Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A novel random environment INAR(1) model with generalized signed thinning for regime-switching integer-valued time series

Wang Yunyan
Jiangxi University of Science and Technology
Tang Mingtian
tangmingtian@jxust.edu.cn

Jiangxi University of Science and Technology

Research Article

Keywords:

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11552555/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A novel random environment INAR(1) model with generalized signed thinning for regime-switching integer-valued time series

Wang Yunyan[1,2] and Tang Mingtian[1*]

[1]School of Mathematics and Physics, Jiangxi University of Science and Technology, Ganzhou, 341000, Jiangxi, China.
[2]Key Laboratory of Low Dimensional Quantum Materials and Sensor Devices of Jiangxi Education Institutes, Ganzhou, 341000, Jiangxi, China.

*Corresponding author(s). E-mail(s): tangmingtian@jxust.edu.cn;
Contributing authors: yywang@jxust.edu.cn;

Abstract: Count time series in criminology, epidemiology and economics often exhibit regime-switching behavior and may take negative values after temporal or cross-series differencing. This paper proposes a novel random environment INAR model with generalized signed thinning operator, in which the innovation term is governed by a finite-state Markov chain. The proposed model accommodates $\mathbb{Z}$-valued observations and captures regime-switching dynamics induced by unobserved environmental changes. We establish the Markovian property and ergodicity of the bivariate process and derive its conditional moments and autocovariance structure. Based on the ergodicity of the bivariate Markov chain, the strong consistency and asymptotic normality of the Yule-Walker estimators are established by applying the additive functional ergodic theorem, without requiring a closed-form marginal distribution. A Monte Carlo simulation study is conducted to assess the finite-sample performance of the estimators. An empirical application to the monthly motor vehicle theft differences between two police beats in Pittsburgh confirms the model's ability to capture regime-dependent dynamics and handle negative-valued count series.

Keywords: Regime switching; Random environment; Generalized sign thinning operator; Integer-valued time series; Strong consistency; Asymptotic normality
Mathematics Subject Classification: Primary 62M10; Secondary 60J10.

1 Introduction

Integer-valued time series data frequently arise across a variety of applied disciplines, including criminology, epidemiology, economics and finance, where stochastic dynamics and regime-switching behaviors are commonly observed. Integer-valued autoregressive (INAR) models have therefore become a standard tool for modeling such data. Since the first-order INAR(1) model of [1] was introduced via the binomial thinning operator, numerous extensions have been developed, including zero-modified INAR(1) processes [2], Pegram-based stationary models [3], integer-valued threshold autoregressive processes [4], and non-linear INAR(1) models under alternative thinning operators [5]. Comprehensive reviews are available in [6] and [7].

The random coefficient framework allows the thinning parameter to vary randomly over time. [8] proposed the first-order random coefficient INAR(1) process. Recent developments include parameter estimation and change-point tests for bivariate random coefficient INAR models [9], as well as a pth-order mixed dependence-driven random coefficient INAR model [10].

Random environment models constitute another important strand of the INAR literature, where the process dynamics are governed by a Markov chain. [11] introduced a random environment process into the random coefficient INAR model based on [8]. [12] extended this line of research by studying non-stationary INAR processes under random environments, establishing their distributional and correlation properties. [13] proposed higher-order random environment INAR models, and [14] developed a first-order random coefficient INAR model under random environment using the negative binomial thinning operator, proving strong consistency of the Yule-Walker estimators. These random environment models have also been extended to non-linear settings [15]. In addition to random environment models, threshold-based INAR models introduce regime switching through a threshold variable. For instance, [16] proposed a time-varying-threshold INAR model with an empirical quantile threshold, which allows the regime classification to adapt to recent observations and captures nonlinear dynamics in count time series.

Classical INAR models are restricted to non-negative integer-valued observations. In practice, temporal differencing is often applied to remove trends and seasonality from non-stationary count data. Cross-series differencing is also common when two related count series are compared, for example, the monthly counts of motor vehicle thefts in two police beats. The resulting differenced series remains integer-valued but may take negative values, a feature that cannot be accommodated by standard non-negative INAR models. To overcome this obstacle, [17] introduced the signed binomial thinning operator. [18] proposed the generalized signed thinning operator and established first-order random coefficient INAR models based on this operator. More recently, [19] introduced a new INAR(1) model for $\mathbb{Z}$-valued time series using the relative binomial thinning operator.

The first-order random coefficient INAR model with the generalized signed thinning operator is given by

$$X_n = \alpha_n \circledast^G X_{n-1} + \varepsilon_n, \tag{1}$$

where $\circledast^G$ denotes the generalized signed thinning operator defined as

$$\alpha_n \circledast^G X_{n-1} = \operatorname{sgn}(\alpha_n)\operatorname{sgn}(X_{n-1}) \sum_{j=1}^{|X_{n-1}|} W_j^{(n)}$$

with $\operatorname{sgn}(x) = 1$ if $x \geq 0$ and $\operatorname{sgn}(x) = -1$ if $x < 0$.

Nevertheless, the existing literature has not yet unified the random coefficient, generalized signed thinning, and random environment mechanisms within a single framework. This paper proposes a novel random environment INAR(1) model with generalized signed thinning. We extend the random disturbance term ε_n from a single white noise process to one governed by a finite-state Markov chain, thereby incorporating time-varying environmental conditions. The resulting model preserves the flexibility of random coefficients and the ability of signed thinning to handle negative values, while capturing regime-switching behavior induced by unobserved environmental changes. Unlike the threshold mechanism, which is driven by observed past values, the regime switching in the proposed model is governed by an hidden Markov chain, allowing the innovation distribution to vary with the latent environmental state.

In this paper, we establish the Markovian property, ergodicity, and derive the moment structure of the proposed process. For parameter estimation, we prove the strong consistency and asymptotic normality of the Yule-Walker estimators through a different route from existing random environment INAR models [12, 14]: instead of relying on the moment structure of a known marginal distribution, which requires the growth of maximal state-specific subsamples, we construct a bivariate Markov chain, prove its ergodicity, and apply the additive functional ergodic theorem directly to the state-specific sample moments. This framework does not require a closed-form marginal or the growth of any particular state-specific subsample, and extends naturally to signed-thinning models on $\mathbb{Z}$. We also assess the finite-sample performance via Monte Carlo simulations. As an illustration in a complex social system, we apply the proposed model to the monthly differences in motor vehicle thefts reported to two police beats in Pittsburgh. The series exhibits regime-switching dynamics and takes both positive and negative values, making it a natural candidate for the proposed $\mathbb{Z}$-valued model.

The remainder of this paper is organized as follows. Section 2 introduces the model, including the random environment process, the generalized signed thinning operator and the required assumptions. Section 3 establishes the Markov property and ergodicity, and derives the conditional moments and autocovariance structure. Section 4 develops Yule-Walker estimation and proves strong consistency and asymptotic normality. Section 5 reports Monte Carlo simulations on finite-sample performance. Section 6 applies the model to the monthly motor vehicle theft differences between two Pittsburgh police beats. Section 7 concludes.

2 Model and Assumptions

In this section, we first give some notations which will be used throughout the paper. Suppose that $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space, and $E_r = \{1, 2, \cdots, r\}$ is a finite set, where $r \in \mathbb{N}$ is a positive integer. Let $\{Z_n\}$, $n \in \mathbb{N}_0$ be an irreducible and aperiodic Markov chain defined on $(\Omega, \mathscr{F}, \mathbb{P})$, taking values in E_r. This Markov chain serves as the random environment process, which governs the time-varying distribution of the innovation term. Let $p_{ij} = \mathbb{P}(Z_n = j \mid$

$Z_{n-1} = i$) denote the one-step transition probability from state i to state j, and let $P = (p_{ij})_{r\times r}$ be the transition probability matrix.

For each $n \in \mathbb{N}$, we define

$$\varepsilon_n(Z_{n-1}, Z_n) = \sum_{i=1}^{r}\sum_{j=1}^{r} \varepsilon_n(i,j)\mathbf{1}_{\{Z_{n-1}=i, Z_n=j\}},$$

where $\mathbf{1}_{\{Z_{n-1}=i,Z_n=j\}}$ denotes the indicator function. Thus, conditional on the environmental states $Z_{n-1} = i$ and $Z_n = j$, the innovation term is given by $\varepsilon_n(i,j)$.

The new model we considered in this paper is defined as follows.

Definition 2.1 *An integer-valued stochastic process $\{X_n(Z_n), n \in \mathbb{N}_0\}$, is called a first-order random coefficient INAR model under random environment with generalized signed thinning (RERCGINARS(1)) if it satisfies*

$$X_n(Z_n) = \alpha_n \circledast^G X_{n-1}(Z_{n-1}) + \varepsilon_n(Z_{n-1}, Z_n),\ n \geq 1, \tag{2}$$

where the generalized signed-thinning operator $\circledast^G$ is defined as

$$\alpha_n \circledast^G X_{n-1}(Z_{n-1}) = \operatorname{sgn}(\alpha_n)\operatorname{sgn}(X_{n-1}(Z_{n-1}))\sum_{j=1}^{|X_{n-1}(Z_{n-1})|} W_j^{(n)}.$$

Here the sign function is given by $\operatorname{sgn}(x) = 1$ for $x \geq 0$ and $\operatorname{sgn}(x) = -1$ for $x < 0$. For each environmental state $z \in E_r$, let $X_n(z)$ denote the value of the process conditional on the event $\{Z_n = z\}$. Then the process can be decomposed as

$$X_n(Z_n) = \sum_{z=1}^{r} X_n(z)\mathbf{1}_{\{Z_n=z\}}.$$

Moreover, the following two conditions are imposed:

(i) *for all $i, j \in E_r$, $\{\varepsilon_n(i,j)\}$ are i.i.d. integer-valued random variables. The sequence $\{\alpha_n\}$ is i.i.d. on $\mathbb{R}$ and independent of $\{\varepsilon_n(i,j)\}$.*

(ii) *Given $\{\alpha_n\}$, the counting sequence $\{W_j^{(n)}\}$ consists of mutually independent and identically distributed random variables following a generalized power series distribution, and is independent of $\{\varepsilon_n(i,j)\}$ for every $i, j \in E_r$. Furthermore,*

$$E\left(W_j^{(n)} \mid \alpha_n\right) = |\alpha_n|, \qquad \operatorname{Var}\left(W_j^{(n)} \mid \alpha_n\right) = \beta_n,$$

where β_n depends only on α_n and is almost surely finite.

The proposed model encompasses several existing models as special cases:

- If the random environment is ignored (i.e., $r = 1$), the model reduces to the generalized random coefficient INAR(1) model of [18].
- If the signed thinning operator is replaced by the standard binomial thinning operator and $r = 1$, it reduces to the classical INAR(1) model of [1].
- If the random coefficient is fixed (i.e., $\alpha_n \equiv \alpha$), the model simplifies to a signed thinning INAR(1) model with random environment.
- If the random environment has a single state and α_n is fixed, the model reduces to the signed INAR(1) model of [17].

The basic assumptions that guarantee the validity of the main results throughout the paper are as follows.

