可通过最小化该准则函数(3)得到, 关于 α 和 β 的最小二乘估计量为:

$$\begin{aligned} \hat{\alpha}_{\text{CLS}} &= \frac{\sum_{t=2}^N X_t - \hat{\beta}_{\text{CLS}} \sum_{t=2}^N (n - X_{t-1})}{\sum_{t=2}^N X_{t-1}}, \\ \hat{\beta}_{\text{CLS}} &= \frac{\sum_{t=2}^N X_t X_{t-1} \sum_{t=2}^N X_{t-1} - \sum_{t=2}^N X_{t-1}^2 \sum_{t=2}^N X_t}{\sum_{t=2}^N (n - X_{t-1}) X_{t-1} \sum_{t=2}^N X_{t-1} - \sum_{t=2}^N X_{t-1}^2 \sum_{t=2}^N (n - X_{t-1})}. \end{aligned}$$

正如 Klimko 和 Nelson [18]的定理 3.1 和 3.2 所示, 记参数真值为 θ_0 , $\hat{\theta}_{\text{CLS}}$ 是强相合且渐近正态的, 即

$$\sqrt{N}(\hat{\theta}_{\text{CLS}} - \theta_0) \xrightarrow{d} N(0, Q^{-1} R Q^{-1}), N \rightarrow \infty,$$

其中

$$Q = E(X_0 X_0^\top), R = E\left((X_1 - \theta^\top X_0)^2 X_0 X_0^\top\right).$$

下面再给出参数向量 γ 的 CLS 估计, 首先定义一步向前的残差为:

$$\hat{\varepsilon}_t := X_t - \hat{\alpha}_{\text{CLS}} X_{t-1} - \hat{\beta}_{\text{CLS}} (n - X_{t-1}), t = 2, \dots, N.$$

基于此构造 CLS 准则函数:

$$S(\gamma) = \sum_{t=2}^N \left[\hat{\varepsilon}_t^2 - \sigma_\alpha^2 X_{t-1}^2 - \sigma_\beta^2 (n - X_{t-1})^2 \right]^2, \quad (4)$$

其中 $\gamma = (\sigma_\alpha^2, \sigma_\beta^2)^\top$ 为待估参数向量。对 $S(\gamma)$ 关于 γ 最小化, 可得 CLS 估计量:

$$\hat{\gamma}_{\text{CLS}} = \begin{pmatrix} \hat{\sigma}_{\alpha, \text{CLS}}^2 \\ \hat{\sigma}_{\beta, \text{CLS}}^2 \end{pmatrix} = \left(\sum_{t=2}^N \mathbf{z}_t \mathbf{z}_t^\top \right)^{-1} \left(\sum_{t=2}^N \hat{\varepsilon}_t^2 \mathbf{z}_t \right),$$

其中向量 $\mathbf{z}_t = (X_{t-1}^2, (n - X_{t-1})^2)^\top$ 。进一步地, 两个方差参数的显式估计公式的展开式为:

$$\hat{\sigma}_{\alpha, \text{CLS}}^2 = \frac{\left(\sum_{t=2}^N (n - X_{t-1})^4 \right) \left(\sum_{t=2}^N \hat{\varepsilon}_t^2 X_{t-1}^2 \right) - \left(\sum_{t=2}^N X_{t-1}^2 (n - X_{t-1})^2 \right) \left(\sum_{t=2}^N \hat{\varepsilon}_t^2 (n - X_{t-1})^2 \right)}{\left(\sum_{t=2}^N X_{t-1}^4 \right) \left(\sum_{t=2}^N (n - X_{t-1})^4 \right) - \left(\sum_{t=2}^N X_{t-1}^2 (n - X_{t-1})^2 \right)^2},$$

$$\hat{\sigma}_{\beta, \text{CLS}}^2 = \frac{\left(\sum_{t=2}^N X_{t-1}^4 \right) \left(\sum_{t=2}^N \hat{\varepsilon}_t^2 (n - X_{t-1})^2 \right) - \left(\sum_{t=2}^N X_{t-1}^2 (n - X_{t-1})^2 \right) \left(\sum_{t=2}^N \hat{\varepsilon}_t^2 X_{t-1}^2 \right)}{\left(\sum_{t=2}^N X_{t-1}^4 \right) \left(\sum_{t=2}^N (n - X_{t-1})^4 \right) - \left(\sum_{t=2}^N X_{t-1}^2 (n - X_{t-1})^2 \right)^2}.$$

在模型设定正确且观测序列 $\{X_t\}$ 满足平稳性与遍历性等基本假设的前提下, 所提出的 CLS 估计量 $\hat{\gamma}_{\text{CLS}}$ 具有强相合性。具体而言, 随着样本长度 N 趋于无穷, 该估计量以概率 1 收敛至真实参数向量 γ_0 , 即

$$\hat{\gamma}_{\text{CLS}} \xrightarrow{\text{a.s.}} \gamma_0, \text{ 当 } N \rightarrow \infty.$$