(A1) The sequences $\{Z_n\}$, $\{\alpha_n\}$, $\{\varepsilon_n(1,1)\},\{\varepsilon_n(1,2)\},\ldots,\{\varepsilon_n(r,r)\}$ are mutually independent. Moreover, given $\{\alpha_n\}$, the counting sequence $\{W_j^{(n)}\}$ is independent of $\{Z_n\}$ and of all past observations $\{X_m(l), m<n\}, \forall l \in E_r$.

(A2) For all $i,j,l \in E_r$, conditional on $Z_{n-1}=i$ and $Z_n=j$, the random variables α_n and $\varepsilon_n(i,j)$ are independent of $\{X_m(l), m<n\}$.

(A3) For $\forall k,l \in E_r$, the following expressions are finite:

$$E(\varepsilon_n(k,l)) = \mu_{\varepsilon(k,l)}, \qquad Var(\varepsilon_n(k,l)) = \sigma^2_{\varepsilon(k,l)},$$

$$E(\alpha_n) = \mu_\alpha, \qquad E\,|\alpha_n| = \mu_{|\alpha|}, \qquad Var(\alpha_n) = \sigma^2_\alpha, \qquad E(\beta_n) = \beta.$$

(A4) The counting sequence $\{W_j^{(n)}\}$ has full support on $\mathbb{N}_0$, i.e., for every $k \in \mathbb{N}_0$, $\mathbb{P}(W_j^{(n)} = k \mid \alpha_n) > 0$ almost surely with respect to α_n. Moreover, the support of α_n contains an interval around zero.

(A5) The persistence probabilities satisfy $p_{kk} > 0$ for all $k \in E_r$.

(A6) For every pair $(i,j) \in E_r \times E_r$, the innovation $\varepsilon_n(i,j)$ satisfies $\mathbb{P}(\varepsilon_n(i,j)=0) > 0$. Moreover, there exists at least one transition pair $(i_0,j_0) \in E_r \times E_r$ such that $p_{i_0,j_0} > 0$, and $\varepsilon_n(i_0,j_0)$ has full support on $\mathbb{Z}$.

Remark 2.1 Assumptions (A4) and (A6) are complementary: (A4) ensures that the thinning layer can realize any integer between $-|x|$ and $|x|$, while (A6) ensures that the noise layer has a point mass at 0 and that the special transition (i_0,j_0) can fill any remaining gap. Since $\{Z_n\}$ is irreducible and aperiodic, j_0 is reachable from any $i \in E_r$, which is used in the irreducibility proof of Theorem 3.2.

3 Properties of the RERCGINARS(1) Process

In this section, we first consider the Markov property and ergodicity of the bivariate Markov chain $\{(X_n(Z_n), Z_n)\}$.

Theorem 3.1 *Suppose assumptions (A1) and (A2) hold. Then the bivariate time series $\{(X_n(Z_n), Z_n)\}_{n\geq 0}$ is a first-order Markov chain.*

Proof For any $n \in \mathbb{N}_0$ and $(x_n, z_n) \in \mathbb{Z} \times E_r$, let $A = \{(X_m(Z_m), Z_m) = (x_m, z_m), 0 \leqslant m < n-1\}$, we have

$$\begin{aligned}
&\mathbb{P}\big((X_n(Z_n), Z_n) = (x_n, z_n) \,\big|\, (X_{n-1}(Z_{n-1}), Z_{n-1}) = (x_{n-1}, z_{n-1}), A\big) \\
=&\mathbb{P}\Big(\alpha_n \circledast^G X_{n-1}(Z_{n-1}) + \varepsilon_n(Z_{n-1}, Z_n) = x_n, Z_n = z_n \,\Big|\, Z_{n-1} = z_{n-1},\, X_{n-1}(Z_{n-1}) = x_{n-1},\, A\Big) \\
=&\mathbb{P}\Big(\alpha_n \circledast^G x_{n-1} + \varepsilon_n(z_{n-1}, z_n) = x_n \,\Big|\, Z_n = z_n,\, Z_{n-1} = z_{n-1},\, X_{n-1}(Z_{n-1}) = x_{n-1},\, A\Big) \\
&\cdot \mathbb{P}\big(Z_n = z_n \,\big|\, Z_{n-1} = z_{n-1},\, X_{n-1}(Z_{n-1}) = x_{n-1},\, A\big).
\end{aligned}$$

$\{Z_n\}$ is a Markov chain, hence

$$\mathbb{P}\big(Z_n = z_n \mid Z_{n-1} = z_{n-1}, X_{n-1}(Z_{n-1}) = x_{n-1}, A\big) = \mathbb{P}(Z_n = z_n \mid Z_{n-1} = z_{n-1}) = p_{z_{n-1}, z_n}.$$

By (A1) and (A2), given $Z_{n-1} = z_{n-1}$ and $Z_n = z_n$, α_n, $\{W_j^{(n)}\}_j$ and innovation $\varepsilon_n(z_{n-1}, z_n)$ are independent of the event A, and x_{n-1} enters only as a fixed deterministic argument. Therefore

$$\begin{aligned}&\mathbb{P}\Big(\alpha_n \circledast^G x_{n-1} + \varepsilon_n(z_{n-1}, z_n) = x_n \,\Big|\, Z_n = z_n,\, Z_{n-1} = z_{n-1},\, X_{n-1}(Z_{n-1}) = x_{n-1},\, A\Big)\\ =&\mathbb{P}\Big(\alpha_n \circledast^G x_{n-1} + \varepsilon_n(z_{n-1}, z_n) = x_n\Big).\end{aligned}$$

Combining the two factors:

$$\begin{aligned}&\mathbb{P}\big((X_n(Z_n), Z_n) = (x_n, z_n) \,\big|\, (X_{n-1}, Z_{n-1}) = (x_{n-1}, z_{n-1}), A\big)\\ =&p_{z_{n-1}, z_n} \cdot \mathbb{P}\Big(\alpha_n \circledast^G x_{n-1} + \varepsilon_n(z_{n-1}, z_n) = x_n\Big).\end{aligned}$$

The right-hand side does not involve the event A. Consequently

$$\begin{aligned}&\mathbb{P}\big((X_n(Z_n), Z_n) = (x_n, z_n) \,\big|\, (X_{n-1}(Z_{n-1}), Z_{n-1}) = (x_{n-1}, z_{n-1}), A\big)\\ =&\mathbb{P}\big((X_n(Z_n), Z_n) = (x_n, z_n) \,\big|\, (X_{n-1}(Z_{n-1}), Z_{n-1}) = (x_{n-1}, z_{n-1})\big).\end{aligned}$$

This establishes the first-order Markov property for $\{(X_n(Z_n), Z_n)\}_{n\geq 0}$. □

The following lemma concerning the drift condition, which will be used to establish positive recurrence.

Lemma 3.1 (Foster-Lyapunov Drift Condition) *Let $\{Y_n\}_{n\geq 0}$ be an irreducible and aperiodic Markov chain on a countable state space $\mathscr{S}$. Suppose there exist a function $g : \mathscr{S} \to [0, +\infty)$ and constants $\lambda \in (0,1)$, $K < +\infty$ such that*

$$E[g(Y_{n+1}) \mid Y_n = y] \leq \lambda g(y) + K, \qquad \forall y \in \mathscr{S}.$$

Then $\{Y_n\}$ is positive recurrent and hence ergodic.

Proof This is the Foster-Lyapunov criterion; see, e.g., [20] (Theorem 11.3.4). □

Lemma 3.2 (Jump-channel Property) *Suppose assumptions (A4) and (A6) hold. Let $(i_0, j_0) \in E_r \times E_r$ be the special transition pair given in assumption (A6). Conditional on the environmental transition $Z_{n-1} = i_0$, $Z_n = j_0$, for any starting value $x \in \mathbb{Z}$ and any target value $x' \in \mathbb{Z}$, the event*

$$\alpha_n \circledast^G x + \varepsilon_n(i_0, j_0) = x'$$

has strictly positive probability. In other words, whenever the environment executes the transition $i_0 \to j_0$, the X-component can jump from any integer x to any integer x' in a single step.

Proof We split into two cases $x = 0$ and $x \neq 0$. If $x = 0$, then $\alpha_n \circledast^G 0 = 0$ almost surely. We require $0 + \varepsilon_n(i_0, j_0) = x'$. By assumption (A6), $\varepsilon_n(i_0, j_0)$ has full support on $\mathbb{Z}$, hence $\mathbb{P}(\varepsilon_n(i_0, j_0) = x') > 0$.

If $x \neq 0$. We aim for

$$\alpha_n \circledast^G x = x \quad \text{and} \quad \varepsilon_n(i_0, j_0) = x' - x.$$

Since $x \in \{-|x|, \ldots, |x|\}$, by assumption (A4) the thinning term can equal x with positive probability. Specifically, choose the counting sequence $\{W_j^{(n)}\}$ such that

$$\sum_{j=1}^{|x|} W_j^{(n)} = |x|,$$

which is possible because $\{W_j^{(n)}\}$ has full support on $\mathbb{N}_0$, and choose α_n such that $\operatorname{sgn}(\alpha_n) = 1$, which is possible because the support of α_n contains an interval around zero. Then

$$\operatorname{sgn}(\alpha_n)\operatorname{sgn}(x)\sum_{j=1}^{|x|} W_j^{(n)} = \operatorname{sgn}(x)|x| = x.$$

By assumption (A6), $\mathbb{P}(\varepsilon_n(i_0, j_0) = x' - x) > 0$. Under assumption (A1), α_n, $\{W_j^{(n)}\}_j$, and $\varepsilon_n(i_0, j_0)$ are mutually independent, so the joint event has strictly positive probability. Hence

$$\mathbb{P}\left(\alpha_n \circledast^G x + \varepsilon_n(i_0, j_0) = x' \mid Z_{n-1} = i_0,\, Z_n = j_0\right) > 0.$$

□

We now establish the ergodicity of the bivariate Markov chain.

Theorem 3.2 *Suppose assumptions (A1)-(A6) hold, the Markov chain $\{Z_n\}$ is irreducible and aperiodic, and the random coefficient α_n satisfies $0 < \mu_{|\alpha|} < 1$. Then the bivariate Markov chain $\{(X_n(Z_n), Z_n)\}$ is irreducible, aperiodic, and positive recurrent. Consequently, it is ergodic.*

Proof First, we prove the irreducibility of bivariate Markov chain $\{(X_n(Z_n), Z_n)\}$. Take arbitrary two states $(x,z), (x',z') \in \mathbb{Z} \times E_r$. We aim to construct a finite-length path with strictly positive transition probability from (x,z) to (x',z').

Since $\{Z_n\}$ is an irreducible finite-state Markov chain, for any $i, j \in E_r$ there exist $m < \infty$ and a sequence $i = z_0, z_1, \ldots, z_m = j$ such that

$$\prod_{k=1}^{m} p_{z_{k-1}, z_k} > 0.$$

We use this fact three times below.

(i) Reach the pre-jump state i_0. By irreducibility of $\{Z_n\}$, there exist $m_1 < \infty$ and a sequence $z = z_0, z_1, \ldots, z_{m_1} = i_0$ with

$$\prod_{k=1}^{m_1} p_{z_{k-1}, z_k} > 0.$$

During these m_1 steps, choose the innovation $\varepsilon_n(z_{k-1}, z_k) = 0$, which has positive probability by assumption (A6). The thinning term $\alpha_n \circledast^G X_{n-1}$ then takes some finite integer value, which is well-defined by assumption (A4). By assumption (A1), the random components are mutually independent, so the joint event

$$\{Z_{m_1} = i_0,\, X_{m_1} = x_{m_1}\}$$

has positive probability for some finite $x_{m_1} \in \mathbb{Z}$.

(ii) Jump to the target value x'. In the (m_1+1)-th step, use the special transition $i_0 \to j_0$, which has positive probability by assumption (A6). By Lemma 3.2, conditional on this environmental transition, the event

$$\alpha_{m_1+1} \circledast^G x_{m_1} + \varepsilon_{m_1+1}(i_0, j_0) = x'$$

has strictly positive probability. Hence the chain reaches (x', j_0) with positive probability.

(iii) Reach the target environment z'. By irreducibility of $\{Z_n\}$, there exist $m_2 < \infty$ and a sequence $j_0 = w_0, w_1, \ldots, w_{m_2} = z'$ with

$$\prod_{k=1}^{m_2} p_{w_{k-1}, w_k} > 0.$$

During these m_2 steps, keep $X = x'$ by choosing $\alpha_n \circledast^G x' = x'$, which is possible by assumption (A4) since $x' \in \{-|x'|, \ldots, |x'|\}$, and choose $\varepsilon_n(w_{k-1}, w_k) = 0$, which is possible by assumption (A6). By assumption (A1), the joint event has positive probability. Then

$$X_{m_1+1+k} = x' + 0 = x'$$

for each $k = 1, \ldots, m_2$, and the environment reaches z'. Hence the chain reaches (x', z') with positive probability.

Combining the three phases, the total number of steps is $m = m_1 + 1 + m_2 < \infty$, and every transition along the path has strictly positive probability. Therefore

$$\mathbb{P}^{(m)}\big((x,z) \to (x',z')\big) > 0.$$

Since (x,z) and (x',z') were arbitrary, the bivariate chain $\{(X_n(Z_n), Z_n)\}$ is irreducible on $\mathscr{S} = \mathbb{Z} \times E_r$.

Next, we consider the aperiodicity. Fix an arbitrary $z \in E_r$. By assumption (A5), the environmental self-transition satisfies $p_{zz} > 0$, so the event $\{Z_{n-1} = z,\ Z_n = z\}$ has positive probability. Consider the state $(0,z) \in \mathscr{S}$. Given $X_{n-1} = 0$, the generalized signed thinning operator satisfies

$$\alpha_n \circledast^G 0 = 0$$

almost surely. By assumption (A6), the innovation satisfies $\mathbb{P}(\varepsilon_n(z,z) = 0) > 0$. Under assumption (A1), α_n, $\{W_j^{(n)}\}$, $\varepsilon_n(z,z)$ and the environmental transition are mutually independent. Hence the joint event

$$\{Z_{n-1} = z,\ Z_n = z,\ \varepsilon_n(z,z) = 0\}$$

has strictly positive probability. Conditional on this event and on $X_{n-1} = 0$, we obtain

$$X_n = \alpha_n \circledast^G 0 + \varepsilon_n(z,z) = 0 + 0 = 0.$$

Hence

$$\mathbb{P}\big((X_n, Z_n) = (0,z) \mid (X_{n-1}, Z_{n-1}) = (0,z)\big) > 0.$$

That is, the state $(0,z)$ admits a one-step self-loop with strictly positive probability, so its period equals one. For an irreducible Markov chain, all states share the identical period. Since the bivariate chain is irreducible, $\{(X_n(Z_n), Z_n)\}$ is aperiodic.

Then, we prove positive recurrence via the Foster-Lyapunov drift condition. Define the Lyapunov function $V : \mathscr{S} \to [0, +\infty)$ by

$$V(x,z) = |x| + 1, \quad (x,z) \in \mathscr{S}.$$

We aim to verify the drift inequality

$$E\big[V(X_{n+1}, Z_{n+1}) \,\big|\, X_n = x, Z_n = z\big] \le \lambda\, V(x,z) + K,$$

for some constants $\lambda \in (0,1)$ and $K < +\infty$. From the model recursion,

$$X_{n+1} = \alpha_{n+1} \circledast^G x + \varepsilon_{n+1}(z, Z_{n+1}).$$

By the triangle inequality for conditional expectation,

$$E\big[|X_{n+1}| \,\big|\, X_n = x, Z_n = z\big] \le E\big[\big|\alpha_{n+1} \circledast^G x\big| \,\big|\, X_n = x, Z_n = z\big] + E\big[\big|\varepsilon_{n+1}(z, Z_{n+1})\big| \,\big|\, Z_n = z\big].$$

For the thinning term, if $x = 0$, then $E[|\alpha_{n+1} \circledast^G 0|] = 0 = |x|\mu_{|\alpha|}$. If $x \neq 0$, then

$$\big|\alpha_{n+1} \circledast^G x\big| = \big|\operatorname{sgn}(\alpha_{n+1}) \operatorname{sgn}(x) \sum_{j=1}^{|x|} W_j^{(n+1)}\big| = \sum_{j=1}^{|x|} W_j^{(n+1)}.$$

Conditioning on α_{n+1} and using assumption (A4), we obtain

$$E\left[\big|\alpha_{n+1} \circledast^G x\big| \,\middle|\, \alpha_{n+1}\right] = \sum_{j=1}^{|x|} E[W_j^{(n+1)} \mid \alpha_{n+1}] = |x|\,|\alpha_{n+1}|.$$

Taking unconditional expectation via the law of total expectation,

$$E\left[\left|\alpha_{n+1} \circledast^G x\right|\right] = |x|\,E[|\alpha_{n+1}|] = |x|\mu_{|\alpha|}.$$

Therefore, by assumptions (A3) and (A4), we have

$$E[|X_{n+1}| \mid X_n = x, Z_n = z] \le \mu_{|\alpha|}|x| + c_1,$$

where $c_1 := \max_{(i,j)\in E_r\times E_r} E\left[|\varepsilon_n(i,j)|\right] < +\infty$. Hence

$$E[V(X_{n+1}, Z_{n+1}) \mid X_n = x, Z_n = z] \le \lambda V(x,z) + K, \quad \forall (x,z) \in \mathscr{S}.$$

with $\lambda = \mu_{|\alpha|} \in (0,1)$ and $K = c_1 + 1 - \mu_{|\alpha|} < +\infty$. By Lemma 3.1, the chain is positive recurrent. Together with irreducibility and aperiodicity, this implies that the chain is ergodic. □

Moments and conditional moments play an important role in parameter estimation of the model. For the process we proposed, the following results are obtained.

Theorem 3.3 *Let $\{X_n(Z_n), n \in \mathbb{N}_0\}$ be the time series model defined by (2), where z_n is the realization of the r-states random environment process $\{Z_n\}$. Suppose assumptions (A1), (A2) and (A3) hold. Then*

(1) $E[X_{n+1}(z_{n+1})|X_n(z_n)] = \mu_\alpha X_n(z_n) + \mu_{\varepsilon(z_n,z_{n+1})}$.

(2) $Var[X_{n+1}(z_{n+1})|X_n(z_n), \alpha_{n+1}] = |X_n(z_n)|\,\beta_{n+1} + \sigma^2_{\varepsilon(z_n,z_{n+1})}$.

(3) $Var[X_{n+1}(z_{n+1})|X_n(z_n)] = X_n^2(z_n)\sigma_\alpha^2 + |X_n(z_n)|\,\beta + \sigma^2_{\varepsilon(z_n,z_{n+1})}$.

Proof (1) By the tower property, for all $n \in \mathbb{N}$,

$$E[X_{n+1}(z_{n+1}) \mid X_n(z_n)] = E\left[E(X_{n+1}(z_{n+1}) \mid X_n(z_n), \alpha_{n+1}) \mid X_n(z_n)\right].$$

Conditioning on $(X_n(z_n), \alpha_{n+1})$ and using the model recursion,

$$\begin{aligned}
&E[X_{n+1}(z_{n+1}) \mid X_n(z_n), \alpha_{n+1}] \\
=&E[\alpha_{n+1} \circledast^G X_n(z_n) + \varepsilon_{n+1}(z_n, z_{n+1}) \mid X_n(z_n), \alpha_{n+1}] \\
=&E[\alpha_{n+1} \circledast^G X_n(z_n) \mid X_n(z_n), \alpha_{n+1}] + E[\varepsilon_{n+1}(z_n, z_{n+1}) \mid X_n(z_n), \alpha_{n+1}].
\end{aligned}$$

By assumptions (A1) and (A2), we have

$$E[\varepsilon_{n+1}(z_n, z_{n+1}) \mid X_n(z_n), \alpha_{n+1}] = E[\varepsilon_{n+1}(z_n, z_{n+1})] = \mu_{\varepsilon(z_n,z_{n+1})},$$

and

$$\begin{aligned}
&E[\alpha_{n+1} \circledast^G X_n(z_n) \mid X_n(z_n), \alpha_{n+1}] \\
=&\operatorname{sgn}(X_n(z_n))\operatorname{sgn}(\alpha_{n+1}) E\left[\sum_{j=1}^{|X_n(z_n)|} W_j^{(n+1)} \mid X_n(z_n), \alpha_{n+1}\right] \\
=&\operatorname{sgn}(X_n(z_n))\operatorname{sgn}(\alpha_{n+1}) \sum_{j=1}^{|X_n(z_n)|} E[W_j^{(n+1)} \mid \alpha_{n+1}] \\
=&\operatorname{sgn}(X_n(z_n))\operatorname{sgn}(\alpha_{n+1})|X_n(z_n)|\,E[W_j^{(n+1)} \mid \alpha_{n+1}] \\
=&X_n(z_n)\operatorname{sgn}(\alpha_{n+1})|\alpha_{n+1}| \\
=&\alpha_{n+1}X_n(z_n).
\end{aligned}$$

Combining the two terms, we have

$$E[X_{n+1}(z_{n+1}) \mid X_n(z_n), \alpha_{n+1}] = \alpha_{n+1}X_n(z_n) + \mu_{\varepsilon(z_n,z_{n+1})}.$$

Using $E(\alpha_{n+1}) = \mu_\alpha$ gives

$$E[X_{n+1}(z_{n+1}) \mid X_n(z_n)] = \mu_\alpha X_n(z_n) + \mu_{\varepsilon(z_n, z_{n+1})}.$$

(2) Given $X_n(z_n)$ and α_{n+1}, the sign factors $\operatorname{sgn}(\alpha_{n+1})$ and $\operatorname{sgn}(X_n(z_n))$ are constants, and the counting variables $\{W_j^{(n+1)}\}$ are mutually independent by assumption (A1). Hence

$$\begin{aligned} Var[\alpha_{n+1} \circledast^G X_n(z_n) \mid X_n(z_n), \alpha_{n+1}] =& Var\left[\operatorname{sgn}(\alpha_{n+1})\operatorname{sgn}(X_n(z_n)) \sum_{j=1}^{|X_n(z_n)|} W_j^{(n+1)} \mid X_n(z_n), \alpha_{n+1}\right] \\ =& \sum_{j=1}^{|X_n(z_n)|} Var(W_j^{(n+1)} \mid \alpha_{n+1}) \\ =& |X_n(z_n)|\beta_{n+1}. \end{aligned}$$