这一定理不仅从理论上保证了估计方法在大样本情形下的可靠性, 也为其在实际数据分析中的稳定性提供了坚实支撑。

3.2. 拟似然(MQL)估计

首先, 在模型定义部分已经给出 RCBAR(1)模型的条件方差, 记:

$$V_\gamma(X_t | X_{t-1}) := \text{Var}(X_t | X_{t-1}) = (\sigma_\alpha^2 + \sigma_\beta^2) X_{t-1}^2 - 2n\sigma_\beta^2 X_{t-1} + n^2 \sigma_\beta^2$$

标准的 MQL 估计方程具有如下形式:

$$\sum_{t=2}^N V_\gamma^{-1}(X_t - \alpha X_{t-1} - \beta(n - X_{t-1})) X_{t-1} = 0,$$

$$\sum_{t=2}^N V_\gamma^{-1}(X_t - \alpha X_{t-1} - \beta(n - X_{t-1}))(n - X_{t-1}) = 0.$$

因为条件方差表达式中包含参数 γ 会导致相应的估计方程在一般情况下变得复杂且难以处理。因此

采用通过其他方法获得的关于 γ 的合适一致估计量 $\hat{\gamma}$ 来代替 γ , 然后求解由此得到的 MQL 估计方程, 以估计主要感兴趣的参数 θ 。这种方法导出了以下 θ 的闭式估计量:

$$\begin{pmatrix} \hat{\alpha}_{\text{MQL}} \\ \hat{\beta}_{\text{MQL}} \end{pmatrix} = \begin{pmatrix} \sum_{t=2}^N V_{\hat{\gamma}}^{-1} X_{t-1}^2 & \sum_{t=2}^N V_{\hat{\gamma}}^{-1} X_{t-1} (n - X_{t-1}) \\ \sum_{t=2}^N V_{\hat{\gamma}}^{-1} X_{t-1} (n - X_{t-1}) & \sum_{t=2}^N V_{\hat{\gamma}}^{-1} (n - X_{t-1})^2 \end{pmatrix}^{-1} \times \begin{pmatrix} \sum_{t=2}^N V_{\hat{\gamma}}^{-1} X_{t-1} X_t \\ \sum_{t=2}^N V_{\hat{\gamma}}^{-1} (n - X_{t-1}) X_t \end{pmatrix}$$

作为一种替代方法, 本文将使用 CLS 估计量来估计 σ_α^2 和 σ_β^2 。因为上述估计量不能保证对所有样本都为正, 因此可能不适用于 σ_α^2 和 σ_β^2 的正性要求, 记 θ_0 为参数真值, 公式中 MQL 估计量的渐近正态性由以下定理建立,

$$\sqrt{N} \begin{pmatrix} \hat{\alpha}_{\text{MQL}} - \alpha \\ \hat{\beta}_{\text{MQL}} - \beta \end{pmatrix} \xrightarrow{d} N \left(0, T^{-1}(\theta_0)' K(\theta_0) T^{-1}(\theta_0) \right),$$

其中,

$$\begin{aligned} K(\theta_0) &= \begin{pmatrix} T_2(\theta_0) & T_3(\theta_0) \\ T_3(\theta_0) & T_1(\theta_0) \end{pmatrix}, \quad T^{-1}(\theta_0) = (T_1 T_2 - T_3^2)^{-1} \begin{pmatrix} T_1 & -T_3 \\ -T_3 & T_2 \end{pmatrix}, \\ T_1(\theta_0) &= E \left[(n - X_0)^2 V_{\gamma}^{-1}(X_1 | X_0) \right], \\ T_2(\theta_0) &= E \left[X_0^2 V_{\gamma}^{-1}(X_1 | X_0) \right], \\ T_3(\theta_0) &= E \left[X_0 (n - X_0) V_{\gamma}^{-1}(X_1 | X_0) \right]. \end{aligned}$$

3.3. 最大经验似然(MEL)估计

接下来, 我们介绍 RCBAR(1)模型的最大经验似然方法。对(3)式中定义的 $S(\theta)$ 关于参数 θ 求导, 可得到估计方程:

$$-\frac{1}{2} \frac{\partial S(\theta)}{\partial \theta} = \sum_{t=2}^N m_t(\theta) = 0 \quad (5)$$

对(4)式中定义的 $S(\gamma)$ 关于参数 γ 求导, 可得到估计方程:

$$-\frac{1}{2} \frac{\partial S(\gamma)}{\partial \gamma} = \sum_{t=2}^N m_t(\gamma) = 0 \quad (6)$$