Adding the innovation variance $\sigma^2_{\varepsilon(z_n, z_{n+1})}$ gives

$$Var[X_{n+1}(z_{n+1}) \mid X_n(z_n), \alpha_{n+1}] = |X_n(z_n)|\beta_{n+1} + \sigma^2_{\varepsilon(z_n, z_{n+1})}.$$

(3) By the law of total variance,

$$\begin{aligned} \operatorname{Var}[X_{n+1}(z_{n+1}) \mid X_n(z_n)] =& E\left[\operatorname{Var}(X_{n+1}(z_{n+1}) \mid X_n(z_n), \alpha_{n+1}) \mid X_n(z_n)\right] \\ &+ \operatorname{Var}\left[E(X_{n+1}(z_{n+1}) \mid X_n(z_n), \alpha_{n+1}) \mid X_n(z_n)\right]. \end{aligned} \tag{3}$$

By part (2), the first term is

$$\begin{aligned} E\left[|X_n(z_n)|\beta_{n+1} + \sigma^2_{\varepsilon(z_n, z_{n+1})} \mid X_n(z_n)\right] =& |X_n(z_n)|E[\beta_{n+1} \mid X_n(z_n)] + \sigma^2_{\varepsilon(z_n, z_{n+1})} \\ =& |X_n(z_n)|\beta + \sigma^2_{\varepsilon(z_n, z_{n+1})}, \end{aligned}$$

where we used assumption (A3) that $E[\beta_{n+1}] = \beta$ and the independence of α_{n+1} from $X_n(z_n)$.

For the second term, from part (1) we have

$$E(X_{n+1}(z_{n+1}) \mid X_n(z_n), \alpha_{n+1}) = \alpha_{n+1} X_n(z_n) + \mu_{\varepsilon(z_n, z_{n+1})}.$$

Hence

$$\operatorname{Var}\left[\alpha_{n+1} X_n(z_n) + \mu_{\varepsilon(z_n, z_{n+1})} \mid X_n(z_n)\right] = X_n^2(z_n) \operatorname{Var}(\alpha_{n+1}) = X_n^2(z_n)\sigma_\alpha^2.$$

Combining the two terms yields

$$\operatorname{Var}[X_{n+1}(z_{n+1}) \mid X_n(z_n)] = X_n^2(z_n)\sigma_\alpha^2 + |X_n(z_n)|\beta + \sigma^2_{\varepsilon(z_n, z_{n+1})}.$$

□

Theorem 3.4 *Let $\{X_n(Z_n), n \in \mathbb{N}_0\}$ be the time series defined by model (2), where z_n is the realization of the r-states random environment process $\{Z_n\}$ at time n. Suppose (A1), (A2) and (A3) hold. Conditional on this environmental path $\{z_n\}$, for any integer $k \geq 1$:*

(1) The conditional expectation of $X_{n+k}(z_{n+k})$ given $X_n(z_n)$ is

$$E[X_{n+k}(z_{n+k}) \mid X_n(z_n)] = \mu_\alpha^k X_n(z_n) + \sum_{l=0}^{k-1} \mu_\alpha^l \mu_{\varepsilon(z_{n+k-l-1}, z_{n+k-l})}.$$

(2) Let $\gamma_{n,k} = \operatorname{Cov}(X_{n+k}(z_{n+k}), X_n(z_n))$. Then

$$\gamma_{n,k} = \mu_\alpha^k \gamma_{n,0}, \qquad k \in \mathbb{N},$$

where $\gamma_{n,0} = \operatorname{Var}(X_n(z_n))$.

Proof (1) By the definition of $X_n(z_n)$, for any $k \in \mathbb{N}$, we have

$$\begin{aligned}
&E\big[X_{n+k}(z_{n+k}) \mid X_n(z_n)\big] \\
=&E\big[\alpha_{n+k} \circledast^G X_{n+k-1}(z_{n+k-1}) + \varepsilon_{n+k}(z_{n+k-1}, z_{n+k}) \mid X_n(z_n)\big] \\
=&E\big[\alpha_{n+k} \circledast^G X_{n+k-1}(z_{n+k-1}) \mid X_n(z_n)\big] + E\big[\varepsilon_{n+k}(z_{n+k-1}, z_{n+k}) \mid X_n(z_n)\big] \\
=&E\Big[E\big(\alpha_{n+k} \circledast^G X_{n+k-1}(z_{n+k-1}) \mid X_n(z_n), X_{n+k-1}(z_{n+k-1}), \alpha_{n+k}\big) \,\Big|\, X_n(z_n)\Big] \\
&+ \mu_{\varepsilon(z_{n+k-1}, z_{n+k})} \\
=&E\Big[E\Big(\operatorname{sgn}(\alpha_{n+k}) \operatorname{sgn}(X_{n+k-1}(z_{n+k-1})) \\
&\times \sum_{j=1}^{|X_{n+k-1}(z_{n+k-1})|} W_j^{(n+k)} \,\Big|\, X_n(z_n), X_{n+k-1}(z_{n+k-1}), \alpha_{n+k}\Big) \,\Big|\, X_n(z_n)\Big] \\
&+ \mu_{\varepsilon(z_{n+k-1}, z_{n+k})} \\
=&E\big[\operatorname{sgn}(\alpha_{n+k}) \operatorname{sgn}(X_{n+k-1}(z_{n+k-1})) |X_{n+k-1}(z_{n+k-1})| E(W_j^{(n+k)} \mid \alpha_{n+k}) \mid X_n(z_n)\big] \\
&+ \mu_{\varepsilon(z_{n+k-1}, z_{n+k})} \\
=&E\big[\alpha_{n+k} X_{n+k-1}(z_{n+k-1}) \mid X_n(z_n)\big] + \mu_{\varepsilon(z_{n+k-1}, z_{n+k})} \\
=&\mu_\alpha E\big[X_{n+k-1}(z_{n+k-1}) \mid X_n(z_n)\big] + \mu_{\varepsilon(z_{n+k-1}, z_{n+k})}.
\end{aligned}$$

By applying the above method $k-1$ times, we can obtain

$$\begin{aligned}
&E[X_{n+k}(z_{n+k}) | X_n(z_n)] \\
=&\mu_\alpha E[X_{n+k-1}(z_{n+k-1}) | X_n(z_n)] + \mu_{\varepsilon(z_{n+k-1}, z_{n+k})} \\
=&\mu_\alpha [\mu_\alpha E(X_{n+k-2}(z_{n+k-2}) | X_n(z_n)) + \mu_{\varepsilon(z_{n+k-2}, z_{n+k-1})}] + \mu_{\varepsilon(z_{n+k-1}, z_{n+k})} \\
=&\mu_\alpha^2 E[X_{n+k-2}(z_{n+k-2}) | X_n(z_n)] + \mu_\alpha \mu_{\varepsilon(z_{n+k-2}, z_{n+k-1})} + \mu_{\varepsilon(z_{n+k-1}, z_{n+k})} \\
=&\cdots \\
=&\mu_\alpha^k E[X_n(z_n) | X_n(z_n)] + \sum_{l=0}^{k-1} \mu_\alpha^l \mu_{\varepsilon(z_{n+k-l-1}, z_{n+k-l})} \\
=&\mu_\alpha^k X_n(z_n) + \sum_{l=0}^{k-1} \mu_\alpha^l \mu_{\varepsilon(z_{n+k-l-1}, z_{n+k-l})}.
\end{aligned}$$

(2) By part (1) of this theorem, for any $k \geq 1$,

$$E\big[X_{n+k}(z_{n+k}) \,\big|\, X_n(z_n)\big] = \mu_\alpha^k X_n(z_n) + \sum_{l=0}^{k-1} \mu_\alpha^l \, \mu_{\varepsilon(z_{n+k-l-1}, z_{n+k-l})}.$$

Denote

$$c_{n,k} := \sum_{l=0}^{k-1} \mu_\alpha^l \, \mu_{\varepsilon(z_{n+k-l-1}, z_{n+k-l})},$$

which is a constant conditional on the environmental path $\{z_n\}$. Hence

$$E\big[X_{n+k}(z_{n+k}) \,\big|\, X_n(z_n)\big] = \mu_\alpha^k X_n(z_n) + c_{n,k}.$$

By the law of total covariance,

$$\operatorname{Cov}(Y, Z) = \operatorname{Cov}\big(E[Y \mid Z], Z\big) + E\big[\operatorname{Cov}(Y, Z \mid Z)\big],$$

and the second term vanishes because $\mathrm{Cov}(Y, Z \mid Z) = 0$. Taking $Y = X_{n+k}(z_{n+k})$ and $Z = X_n(z_n)$, we obtain

$$\begin{aligned}\gamma_{n,k} &= \mathrm{Cov}\left(X_{n+k}(z_{n+k}), X_n(z_n)\right) \\ &= \mathrm{Cov}\left(E\left[X_{n+k}(z_{n+k}) \,\middle|\, X_n(z_n)\right], X_n(z_n)\right) \\ &= \mathrm{Cov}\left(\mu_\alpha^k X_n(z_n) + c_{n,k}, X_n(z_n)\right) \\ &= \mu_\alpha^k \,\mathrm{Cov}\left(X_n(z_n), X_n(z_n)\right) + \mathrm{Cov}\left(c_{n,k}, X_n(z_n)\right).\end{aligned}$$

Since $c_{n,k}$ is constant, $\mathrm{Cov}(c_{n,k}, X_n(z_n)) = 0$, and $\mathrm{Cov}(X_n(z_n), X_n(z_n)) = \mathrm{Var}(X_n(z_n)) = \gamma_{n,0}$. Therefore

$$\gamma_{n,k} = \mu_\alpha^k \gamma_{n,0}.$$

□

Remark 3.1 The results in Theorem 3.4 hold conditionally on a fixed environmental path $\{z_n\}$. When $\{Z_n\}$ is random, the unconditional autocovariance decomposes via the law of total covariance into

$$\begin{aligned}\mathrm{Cov}\left(X_{n+k}(Z_{n+k}), X_n(Z_n)\right) =& E\Big[\mathrm{Cov}\left(X_{n+k}(Z_{n+k}), X_n(Z_n) \mid \{z_n\}\right)\Big] \\ &+ \mathrm{Cov}\Big(E\left[X_{n+k}(Z_{n+k}) \mid \{z_n\}\right], E\left[X_n(Z_n) \mid \{z_n\}\right]\Big).\end{aligned}$$

By Theorem 3.4 (2), the first term equals $\mu_\alpha^k E[\gamma_{n,0}]$, while the second term is generally non-zero because the conditional means depend on the environmental path. Consequently, the unconditional autocovariance does not admit the simple multiplicative form $\mu_\alpha^k \gamma_{n,0}$, owing to serial variation in the environmental chain.