为进一步提升估计效率, 从而引入一个额外的矩条件。该条件在期望为零的前提下, 能够提供关于模型结构的更多信息, 其构造方法详见 Lu [19]。此条件将被纳入经验似然框架, 用于构建带附加约束的似然函数。具体而言, 有:

$$m_t(\theta, \gamma) = \begin{pmatrix} m_{1t}(\theta) \\ m_{2t}(\theta) \\ m_{3t}(\gamma) \\ m_{4t}(\gamma) \end{pmatrix} = \begin{pmatrix} e_t \cdot X_{t-1} \\ e_t \cdot (n - X_{t-1}) \\ (e_t^2 - v_t) \cdot X_{t-1}^2 \\ (e_t^2 - v_t) \cdot (n - X_{t-1})^2 \end{pmatrix} = 0$$

其中,

$$e_t = X_t - \alpha X_{t-1} - \beta(n - X_{t-1}), \quad v_t = \sigma_\alpha^2 X_{t-1}^2 + \sigma_\beta^2 (n - X_{t-1})^2$$

根据 Mykland [20], 剖面经验似然比(ELR)函数为:

$$R(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \max \left\{ \prod_{t=2}^N N p_t \mid p_t \geq 0, \sum_{t=2}^N p_t = 1, \sum_{t=2}^N p_t m_t(\boldsymbol{\theta}, \boldsymbol{\gamma}) = 0 \right\}$$

其中, 公式(5)中 $m_t(\boldsymbol{\theta}) = (X_t - \boldsymbol{\theta}^\top X_{t-1}) X_{t-1}$ 。通过采用标准的拉格朗日乘子法, 可以得到最优的 p_t 的取值为:

$$p_t = \frac{1}{N} \frac{1}{1 + \mathbf{b}^\top(\boldsymbol{\theta}) m_t(\boldsymbol{\theta})}, \quad t = 1, 2, \dots, N$$

其中 $\mathbf{b}(\boldsymbol{\theta})$ 是对应于拉格朗日约束条件的乘子:

$$g(\mathbf{b}(\boldsymbol{\theta})) = \frac{1}{N} \sum_{t=2}^N \frac{m_t(\boldsymbol{\theta})}{1 + \mathbf{b}^\top(\boldsymbol{\theta}) m_t(\boldsymbol{\theta})} = 0$$

由此, 可以得到对数经验似然比统计量:

$$l(\boldsymbol{\theta}) = -2 \log R(\boldsymbol{\theta}) = 2 \sum_{t=2}^N \log(1 + \mathbf{b}^\top(\boldsymbol{\theta}) m_t(\boldsymbol{\theta}))$$

为了研究估计量的渐近性质, 我们对基础过程作如下假设:

(C1) $\{X_t\}$ 是一个平稳过程;

(C2) $E|X_t|^6 < \infty$ 。

对于 RCBAR(1)模型, 我们有以下定理: 在假设(C1)~(C2)下, 记 $\boldsymbol{\theta}_0$ 为参数真值, $\hat{\boldsymbol{\theta}}_{\text{MEL}}$ 是 $\boldsymbol{\theta}$ 的最大经验似然估计量, 当 $n \rightarrow \infty$ 时, 有 $l(\boldsymbol{\theta}_0) \xrightarrow{L} \chi_{(4)}^2$ 。最大经验似然估计量渐近正态,

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{\text{MEL}} - \boldsymbol{\theta}_0) \xrightarrow{L} N(0, \mathbf{J}^{-1} \mathbf{I}(\boldsymbol{\theta}_0) \mathbf{J}^{-1}),$$

其中 $\mathbf{I}(\boldsymbol{\theta}_0) = E[m_t(\boldsymbol{\theta}_0) m_t(\boldsymbol{\theta}_0)^\top]$, 且 $\mathbf{J} = E\left[\frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right]$ 。

3.4. 估计方法的渐近性说明

本文提出的 MQL 与 MEL 估计量均依赖于 CLS 估计量提供的方差参数预估计。在渐近理论中, 若考虑 $\hat{\boldsymbol{\gamma}}_{\text{CLS}}$ 的估计误差, 则 $\hat{\boldsymbol{\theta}}_{\text{MQL}}$ 与 $\hat{\boldsymbol{\theta}}_{\text{MEL}}$ 的渐近分布应修正为:

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{\text{two-stage}}),$$

其中 $\boldsymbol{\Sigma}_{\text{two-stage}}$ 包含来自 $\hat{\boldsymbol{\gamma}}_{\text{CLS}}$ 的额外变异项。由于 CLS 估计量具有相合性, 且模拟结果显示其偏差较小, 本文在理论部分给出的渐近方差矩阵可视为在 $\boldsymbol{\gamma}$ 已知条件下的近似。在实际应用中, 建议使用 Bootstrap 方法构建置信区间以更稳健地反映估计不确定性。