4 Yule-Walker Estimation

In this section, we develop the Yule-Walker estimation of the parameters in the proposed model $X_n(Z_n)$ defined by (2). Since the random environment process $\{Z_n\}$ makes the marginal process $\{X_n\}$ non-stationary, ergodic theorems cannot be applied directly to the whole sample of X_n. However, the bivariate chain $\{(X_n(Z_n), Z_n)\}$ is ergodic by Theorem 3.2, and the state-specific sample moments can be handled by applying the additive functional ergodic theorem to this bivariate chain. Within each consecutive block where the environment stays in state k, the process follows a stationary INAR(1) model with state-specific innovation distribution.

Let $X_1(z_1), X_2(z_2), \ldots, X_N(z_N)$ be a sample of size N from the process $\{X_n(Z_n)\}$. For each environment state $k \in E_r$, define

$$I_k = \big\{i \in \{1, 2, \ldots, N\} | Z_i = k\big\}, \quad n_k = |I_k|,$$

$$n_{k,k} = \#\big\{i \in \{1, \ldots, N-1\} : Z_i = k,\, Z_{i+1} = k\big\}.$$

Under assumption (A5), we have $p_{kk} > 0$, which ensures that $n_{k,k} \to \infty$ almost surely as $N \to \infty$. The subsample belonging to state k is

$$U_k = \big\{X_i(k) : i \in I_k\big\},$$

where the observations are ordered according to the original time indices.

We now define the pooled state-specific sample moments for the interrupted subsample U_k. Let

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i \in I_k} X_i(k),$$

$$\hat{\gamma}_0^{(k)} = \frac{1}{n_k} \sum_{i \in I_k} \big(X_i(k) - \hat{\mu}_k\big)^2,$$

$$\hat{\gamma}_1^{(k)} = \frac{1}{n_{k,k}} \sum_{\{i,i+1\} \subseteq I_k} \big(X_{i+1}(k) - \hat{\mu}_k\big)\big(X_i(k) - \hat{\mu}_k\big).$$

Theorem 4.1 *Suppose assumptions (A1)–(A6) hold and additionally* $0 < \mu_\alpha^2 + \sigma_\alpha^2 < 1$. *Then the estimators* $\hat{\mu}_k$, $\hat{\gamma}_0^{(k)}$ *and* $\hat{\gamma}_1^{(k)}$ *are strongly consistent, i.e.*

$$\hat{\mu}_k \xrightarrow{a.s.} \mu(k), \quad \hat{\gamma}_0^{(k)} \xrightarrow{a.s.} \gamma_0(k), \quad \hat{\gamma}_1^{(k)} \xrightarrow{a.s.} \gamma_1(k) \qquad \text{as } n_k \to \infty,$$

where $\mu(k) = E(X_n(k))$, $\gamma_0(k) = Var(X_n(k))$ *and* $\gamma_1(k) = Cov(X_{n+1}(k), X_n(k))$.

Proof Let $\{Y_n\} = \{(X_n(Z_n), Z_n)\}$ denote the bivariate ergodic Markov chain with stationary distribution π. For measurable functionals on the state space of Y_n, Meyn-Tweedie additive functional ergodic theorem states: if $E_\pi[|\psi(Y)|] < \infty$, then

$$\frac{1}{N} \sum_{n=1}^{N} \psi(Y_n) \xrightarrow{a.s.} E_\pi\big[\psi(Y)\big].$$

For two-step functionals $\psi(Y_n, Y_{n+1})$, apply the theorem to the augmented Markov chain (Y_n, Y_{n+1}). Under assumptions (A1)-(A4) and the ergodicity conditions of Theorem 3.2, the following moment bounds hold:

$$E_\pi\big[|X|\big] < \infty, \quad E_\pi\big[X^2\big] < \infty, \quad E_\pi\big[|X X'|\big] < \infty,$$

where (X, Z) and (X', Z') denote two consecutive draws from the stationary distribution of $\{(X_n(Z_n), Z_n)\}$.

Consistency of $\hat{\mu}_k$. Define functional $\psi_{\mu,k}(Y_n) = X_n(k)\mathbf{1}_{\{Z_n = k\}}$. Then

$$\hat{\mu}_k = \frac{\frac{1}{N}\sum_{n=1}^{N} \psi_{\mu,k}(Y_n)}{\frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{\{Z_n = k\}}}.$$

By the ergodic theorem:

$$\frac{1}{N} \sum_{n=1}^{N} \psi_{\mu,k}(Y_n) \xrightarrow{a.s.} E_\pi\big[X(k)\mathbf{1}_{\{Z=k\}}\big] = \pi_k\, \mu(k).$$

For denominator:

$$\frac{n_k}{N} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{\{Z_n = k\}} \xrightarrow{a.s.} \pi_k > 0.$$

By continuous mapping theorem for almost sure convergence (ratio map continuous at non-zero denominator):

$$\hat{\mu}_k \xrightarrow{a.s.} \frac{\pi_k \mu(k)}{\pi_k} = \mu(k).$$

Consistency of $\hat{\gamma}_0^{(k)}$. Standard algebraic decomposition for sample variance:

$$\hat{\gamma}_0^{(k)} = \frac{1}{n_k} \sum_{i \in I_k} \big(X_i(k) - \hat{\mu}_k\big)^2 = \frac{1}{n_k} \sum_{i \in I_k} \big(X_i(k) - \mu(k)\big)^2 - \big(\hat{\mu}_k - \mu(k)\big)^2.$$

The second term satisfies $(\hat{\mu}_k - \mu(k))^2 \xrightarrow{a.s.} 0$ from the consistency of $\hat{\mu}_k$. Define functional

$$\psi_{\gamma_0,k}(Y_n) = \big(X_n(k) - \mu(k)\big)^2 \mathbf{1}_{\{Z_n = k\}}.$$

Apply additive functional ergodic theorem:

$$\frac{1}{N} \sum_{n=1}^{N} \psi_{\gamma_0,k}(Y_n) \xrightarrow{a.s.} \mathbb{E}_\pi\big[(X(k) - \mu(k))^2 \mathbf{1}_{\{Z=k\}}\big] = \pi_k \gamma_0(k).$$

Divide by $n_k/N \xrightarrow{a.s.} \pi_k > 0$:

$$\frac{1}{n_k} \sum_{i \in I_k} \big(X_i(k) - \mu(k)\big)^2 \xrightarrow{a.s.} \gamma_0(k).$$

Combining two terms yields $\hat{\gamma}_0^{(k)} \xrightarrow{a.s.} \gamma_0(k)$.

Consistency of $\hat{\gamma}_1^{(k)}$. Define indicator for consecutive same state pairs:

$$J_n^{(k)} = \mathbf{1}_{\{Z_n = k,\, Z_{n+1} = k\}}, \quad n = 1, \ldots, N-1.$$

Then

$$\hat{\gamma}_1^{(k)} = \frac{\sum_{n=1}^{N-1} J_n^{(k)} \big(X_{n+1}(k) - \hat{\mu}_k\big)\big(X_n(k) - \hat{\mu}_k\big)}{\sum_{n=1}^{N-1} J_n^{(k)}}.$$

Expand the product by separating population mean $\mu(k)$ and random deviation $\hat{\mu}_k - \mu(k)$:

$$\begin{aligned}
&J_n^{(k)} \big(X_{n+1}(k) - \hat{\mu}_k\big)\big(X_n(k) - \hat{\mu}_k\big) \\
&= J_n^{(k)} \big(X_{n+1}(k) - \mu(k)\big)\big(X_n(k) - \mu(k)\big) \\
&\quad - (\hat{\mu}_k - \mu(k)) J_n^{(k)} \Big[\big(X_{n+1}(k) - \mu(k)\big) + \big(X_n(k) - \mu(k)\big)\Big] \\
&\quad + (\hat{\mu}_k - \mu(k))^2 J_n^{(k)}.
\end{aligned}$$

Define two-step main term functional:

$$\psi_{\gamma_1,k}(Y_n, Y_{n+1}) = J_n^{(k)} \big(X_{n+1}(k) - \mu(k)\big)\big(X_n(k) - \mu(k)\big).$$

Apply additive functional ergodic theorem to augmented chain (Y_n, Y_{n+1}):

$$\frac{1}{N-1} \sum_{n=1}^{N-1} \psi_{\gamma_1,k}(Y_n, Y_{n+1}) \xrightarrow{a.s.} E_\pi\big[\psi_{\gamma_1,k}(Y, Y')\big] = \pi_k p_{kk} \gamma_1(k).$$

Denominator:

$$\frac{n_{k,k}}{N-1} = \frac{1}{N-1} \sum_{n=1}^{N-1} J_n^{(k)} \xrightarrow{a.s.} \pi_k p_{kk} > 0.$$

Thus the main term ratio converges almost surely:

$$\frac{\frac{1}{N-1} \sum \psi_{\gamma_1,k}}{\frac{n_{k,k}}{N-1}} \xrightarrow{a.s.} \frac{\pi_k p_{kk} \gamma_1(k)}{\pi_k p_{kk}} = \gamma_1(k).$$

The two remainder terms both contain factor $\hat{\mu}_k - \mu(k) \xrightarrow{a.s.} 0$, while the remaining sample averages stay almost surely bounded under the moment conditions. Hence remainder terms converge to zero almost surely. Altogether:

$$\hat{\gamma}_1^{(k)} \xrightarrow{a.s.} \gamma_1(k).$$

□

Yule-Walker estimator for μ_α

From Theorem 3.4 (2) applied to a block with constant environment k, we have $\gamma_1(k) = \mu_\alpha \gamma_0(k)$. Hence we define the Yule-Walker estimator of μ_α as

$$\hat{\mu}_\alpha = \sum_{k=1}^{r} \frac{n_k}{N} \frac{\hat{\gamma}_1^{(k)}}{\hat{\gamma}_0^{(k)}}. \tag{4}$$

Since $\gamma_0(k) = \mathrm{Var}(X_n(k)) > 0$ under model distributional assumptions, the ratio is well-defined in population.

Theorem 4.2 *Suppose assumptions (A1)–(A6) hold, and additionally* $0 < \mu_\alpha^2 + \sigma_\alpha^2 < 1$. *Then* $\hat{\mu}_\alpha$ *is strongly consistent:*

$$\hat{\mu}_\alpha \xrightarrow{a.s.} \mu_\alpha \quad \text{as } N \to \infty.$$

Proof By Theorem 4.1, for each $k \in E_r$,

$$\hat{\mu}_{\alpha,k} = \frac{\hat{\gamma}_1^{(k)}}{\hat{\gamma}_0^{(k)}} \xrightarrow{\text{a.s.}} \frac{\gamma_1(k)}{\gamma_0(k)} = \mu_\alpha \quad \text{as } n_k \to \infty.$$

Let Ω_k denote the null-probability set on which $\hat{\mu}_{\alpha,k} \nrightarrow \mu_\alpha$. Define $\Omega = \bigcup_{k=1}^{r} \Omega_k$, then $P(\Omega) \le \sum_{k=1}^{r} P(\Omega_k) = 0$. On Ω^c, every $\hat{\mu}_{\alpha,k} \to \mu_\alpha$.