4. 模拟研究

为了评估所提出估计方法的有限样本性质, 本文设计了一系列蒙特卡洛模拟实验。数据生成基于如下带有随机系数的二项自回归(RCBAR(1))模型:

$$X_t = \alpha_t \circ X_{t-1} + \beta_t \circ (n - X_{t-1}), \quad t \geq 1,$$

其中 n 为状态空间上限(固定常数), $\alpha_t \sim \text{Beta}(a_1, b_1)$, $\beta_t \sim \text{Beta}(a_2, b_2)$, 且 α_t 与 β_t 相互独立。初始值 X_0 服从参数为 $(n, 0.5)$ 的二项分布。我们考虑以下四种模型设定:

$$\begin{aligned} \text{M1: } (n; a_1, a_2, b_1, b_2) &= (10; 0.2, 0.5, 0.3, 0.2); \\ \text{M2: } (n; a_1, a_2, b_1, b_2) &= (20; 0.2, 0.3, 0.4, 0.7); \\ \text{M3: } (n; a_1, a_2, b_1, b_2) &= (30; 0.3, 0.3, 0.2, 0.4); \\ \text{M4: } (n; a_1, a_2, b_1, b_2) &= (40; 0.3, 0.5, 0.1, 0.4). \end{aligned}$$

其中,

$$\mu_\alpha = \frac{a_1}{a_1 + b_1}, \sigma_\alpha^2 = \frac{a_1 b_1}{(a_1 + b_1)^2 (a_1 + b_1 + 1)}, \mu_\beta = \frac{a_2}{a_2 + b_2}, \sigma_\beta^2 = \frac{a_2 b_2}{(a_2 + b_2)^2 (a_2 + b_2 + 1)}.$$

本文设置了 M5 为了检验估计方法对随机系数分布形式的敏感性, 考虑 (α_t, β_t) 服从截断正态分布, 即:

$$\alpha_t \sim \text{TN}_{(0,1)}(\mu_\alpha, \sigma_\alpha^2), \beta_t \sim \text{TN}_{(0,1)}(\mu_\beta, \sigma_\beta^2),$$

其中 $\text{TN}_{(0,1)}$ 表示在区间 $(0,1)$ 内截断的正态分布, 参数设置为 $(\mu_\alpha, \sigma_\alpha^2, \mu_\beta, \sigma_\beta^2) = (0.4, 0.16, 0.7143, 0.1205)$, 对应与 M1 相近的边缘均值和方差, 其余设定与 M1 相同。

基于 $M = 1000$ 次重复模拟, 计算偏差(Bias)和均方根误差(RMSE)以评估估计量 ψ (代表 $\hat{\alpha}$ 和 $\hat{\beta}$ 及感兴趣的方差估计量)的表现:

$$\text{Bias} = \frac{1}{M} \sum_{i=1}^M (\psi^{(i)} - \psi), \text{RMSE} = \sqrt{\frac{1}{M} \sum_{i=1}^M (\psi^{(i)} - \psi)^2}.$$

在表格中, 为系统评估不同估计方法在有限样本下的表现, 本文在四种模型下开展了模拟实验。表 1 和表 2 报告了参数 α 和 β 的 Bias 与 RMSE, 其中 N 表示样本量, n 表示状态空间上限。我们考察了 $N \in \{50, 100, 200, 500\}$ 与 $n \in \{10, 20, 30, 40\}$ 的多种组合, 以反映不同样本规模和数据结构下的估计性能。从两表中可以看出, 随着波动程度变化, 各估计方法的偏差与稳定性呈现出明显差异。表 3 和表 4 则比较了在不同模型情境下的不同组合中, 报告了参数 σ_α^2 和 σ_β^2 的 Bias 与 RMSE, 反映了方法在不确定性量化方面的稳健性。表 5 则考虑到对随机系数分布形式的敏感性分析, 假设参数服从截断正态分布并在表格中报告了模拟结果。这些结果为选择适用于不同波动场景的估计方法提供了依据。此外, 图 1 通过样本路径图、自相关函数(ACF)图与偏自相关函数(PACF)图, 直观呈现了所研究模型的动态特征与依赖结构, 为后续的参数估计与模型诊断提供了可视化基础。

图 1 显示, RCBAR(1)模型的样本路径具有典型计数时间序列的波动特征: 自相关函数在滞后一阶显著为正并指数衰减, 表明短期记忆性; 偏自相关函数在滞后二阶后迅速截尾, 支持二阶自回归结构的合理性。表 1 和表 2 结果表明, 随着样本量从 50 增至 500, CLS、MQL 和 MEL 三种方法的偏差与 RMSE 均显著下降, 其中 MEL 表现最优, MQL 次之, CLS 相对较差但计算简便; 参数 α 普遍呈轻微负偏差, β 多为正偏差, 且 β 估计精度略高; 各方法在不同波动水平下均稳健。表 3 和表 4 显示, 方差参数 σ_α^2 与 σ_β^2 的估计误差亦随样本量增大而快速收敛; MEL 仍最优, σ_α^2 误差略大于 σ_β^2 , 验证了方差参数的可估性与方法有效性。从表 5 结果可见, 即使在随机系数服从非 Beta 分布时, CLS、MQL 与 MEL 估计量仍保持良好精度与稳健性, Bias 与 RMSE 未见显著上升, 表明方法对分布形式具有鲁棒性。MEL 在分布误设下仍表现最优, 进一步验证了其实际适用性。

5. 总结

本文通过扩展固定系数模型, 研究了一种新的计数数据模型。该模型允许自回归参数随时间随机变动, 同时推导了过程的平稳性与遍历性, 也推导了模型参数的 CLS 估计量、MQL 估计量与 MEL 估计量,

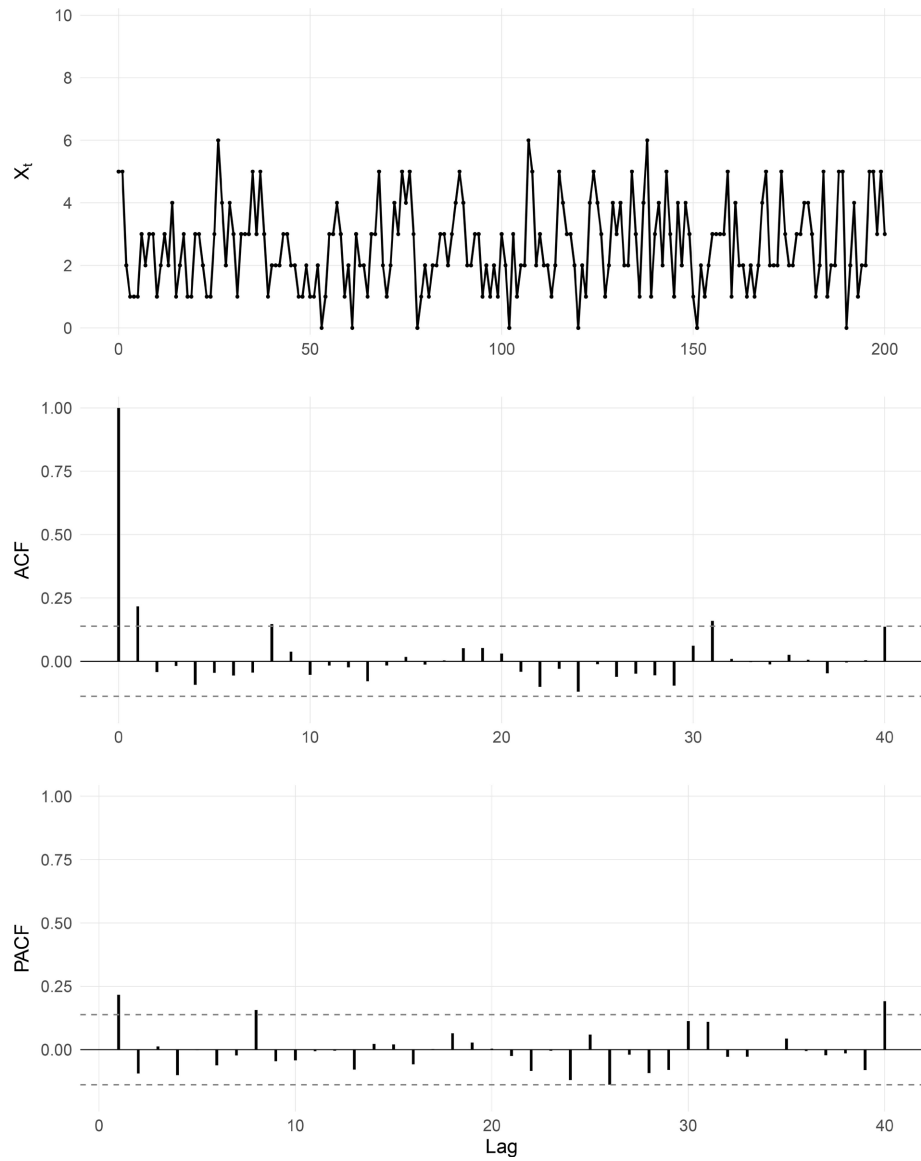


Figure 1. Sample path plot, ACF plot, and PACF plot of the RCBAR(1) model
图 1. RCBAR(1)模型样本路径图、ACF 图和 PACF 图

Table 1. Bias and RMSE of the parameters to be estimated under model M1, the true parameter values are $(\alpha, \beta, \sigma_\alpha^2, \sigma_\beta^2) = (0.4, 0.7143, 0.16, 0.1205)$

表 1. 模型 M1 下待估参数的偏差与均方根误差, 参数真值为 $(\alpha, \beta, \sigma_\alpha^2, \sigma_\beta^2) = (0.4, 0.7143, 0.16, 0.1205)$