By environment chain ergodic theorem:

$$\frac{n_k}{N} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{\{Z_n = k\}} \xrightarrow{a.s.} \pi_k, \qquad \sum_{k=1}^{r} \pi_k = 1.$$

On the probability-one set Ω^c:

$$\hat{\mu}_\alpha = \sum_{k=1}^{r} \frac{n_k}{N} \hat{\mu}_{\alpha,k} \xrightarrow{a.s.} \sum_{k=1}^{r} \pi_k \mu_\alpha = \mu_\alpha \sum_{k=1}^{r} \pi_k = \mu_\alpha.$$

□

We next derive the asymptotic distribution of $\hat{\mu}_\alpha$. Under the same assumptions, the delta method applied to the state-specific sample moments yields the asymptotic normality of $\hat{\mu}_\alpha$.

Theorem 4.3 (Asymptotic Normality) *Suppose assumptions (A1)-(A6) hold, assume that* $0 < \mu_\alpha^2 + \sigma_\alpha^2 < 1$, $E[\alpha_n^4] < 1$, $E[(W_j^{(n)})^4] < \infty$ *and* $E|\varepsilon_n(k,l)|^4 < \infty$ *for all* $(k,l) \in E_r \times E_r$. *Assume further that* $\gamma_0(k) = \mathrm{Var}(X_n(k)) > 0$ *for all* $k \in E_r$. *Then, as* $N \to \infty$,

$$\sqrt{N}(\hat{\mu}_\alpha - \mu_\alpha) \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{as}}^2),$$

where

$$\sigma_{\text{as}}^2 = \sum_{k=1}^{r} \pi_k^2 \sigma_k^2 + \sum_{k \neq l} \pi_k \pi_l \sigma_{kl}.$$

Here $V = (\sigma_{kl})_{1 \le k,l \le r}$ *is the covariance matrix of the limiting normal distribution of the random vector* $\sqrt{N}(\hat{\mu}_\alpha - \mu_\alpha \mathbf{1}_r)$, *with* $\hat{\mu}_\alpha = (\hat{\mu}_{\alpha,1}, \ldots, \hat{\mu}_{\alpha,r})^\top$. *In particular,* $\sigma_k^2 = \sigma_{kk}$, *and* $\hat{\mu}_{\alpha,k} = \hat{\gamma}_1^{(k)} / \hat{\gamma}_0^{(k)}$.

Proof The proof proceeds in four steps.

Step 1: Joint asymptotic normality of the state-specific moments. Let $\{(X_n(Z_n), Z_n)\}$ denote the bivariate chain, which is ergodic by Theorem 3.2. For each $k \in E_r$, define the state-specific moment functionals

$$\psi_{1,k}(Y_n) = J_n^{(k)} X_{n+1}(k) X_n(k), \qquad \psi_{0,k}(Y_n) = J_n^{(k)} X_n^2(k),$$

where $J_n^{(k)} = \mathbf{1}_{\{Z_n=k, Z_{n+1}=k\}}$. These are additive functionals of the augmented chain $\{(Y_n, Y_{n+1})\} = \{(X_n, Z_n, X_{n+1}, Z_{n+1})\}$, which is ergodic under the same assumptions. Under the assumed fourth-moment conditions, a fourth-order moment recursion analogous to the drift condition in Theorem 3.2 yields $E_\pi[X^4] < \infty$. Let (X, Z) and (X', Z') denote two consecutive draws from the stationary distribution of $\{(X_n(Z_n), Z_n)\}$. Then, by Cauchy-Schwarz Inequality,

$$E_\pi[X^2 X'^2] \le \sqrt{E_\pi[X^4] E_\pi[(X')^4]} = E_\pi[X^4] < \infty,$$

and hence the augmented chain has finite fourth moments.

Therefore, the multivariate central limit theorem for additive functionals of ergodic Markov chains [20, Theorem 17.0.1] yields

$$\sqrt{N}\left(\hat{\gamma}_1^{(1)} - \gamma_1(1), \ldots, \hat{\gamma}_1^{(r)} - \gamma_1(r), \hat{\gamma}_0^{(1)} - \gamma_0(1), \ldots, \hat{\gamma}_0^{(r)} - \gamma_0(r)\right) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where Σ is the $2r \times 2r$ long-run covariance matrix of the augmented chain, whose (i, j)-entry is

$$\Sigma_{ij} = \sum_{m=-\infty}^{\infty} Cov\left(\psi_i(Y_0), \psi_j(Y_m)\right),$$

with ψ_i denoting the corresponding functional.

Step 2: Delta method for each state-specific ratio. Recall that under the model specification, $E(\alpha_n) = \mu_\alpha$ does not depend on the environmental state Z_n. By Theorem 3.4 (2), the state-specific ratio satisfies

$$\frac{\gamma_1(k)}{\gamma_0(k)} = \mu_\alpha, \qquad \forall k \in E_r.$$

Hence each state-specific ratio estimator $\hat{\mu}_{\alpha,k}$ targets the same common parameter μ_α. For each $k \in E_r$, define the smooth function $g_k(a, b) = a/b$, which is differentiable at$(\gamma_1(k), \gamma_0(k))$ with gradient

$$\nabla g_k = \left(\frac{1}{\gamma_0(k)}, -\frac{\gamma_1(k)}{(\gamma_0(k))^2}\right).$$

Since $\gamma_0(k) > 0$, the delta method gives

$$\sqrt{N}\left(\hat{\mu}_{\alpha,k} - \mu_\alpha\right) \xrightarrow{d} \mathcal{N}(0, \sigma_k^2),$$

where $\sigma_k^2 = \nabla g_k^\top \Sigma_k \nabla g_k$, and Σ_k is the 2×2 submatrix of Σ corresponding to state k.

Step 3: Joint asymptotic normality of the ratio vector. Applying the multivariate delta method to the function $\mathbf{g} = (g_1, \ldots, g_r)$, which has Jacobian

$$J = \mathrm{diag}(\nabla g_1, \ldots, \nabla g_r),$$

yields

$$\sqrt{N}\left(\hat{\mu}_{\alpha,1} - \mu_\alpha, \ldots, \hat{\mu}_{\alpha,r} - \mu_\alpha\right) \xrightarrow{d} \mathcal{N}(0, J\Sigma J^\top).$$

Denote this asymptotic covariance matrix by V. Its entries satisfy $V_{kk} = \sigma_k^2$ and $V_{kl} = \sigma_{kl}$ for $k \neq l$.

Step 4: Slutsky's theorem for the pooled estimator. The pooled estimator is

$$\hat{\mu}_\alpha = \sum_{k=1}^{r} \frac{n_k}{N} \hat{\mu}_{\alpha,k}.$$

Since $\{Z_n\}$ is ergodic, the occupation frequencies satisfy

$$\frac{n_k}{N} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{\{Z_n=k\}} \xrightarrow{a.s.} \pi_k.$$

Therefore, by the Cramér-Wold device and Slutsky's theorem,

$$\sqrt{N}(\hat{\mu}_\alpha - \mu_\alpha) = \sum_{k=1}^{r} \frac{n_k}{N}\sqrt{N}(\hat{\mu}_{\alpha,k} - \mu_\alpha) \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{as}}^2),$$

where

$$\sigma_{\text{as}}^2 = \sum_{k=1}^{r} \pi_k^2 \sigma_k^2 + \sum_{k\neq l} \pi_k \pi_l \sigma_{kl}.$$

□

Remark 4.1 A consistent estimator $\hat{\sigma}_{\text{as}}^2$ of the asymptotic variance can be obtained by replacing the population quantities $(\gamma_1(k), \gamma_0(k), \Sigma_k, \pi_k)$ with their sample analogues. Alternatively, the parametric bootstrap described in Section 6 provides a finite-sample approximation, which is particularly useful when some environmental states have small occupancy n_k.

Estimation of innovation means $\mu_{\varepsilon(i,j)}$

For any pair of environment states $(i, j) \in E_r \times E_r$, consider a transition from i to j. Taking unconditional expectations in Theorem 3.3(1) gives

$$\mu_j = \mu_\alpha \mu_i + \mu_{\varepsilon(i,j)},$$

where $\mu_i = E[X_n(i)]$ and $\mu_j = E[X_n(j)]$. Replacing the population quantities by their strongly consistent estimators (see Theorem 4.1 and Theorem 4.2) we obtain

$$\hat{\mu}_{\varepsilon(i,j)} = \hat{\mu}_j - \hat{\mu}_\alpha \hat{\mu}_i.$$

Corollary 4.1 *Under assumptions (A1)–(A5), for every $(i, j) \in E_r \times E_r$,*

$$\hat{\mu}_{\varepsilon(i,j)} \xrightarrow{a.s.} \mu_{\varepsilon(i,j)}, \qquad N \to \infty.$$

Proof $\hat{\mu}_i \xrightarrow{a.s.} \mu_i$, $\hat{\mu}_j \xrightarrow{a.s.} \mu_j$, $\hat{\mu}_\alpha \xrightarrow{a.s.} \mu_\alpha$. Apply continuous mapping theorem for almost sure convergence to the mapping $(a, b, c) \mapsto b - c \cdot a$. □

Remark 4.2 The asymptotic theory established above holds as $N \to \infty$. In finite samples, if some environmental state has small occupancy n_k or few consecutive same-state pairs $n_{k,k}$, the finite-sample variance of $\hat{\gamma}_1^{(k)}$ and hence of $\hat{\mu}_{\alpha,k}$ can be large, even though strong consistency holds asymptotically. This is especially relevant when environment switching is very frequent (i.e., p_{kk} small).

5 Simulations

In this section, we investigate the finite-sample performance of the Yule-Walker estimators via Monte Carlo simulation. The data are generated from the RERCGINARS(1) process defined in (2). The simulation proceeds as follows.

(i) The environment chain $\{Z_n\}$ is generated from the initial distribution π_0 and the transition probability matrix P. Specifically, Z_1 is drawn from π_0, and for $n \geq 2$, Z_n is drawn from the conditional distribution given Z_{n-1} according to the corresponding row of P.

(ii) The random coefficient α_n is drawn independently from a uniform distribution on (c,d), where $c,d \in \mathbb{R}$ with $c < d$. We set $\mu_\alpha = E(\alpha_n) = (c+d)/2$.

(iii) Given $\{\alpha_n\}$, the counting sequence $\{W_j^{(n)}\}$ consists of independent and identically distributed random variables with probability mass function

$$P(W_j^{(n)} = x) = \frac{|\alpha_n|^x}{(1+|\alpha_n|)^{x+1}}, \qquad x \in \mathbb{N}_0,$$

which is a special case of the generalized power series distribution with mean $|\alpha_n|$, namely the geometric distribution with success probability $1/(1+|\alpha_n|)$.