N	Para.	CLS		MEL		MQL	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
50	α	-0.0083	0.0801	-0.0021	0.0383	-0.0042	0.0587
	β	0.0742	0.0759	0.0230	0.0455	0.0323	0.0539
	σ_α^2	0.0257	0.0124	0.0068	0.0030	0.0141	0.0069
	σ_β^2	0.0375	0.0550	0.0121	0.0179	0.0164	0.0242

续表

100	α	-0.0062	0.0602	-0.0012	0.0240	-0.0034	0.0412
	β	0.0487	0.0572	0.0145	0.0367	0.0186	0.0386
	σ_α^2	0.0159	0.0083	0.0043	0.0029	0.0071	0.0049
	σ_β^2	0.0216	0.0281	0.0063	0.0122	0.0082	0.0168
200	α	-0.0032	0.0444	-0.0006	0.0207	-0.0017	0.0356
	β	0.0323	0.0443	0.0081	0.0220	0.0104	0.0285
	σ_α^2	0.0093	0.0056	0.0032	0.0025	0.0056	0.0034
	σ_β^2	0.0148	0.0199	0.0045	0.0089	0.0057	0.0128
500	α	-0.0015	0.0317	-0.0003	0.0115	-0.0009	0.0200
	β	0.0109	0.0287	0.0029	0.0087	0.0041	0.0126
	σ_α^2	0.0056	0.0032	0.0016	0.0012	0.0029	0.0021
	σ_β^2	0.0071	0.0110	0.0021	0.0050	0.0034	0.0073

Table 2. Bias and RMSE of the parameters to be estimated under model M2, the true parameter values are $(\alpha, \beta, \sigma_\alpha^2, \sigma_\beta^2) = (0.3333, 0.3, 0.1389, 0.105)$

表 2. 模型 M2 下待估参数的偏差与均方根误差, 参数真值为 $(\alpha, \beta, \sigma_\alpha^2, \sigma_\beta^2) = (0.3333, 0.3, 0.1389, 0.105)$

N	Para.	CLS		MEL		MQL	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
50	α	-0.0078	0.0721	-0.0031	0.0415	-0.0054	0.0568
	β	0.0685	0.0821	0.0196	0.0483	0.0289	0.0497
	σ_α^2	0.0229	0.0115	0.0075	0.0041	0.0136	0.0072
	σ_β^2	0.0342	0.0497	0.0135	0.0192	0.0179	0.0258
100	α	-0.0057	0.0553	-0.0016	0.0268	-0.0039	0.0379
	β	0.0438	0.0615	0.0132	0.0392	0.0173	0.0361
	σ_α^2	0.0143	0.0076	0.0052	0.0034	0.0084	0.0053
	σ_β^2	0.0238	0.0305	0.0078	0.0135	0.0097	0.0179
200	α	-0.0029	0.0398	-0.0009	0.0186	-0.0019	0.0314
	β	0.0286	0.0382	0.0074	0.0243	0.0097	0.0268
	σ_α^2	0.0086	0.0051	0.0036	0.0028	0.0058	0.0037
	σ_β^2	0.0159	0.0217	0.0051	0.0096	0.0062	0.0135
500	α	-0.0012	0.0284	-0.0004	0.0132	-0.0011	0.0189
	β	0.0094	0.0253	0.0024	0.0095	0.0038	0.0119
	σ_α^2	0.0049	0.0028	0.0018	0.0014	0.0026	0.0023
	σ_β^2	0.0082	0.0124	0.0025	0.0057	0.0037	0.0081

Table 3. Bias and RMSE of the parameters to be estimated under model M3, the true parameter values are $(\alpha, \beta, \sigma_\alpha^2, \sigma_\beta^2) = (0.6, 0.4286, 0.16, 0.1442)$

表 3. 模型 M3 下待估参数的偏差与均方根误差, 参数真值为 $(\alpha, \beta, \sigma_\alpha^2, \sigma_\beta^2) = (0.6, 0.4286, 0.16, 0.1442)$

N	Para.	CLS		MEL		MQL	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
50	α	-0.0092	0.0752	-0.0028	0.0436	-0.0048	0.0621
	β	0.0714	0.0785	0.0208	0.0468	0.0301	0.0513
	σ_α^2	0.0248	0.0132	0.0082	0.0045	0.0153	0.0078
	σ_β^2	0.0362	0.0513	0.0128	0.0187	0.0185	0.0249
100	α	-0.0068	0.0581	-0.0017	0.0259	-0.0037	0.0395
	β	0.0459	0.0598	0.0139	0.0379	0.0182	0.0374
	σ_α^2	0.0157	0.0091	0.0058	0.0039	0.0091	0.0057
	σ_β^2	0.0227	0.0328	0.0071	0.0142	0.0092	0.0184
200	α	-0.0031	0.0421	-0.0008	0.0194	-0.0018	0.0335
	β	0.0305	0.0418	0.0079	0.0236	0.0101	0.0276
	σ_α^2	0.0092	0.0063	0.0041	0.0032	0.0064	0.0041
	σ_β^2	0.0153	0.0231	0.0048	0.0103	0.0059	0.0142
500	α	-0.0014	0.0302	-0.0005	0.0141	-0.0012	0.0194
	β	0.0102	0.0271	0.0026	0.0101	0.0039	0.0124
	σ_α^2	0.0053	0.0035	0.0021	0.0017	0.0031	0.0025
	σ_β^2	0.0078	0.0138	0.0023	0.0063	0.0033	0.0086