(iv) For a transition from environment state i to state j, the innovation term $\varepsilon_n(i,j)$ is generated as the difference of two independent Poisson random variables:

$$\varepsilon_n(i,j) = P_1 - P_2,$$

where $P_1 \sim \text{Poisson}(\lambda_1), P_2 \sim \text{Poisson}(\lambda_2)$, with P_1 and P_2 independent. Let

$$\mu_{\varepsilon(i,j)} = \mu_j - \mu_\alpha \mu_i$$

denote the target mean. To allow flexible noise intensity, we introduce a scaling parameter $\tau \geq 1$ and set the variance of $\varepsilon_n(i,j)$ to $\tau^2|\mu_{\varepsilon(i,j)}|$. The Poisson rates are then chosen as

$$\lambda_1 = \frac{\tau^2|\mu_{\varepsilon(i,j)}| + \mu_{\varepsilon(i,j)}}{2}, \qquad \lambda_2 = \frac{\tau^2|\mu_{\varepsilon(i,j)}| - \mu_{\varepsilon(i,j)}}{2}.$$

This ensures $\lambda_1, \lambda_2 \geq 0$ for all $\tau \geq 1$. Consequently,

$$E[\varepsilon_n(i,j)] = \lambda_1 - \lambda_2 = \mu_{\varepsilon(i,j)} = \mu_j - \mu_\alpha \mu_i,$$

and

$$\text{Var}[\varepsilon_n(i,j)] = \lambda_1 + \lambda_2 = \tau^2|\mu_{\varepsilon(i,j)}|.$$

The case $\tau = 1$ corresponds to the minimal variance achievable by a Poisson difference, while larger values of τ produce heavier noise.

For each replication, we set the initial value $X_0 = 0$ and discard the first $B = 500$ observations as burn-in to eliminate initial transients. All random components are drawn independently across replications.

We consider five different parameter configurations. For each configuration, sample sizes $N = 100, 200, 500, 1000$ are used, and the simulation results are based on 500 independent replications.

The five cases are specified as follows. For cases with multiple τ values, the simulation is repeated independently for each value.

Case (a): $E_r = \{1,2\}$, $\mu = (\mu_1, \mu_2) = (1,2)$, $\alpha_n \sim U(-0.05, 0.15)$ (so $\mu_\alpha = 0.05$), noise scaling parameter $\tau = 1.0$, initial distribution $\pi_0 = (0.8, 0.2)$, transition probability matrix

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix}.$$

Case (b): $E_r = \{1,2\}$, $\mu = (\mu_1, \mu_2) = (-0.5, 1.0)$, $\alpha_n \sim U(-0.6, -0.2)$ (so $\mu_\alpha = -0.40$), noise scaling parameter $\tau \in \{1.0, 2.0, 3.0\}$, initial distribution $\pi_0 = (0.5, 0.5)$, transition probability matrix

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}.$$

Case (c): $E_r = \{1,2\}$, $\mu = (\mu_1, \mu_2) = (0.1, -0.7)$, $\alpha_n \sim U(-0.55, -0.35)$ (so $\mu_\alpha = -0.45$), noise scaling parameter $\tau = 1.0$, initial distribution $\pi_0 = (0.5, 0.5)$, transition probability matrix

$$P = \begin{pmatrix} 0.4 & 0.6 \\ 0.7 & 0.3 \end{pmatrix}.$$

Case (d): $E_r = \{1,2,3\}$, $\mu = (\mu_1, \mu_2, \mu_3) = (0.1, 0.3, -0.8)$, $\alpha_n \sim U(-0.55, -0.35)$ (so $\mu_\alpha = -0.45$), noise scaling parameter $\tau \in \{1.0, 2.0, 3.0\}$, initial distribution $\pi_0 = (0.33, 0.33, 0.34)$, transition probability matrix

$$P = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.2 & 0.2 & 0.6 \end{pmatrix}.$$

Case (e): $E_r = \{1,2,3\}$, $\mu = (\mu_1, \mu_2, \mu_3) = (1,2,3)$, $\alpha_n \sim U(0.10, 0.30)$ (so $\mu_\alpha = 0.20$), noise scaling parameter $\tau \in \{1.0, 2.0, 3.0\}$, initial distribution $\pi_0 = (0.4, 0.3, 0.3)$, transition probability matrix

$$P = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}.$$

Figure 1 displays sample paths of length $N = 500$ for the five simulation cases, all generated with $\tau = 1.0$, with each observation coloured according to its latent state Z_n. Cases (a) and (e) remain mostly positive and oscillate around their state means without prolonged excursions, whereas Cases (b)-(d) repeatedly cross the zero line with rapid sign alternation. Among the two-state cases, Case (c) shows the most frequent colour alternation, reflecting its low persistence probabilities. In Case (d), the purple points (state 3) cluster in the negative region with large fluctuations while the blue and orange points remain close to zero. In Case (e), all three states are positive but the magnitude of fluctuations increases with the state index, consistent with larger state means and higher innovation variance.

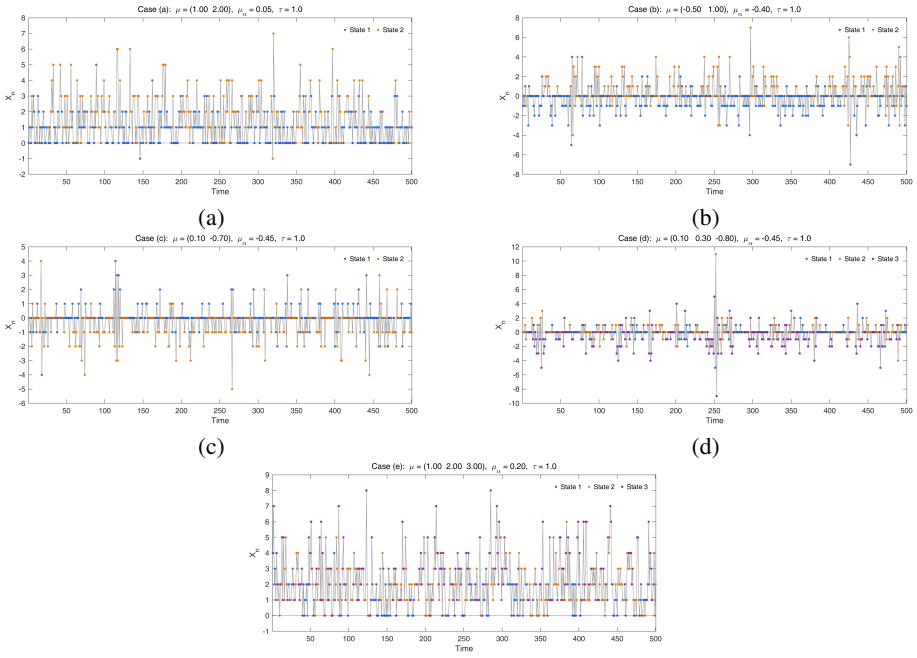


Fig. 1: Sample path diagram with a sample size of 500.

Table 1: Simulation results for Case (a): $\mu = (1,2)$, $\mu_\alpha = 0.05$, $\tau = 1.0$.

Parameter	Statistic	$N = 100$	$N = 200$	$N = 500$	$N = 1000$
μ_1	Mean	0.9960	0.9971	1.0024	0.9984
	Bias	-0.0040	-0.0029	0.0024	-0.0016
	SD	0.1300	0.0911	0.0546	0.0389
	MSE	0.0169	0.0083	0.0030	0.0015
μ_2	Mean	2.0053	1.9918	2.0037	2.0056
	Bias	0.0053	-0.0082	0.0037	0.0056
	SD	0.2503	0.1846	0.1166	0.0798
	MSE	0.0626	0.0341	0.0136	0.0064
μ_α	Mean	0.0273	0.0383	0.0461	0.0472
	Bias	-0.0227	-0.0117	-0.0039	-0.0028
	SD	0.1198	0.0890	0.0542	0.0390
	MSE	0.0148	0.0080	0.0029	0.0015

The simulation results are summarized in Tables 1-5. For each parameter, we report the sample mean of the estimates over 500 replications, the bias (mean minus true value), the sample standard deviation (SD), and the mean squared error (MSE), which equals $\text{Bias}^2 + \text{SD}^2$. Across all configurations, the mean squared errors (MSEs) decrease steadily as the sample size N increases, which is consistent with the strong consistency established in Theorem 4.2.

Table 2: Simulation results for Case (b): $\mu = (-0.5, 1.0)$, $\mu_\alpha = -0.40$.

τ	Parameter	Statistic	$N = 100$	$N = 200$	$N = 500$	$N = 1000$
1.0	μ_1	Mean	-0.4977	-0.5059	-0.4987	-0.4992
		Bias	0.0023	-0.0059	0.0013	0.0008
		SD	0.1122	0.0776	0.0480	0.0355
		MSE	0.0126	0.0060	0.0023	0.0013
	μ_2	Mean	1.0002	0.9976	0.9971	0.9943
		Bias	0.0002	-0.0024	-0.0029	-0.0057
		SD	0.1883	0.1313	0.0772	0.0543
		MSE	0.0354	0.0172	0.0059	0.0030
	μ_α	Mean	-0.3736	-0.3810	-0.3875	-0.3949
		Bias	0.0264	0.0190	0.0125	0.0051
		SD	0.1331	0.1002	0.0658	0.0473
		MSE	0.0184	0.0104	0.0045	0.0023
2.0	μ_1	Mean	-0.4988	-0.5081	-0.5045	-0.5017
		Bias	0.0012	-0.0081	-0.0045	-0.0017
		SD	0.1961	0.1304	0.0867	0.0595
		MSE	0.0384	0.0170	0.0075	0.0035
	μ_2	Mean	1.0138	1.0210	1.0005	1.0020
		Bias	0.0138	0.0210	0.0005	0.0020
		SD	0.3218	0.2268	0.1403	0.1001
		MSE	0.1035	0.0518	0.0196	0.0100
	μ_α	Mean	-0.3756	-0.3875	-0.3983	-0.3941
		Bias	0.0244	0.0125	0.0017	0.0059
		SD	0.1287	0.0912	0.0611	0.0435
		MSE	0.0171	0.0085	0.0037	0.0019
3.0	μ_1	Mean	-0.4959	-0.5022	-0.4964	-0.4984
		Bias	0.0041	-0.0022	0.0036	0.0016
		SD	0.2767	0.1985	0.1212	0.0838
		MSE	0.0764	0.0393	0.0147	0.0070
	μ_2	Mean	0.9734	0.9967	1.0009	1.0067
		Bias	-0.0266	-0.0033	0.0009	0.0067
		SD	0.4819	0.3052	0.2089	0.1532
		MSE	0.2324	0.0930	0.0436	0.0235
	μ_α	Mean	-0.3912	-0.3961	-0.3943	-0.3996
		Bias	0.0088	0.0039	0.0057	0.0004
		SD	0.1183	0.0927	0.0601	0.0414
		MSE	0.0140	0.0086	0.0036	0.0017

The state-specific means μ_k are estimated with negligible bias in all cases, and the bias diminishes further as N grows.

Table 3: Simulation results for Case (c): $\mu = (0.1, -0.7)$, $\mu_\alpha = -0.45$, $\tau = 1.0$ (frequent regime switching).