Table 4. Bias and RMSE of the parameters to be estimated under model M4, the true parameter values are $(\alpha, \beta, \sigma_\alpha^2, \sigma_\beta^2) = (0.75, 0.5556, 0.1339, 0.1301)$

表 4. 模型 M4 下待估参数的偏差与均方根误差, 参数真值为 $(\alpha, \beta, \sigma_\alpha^2, \sigma_\beta^2) = (0.75, 0.5556, 0.1339, 0.1301)$

N	Para.	CLS		MEL		MQL	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
50	α	-0.0107	0.0779	-0.0035	0.0452	-0.0052	0.0614
	β	0.0691	0.0810	0.0216	0.0491	0.0298	0.0526
	σ_α^2	0.0263	0.0145	0.0091	0.0052	0.0168	0.0084
	σ_β^2	0.0387	0.0539	0.0139	0.0203	0.0197	0.0267
100	α	-0.0074	0.0608	-0.0019	0.0276	-0.0041	0.0409
	β	0.0472	0.0625	0.0146	0.0405	0.0191	0.0391
	σ_α^2	0.0172	0.0103	0.0065	0.0044	0.0098	0.0063
	σ_β^2	0.0243	0.0351	0.0078	0.0156	0.0099	0.0198
200	α	-0.0035	0.0453	-0.0009	0.0211	-0.0020	0.0352
	β	0.0318	0.0435	0.0084	0.0253	0.0107	0.0293

续表

500	σ_α^2	0.0101	0.0075	0.0047	0.0037	0.0071	0.0047
	σ_β^2	0.0164	0.0253	0.0053	0.0115	0.0064	0.0155
	α	-0.0016	0.0328	-0.0006	0.0157	-0.0013	0.0210
	β	0.0108	0.0296	0.0028	0.0113	0.0041	0.0137
	σ_α^2	0.0059	0.0042	0.0024	0.0020	0.0035	0.0029
	σ_β^2	0.0085	0.0154	0.0026	0.0071	0.0037	0.0094

Table 5. Bias and RMSE of the parameters to be estimated under model M5 (truncated normal distribution), the true parameter values are $(\alpha, \beta, \sigma_\alpha^2, \sigma_\beta^2) = (0.4, 0.7143, 0.16, 0.1205)$

表 5. 模型 M5 (截断正态分布) 下待估参数的偏差与均方根误差, 参数真值为 $(\alpha, \beta, \sigma_\alpha^2, \sigma_\beta^2) = (0.4, 0.7143, 0.16, 0.1205)$

N	Para.	CLS		MEL		MQL	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
50	α	-0.0091	0.0825	-0.0024	0.0398	-0.0048	0.0602
	β	0.0765	0.0783	0.0243	0.0471	0.0341	0.0556
	σ_α^2	0.0271	0.0138	0.0072	0.0034	0.0153	0.0076
	σ_β^2	0.0392	0.0574	0.0129	0.0188	0.0171	0.0254
100	α	-0.0069	0.0623	-0.0015	0.0251	-0.0038	0.0427
	β	0.0512	0.0595	0.0152	0.0382	0.0195	0.0403
	σ_α^2	0.0168	0.0092	0.0048	0.0032	0.0078	0.0054
	σ_β^2	0.0227	0.0295	0.0068	0.0129	0.0089	0.0176
200	α	-0.0036	0.0459	-0.0008	0.0214	-0.0019	0.0368
	β	0.0338	0.0458	0.0086	0.0231	0.0109	0.0298
	σ_α^2	0.0098	0.0061	0.0035	0.0027	0.0059	0.0038
	σ_β^2	0.0155	0.0208	0.0048	0.0093	0.0060	0.0134
500	α	-0.0018	0.0328	-0.0004	0.0122	-0.0011	0.0209
	β	0.0115	0.0298	0.0031	0.0091	0.0044	0.0132
	σ_α^2	0.0059	0.0035	0.0018	0.0013	0.0031	0.0023
	σ_β^2	0.0075	0.0115	0.0023	0.0053	0.0036	0.0077

并获得了它们的渐近分布。通过模拟研究, 比较了 CLS、MQL 和 MEL 估计量的表现。模拟结果表明: 三种方法在大样本下均能获得良好的估计结果; CLS 方法因其简洁性值得采用, 但 MEL 方法能提供比 CLS、MQL 更有效的估计量。且 MEL 方法展现出更强的稳健性, 虽然 MEL 方法是计算最复杂的, 但正如预期那样, 它具有最高的估计效率。因此综合来看, 推荐将 MEL 方法应用于本文研究的模型。

基金项目

辽宁科技大学博士启动基金(6003000310)。

利益冲突声明

作者声明无任何潜在利益冲突。

参考文献

- [1] McKenzie, E. (1988) Some ARMA Models for Dependent Sequences of Poisson Counts. *Advances in Applied Probability*, **20**, 822-835. <https://doi.org/10.2307/1427362>
- [2] Al-Osh, M.A. and Alzaid, A.A. (1987) First-Order Integer-Valued Autoregressive (INAR(1)) Process. *Journal of Time Series Analysis*, **8**, 261-275. <https://doi.org/10.1111/j.1467-9892.1987.tb00438.x>
- [3] Steutel, F.W. and van Harn, K. (1979) Discrete Analogues of Self-Decomposability and Stability. *The Annals of Probability*, **7**, 893-899. <https://doi.org/10.1214/aop/1176994950>
- [4] Weiß, C.H. and Pollett, P.K. (2013) Binomial Autoregressive Processes with Density-Dependent Thinning. *Journal of Time Series Analysis*, **35**, 115-132. <https://doi.org/10.1002/jtsa.12054>
- [5] Weiß, C.H. and Kim, H. (2012) Parameter Estimation for Binomial AR(1) Models with Applications in Finance and Industry. *Statistical Papers*, **54**, 563-590. <https://doi.org/10.1007/s00362-012-0449-y>
- [6] Zheng, H., Basawa, I.V. and Datta, S. (2007) First-order Random Coefficient Integer-Valued Autoregressive Processes. *Journal of Statistical Planning and Inference*, **137**, 212-229. <https://doi.org/10.1016/j.jspi.2005.12.003>
- [7] Owen, A.B. (1988) Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika*, **75**, 237-249. <https://doi.org/10.1093/biomet/75.2.237>
- [8] Owen, A. (1990) Empirical Likelihood Ratio Confidence Regions. *The Annals of Statistics*, **18**, 90-120. <https://doi.org/10.1214/aos/1176347494>
- [9] Owen, A. (1991) Empirical Likelihood for Linear Models. *The Annals of Statistics*, **19**, 1725-1747. <https://doi.org/10.1214/aos/1176348368>
- [10] Qin, J. and Lawless, J. (1994) Empirical Likelihood and General Estimating Equations. *The Annals of Statistics*, **22**, 300-325. <https://doi.org/10.1214/aos/1176325370>
- [11] Chuang, C.S. and Chan, N.H. (2002) Empirical Likelihood for Autoregressive Models, with Applications to Unstable Time Series. *Statistica Sinica*, **12**, 387-407.
- [12] Chan, N.H. and Ling, S. (2006) Empirical Likelihood for GARCH Models. *Econometric Theory*, **22**, 403-428. <https://doi.org/10.1017/s0266466606060208>
- [13] Zhang, R. and Wang, D. (2022) Bayesian Empirical Likelihood Inference for the Generalized Binomial AR(1) Model. *Journal of the Korean Statistical Society*, **51**, 977-1004. <https://doi.org/10.1007/s42952-022-00171-7>
- [14] Yang, K., Wang, D. and Li, H. (2017) Threshold Autoregression Analysis for Finite-Range Time Series of Counts with an Application on Measles Data. *Journal of Statistical Computation and Simulation*, **88**, 597-614. <https://doi.org/10.1080/00949655.2017.1400032>
- [15] Weiß, C.H. (2009) A New Class of Autoregressive Models for Time Series of Binomial Counts. *Communications in Statistics—Theory and Methods*, **38**, 447-460. <https://doi.org/10.1080/03610920802233937>
- [16] Ma, T. and Wang, D. (2023) First-Order Random Coefficient Binomial AR Process with Dependent Counting Series. *Communications in Statistics—Simulation and Computation*, **54**, 71-90. <https://doi.org/10.1080/03610918.2023.2245177>
- [17] Weiß, C.H. and Kim, H.Y. (2011) Binomial AR(1) Processes: Moments, Cumulants, and Estimation. *Statistics*, **45**, 279-295.
- [18] Klimko, L.A. and Nelson, P.I. (1978) On Conditional Least Squares Estimation for Stochastic Processes. *The Annals of Statistics*, **6**, 629-642. <https://doi.org/10.1214/aos/1176344207>
- [19] Lu, F. and Wang, D. (2021) A New Estimation for INAR(1) Process with Poisson Distribution. *Computational Statistics*, **37**, 1185-1201. <https://doi.org/10.1007/s00180-021-01157-5>
- [20] Mykland, P.A. (1995) Dual Likelihood. *The Annals of Statistics*, **23**, 396-421.