Parameter	Statistic	$N = 100$	$N = 200$	$N = 500$	$N = 1000$
μ_1	Mean	0.0902	0.1036	0.1018	0.1010
	Bias	−0.0098	0.0036	0.0018	0.0010
	SD	0.1178	0.0810	0.0510	0.0370
	MSE	0.0139	0.0066	0.0026	0.0014
μ_2	Mean	−0.6959	−0.7024	−0.7003	−0.6985
	Bias	0.0041	−0.0024	−0.0003	0.0015
	SD	0.1645	0.1134	0.0711	0.0514
	MSE	0.0270	0.0128	0.0050	0.0026
μ_α	Mean	−0.3864	−0.4239	−0.4353	−0.4433
	Bias	0.0636	0.0261	0.0147	0.0067
	SD	0.2502	0.1842	0.1328	0.1000
	MSE	0.0665	0.0345	0.0178	0.0100

For Case (a) with $\mu_\alpha = 0.05$, the estimator $\hat{\mu}_\alpha$ exhibits a small negative bias at $N = 100$ (-0.0227) and $N = 200$ (-0.0117), which essentially vanishes at $N = 1000$ (-0.0028). The MSE decreases from 0.0148 at $N = 100$ to 0.0015 at $N = 1000$. This confirms that under weak positive autocorrelation and balanced state probabilities, the Yule-Walker estimator performs well even for moderate sample sizes.

In the negative μ_α cases (b), (c), and (d), a small positive finite-sample bias is observed, which diminishes as N increases. For instance, in Case (d) with $\tau = 1.0$, the bias of $\hat{\mu}_\alpha$ is 0.0632 at $N = 100$ but reduces to 0.0070 at $N = 1000$; the corresponding MSE drops from 0.0291 to 0.0040. This positive bias is more pronounced when regime switching is frequent, as in Case (c), where the bias at $N = 100$ is 0.0636 and the MSE is 0.0665. These findings empirically support Remark 4.2 that a small number of consecutive same-state pairs $n_{k,k}$ can substantially inflate the finite-sample variance of $\hat{\gamma}_1^{(k)}$ and hence of $\hat{\mu}_\alpha$.

The noise scaling parameter τ affects the estimator in different ways depending on the target parameter. For the state-specific means μ_k, increasing τ inflates the innovation variance and hence the variance of $X_n(k)$. Since $\hat{\mu}_k$ is a simple state-wise sample average, its sampling variance is approximately proportional to $\mathrm{Var}(X_n(k))/n_k$. Consequently, both the standard deviation and MSE of $\hat{\mu}_k$ grow markedly as τ increases. For instance, in Case (d) at $N = 100$, the MSE of μ_3 rises from 0.0376 at $\tau = 1.0$ to 0.2389 at $\tau = 3.0$, a more than six-fold increase.

In contrast, the finite-sample performance of $\hat{\mu}_\alpha$ behaves quite differently. The following mechanism applies specifically to the negative μ_α case, where the finite-sample bias is positive and dominates the MSE at small τ. Although the innovation variance increases with τ, the overall MSE of $\hat{\mu}_\alpha$ decreases in Cases (b)-(d). For example, in Case (d) at $N = 100$, the MSE of $\hat{\mu}_\alpha$ drops from 0.0291 at $\tau = 1.0$ to 0.0196 at $\tau = 3.0$. This pattern arises because, intuitively, the positive finite-sample bias of the ratio-type estimator $\hat{\gamma}_1^{(k)}/\hat{\gamma}_0^{(k)}$ is reduced when the state-specific variance $\gamma_0(k)$ becomes larger. As noise grows, the reduction in squared bias outweighs the increase in variance. In short, higher innovation noise can mitigate the finite-sample bias of the Yule-Walker ratio estimator under negative autocorrelation, while

Table 4: Simulation results for Case (d): $\mu = (0.1, 0.3, -0.8)$, $\mu_\alpha = -0.45$.

τ	Parameter	Statistic	$N = 100$	$N = 200$	$N = 500$	$N = 1000$
1.0	μ_1	Mean	0.0988	0.1067	0.0972	0.1001
		Bias	−0.0012	0.0067	−0.0028	0.0001
		SD	0.1172	0.0749	0.0465	0.0348
		MSE	0.0137	0.0057	0.0022	0.0012
	μ_2	Mean	0.2963	0.2946	0.3018	0.3005
		Bias	−0.0037	−0.0054	0.0018	0.0005
		SD	0.1360	0.0973	0.0600	0.0423
		MSE	0.0185	0.0095	0.0036	0.0018
	μ_3	Mean	−0.7978	−0.7948	−0.7973	−0.7976
		Bias	0.0022	0.0052	0.0027	0.0024
		SD	0.1941	0.1303	0.0800	0.0568
		MSE	0.0376	0.0170	0.0064	0.0032
	μ_α	Mean	−0.3868	−0.4027	−0.4277	−0.4430
		Bias	0.0632	0.0473	0.0223	0.0070
		SD	0.1585	0.1220	0.0866	0.0627
		MSE	0.0291	0.0171	0.0080	0.0040
2.0	μ_1	Mean	0.0955	0.1094	0.1055	0.1008
		Bias	−0.0045	0.0094	0.0055	0.0008
		SD	0.2051	0.1403	0.0836	0.0560
		MSE	0.0420	0.0197	0.0070	0.0031
	μ_2	Mean	0.3133	0.2983	0.3013	0.3057
		Bias	0.0133	−0.0017	0.0013	0.0057
		SD	0.2301	0.1655	0.1036	0.0727
		MSE	0.0530	0.0273	0.0107	0.0053
	μ_3	Mean	−0.7763	−0.7905	−0.7950	−0.8065
		Bias	0.0237	0.0095	0.0050	−0.0065
		SD	0.3450	0.2323	0.1426	0.1064
		MSE	0.1194	0.0539	0.0203	0.0113
	μ_α	Mean	−0.4134	−0.4295	−0.4399	−0.4456
		Bias	0.0366	0.0205	0.0101	0.0044
		SD	0.1442	0.1066	0.0713	0.0495
		MSE	0.0221	0.0118	0.0052	0.0025
3.0	μ_1	Mean	0.0975	0.1115	0.1029	0.0981
		Bias	−0.0025	0.0115	0.0029	−0.0019
		SD	0.2829	0.1822	0.1094	0.0785
		MSE	0.0799	0.0333	0.0119	0.0062
	μ_2	Mean	0.3075	0.3091	0.3015	0.3044
		Bias	0.0075	0.0091	0.0015	0.0044
		SD	0.3362	0.2411	0.1427	0.1041
		MSE	0.1129	0.0581	0.0203	0.0108
	μ_3	Mean	−0.8224	−0.8178	−0.7931	−0.7989
		Bias	−0.0224	−0.0178	0.0069	0.0011
		SD	0.4888	0.3264	0.2061	0.1455
		MSE	0.2389	0.1067	0.0424	0.0211
	μ_α	Mean	−0.4284	−0.4441	−0.4408	−0.4487
		Bias	0.0216	0.0059	0.0092	0.0013
		SD	0.1383	0.1002	0.0652	0.0472
		MSE	0.0196	0.0101	0.0043	0.0022

increasing the sampling variability of the state-mean estimates. By contrast, in the positive μ_α case (e), the bias is negative and small, and the MSE of $\hat{\mu}_\alpha$ is nearly insensitive to τ.

These results have practical implications. If the primary target of inference is the autoregressive coefficient μ_α (especially for negative mean-reverting values), moderate increases in

Table 5: Simulation results for Case (e): $\mu = (1,2,3)$, $\mu_\alpha = 0.20$.

τ	Parameter	Statistic	$N = 100$	$N = 200$	$N = 500$	$N = 1000$
1.0	μ_1	Mean	1.0055	1.0037	0.9996	1.0013
		Bias	0.0055	0.0037	−0.0004	0.0013
		SD	0.2410	0.1662	0.1010	0.0738
		MSE	0.0580	0.0276	0.0102	0.0054
	μ_2	Mean	2.0219	1.9919	1.9932	2.0020
		Bias	0.0219	−0.0081	−0.0068	0.0020
		SD	0.2619	0.2004	0.1190	0.0815
		MSE	0.0689	0.0401	0.0142	0.0066
	μ_3	Mean	3.0052	3.0007	2.9867	3.0053
		Bias	0.0052	0.0007	−0.0133	0.0053
		SD	0.3512	0.2528	0.1528	0.1129
		MSE	0.1232	0.0638	0.0235	0.0127
	μ_α	Mean	0.1502	0.1725	0.1907	0.1936
		Bias	−0.0498	−0.0275	−0.0093	−0.0064
		SD	0.1315	0.1020	0.0624	0.0444
		MSE	0.0197	0.0111	0.0040	0.0020
2.0	μ_1	Mean	1.0158	0.9842	0.9918	1.0006
		Bias	0.0158	−0.0158	−0.0082	0.0006
		SD	0.4116	0.2810	0.1771	0.1327
		MSE	0.1693	0.0790	0.0314	0.0176
	μ_2	Mean	1.9690	2.0205	2.0068	2.0026
		Bias	−0.0310	0.0205	0.0068	0.0026
		SD	0.5175	0.3612	0.2276	0.1595
		MSE	0.2682	0.1306	0.0517	0.0254
	μ_3	Mean	2.9520	2.9975	3.0034	2.9930
		Bias	−0.0480	−0.0025	0.0034	−0.0070
		SD	0.6727	0.4368	0.2826	0.2005
		MSE	0.4539	0.1904	0.0797	0.0402
	μ_α	Mean	0.1527	0.1723	0.1872	0.1910
		Bias	−0.0473	−0.0277	−0.0128	−0.0090
		SD	0.1305	0.0937	0.0632	0.0445
		MSE	0.0192	0.0095	0.0042	0.0021
3.0	μ_1	Mean	0.9956	1.0088	0.9977	1.0073
		Bias	−0.0044	0.0088	−0.0023	0.0073
		SD	0.5955	0.4060	0.2652	0.1831
		MSE	0.3539	0.1646	0.0702	0.0335
	μ_2	Mean	1.9941	1.9539	2.0179	2.0106
		Bias	−0.0059	−0.0461	0.0179	0.0106
		SD	0.7438	0.5123	0.3263	0.2232
		MSE	0.5522	0.2641	0.1066	0.0498
	μ_3	Mean	2.9457	3.0010	3.0115	2.9963
		Bias	−0.0543	0.0010	0.0115	−0.0037
		SD	0.9258	0.6740	0.4357	0.2974
		MSE	0.8583	0.4534	0.1896	0.0883
	μ_α	Mean	0.1549	0.1714	0.1891	0.1948
		Bias	−0.0451	−0.0286	−0.0109	−0.0052
		SD	0.1294	0.0932	0.0596	0.0414
		MSE	0.0187	0.0095	0.0037	0.0017

innovation noise do not necessarily degrade estimator performance and may even yield better finite-sample MSE. If, however, the state-specific means are the quantities of main interest, stronger noise substantially reduces their estimation accuracy, and larger sample sizes are required to maintain satisfactory precision.

Overall, the estimators are consistent and perform well in finite samples. The negative μ_α cases, which are particularly relevant to the empirical application, show a small positive bias that diminishes toward zero as $N \to \infty$, and the estimator remains reasonably accurate even under strong mean reversion and regime switching.

To visualize the convergence behaviour, Figure 2 shows the convergence of $\hat{\mu}_\alpha$ across the five cases. All curves approach the true value as N increases, providing empirical support for the consistency established in Theorem 4.2. A small finite-sample bias persists at $N = 1000$, positive for negative μ_α and negative for positive μ_α, which is consistent with the bias reported in Tables 1-5. Larger τ tends to reduce the small-sample bias in the negative μ_α cases, while Case (e) is less sensitive to noise intensity.

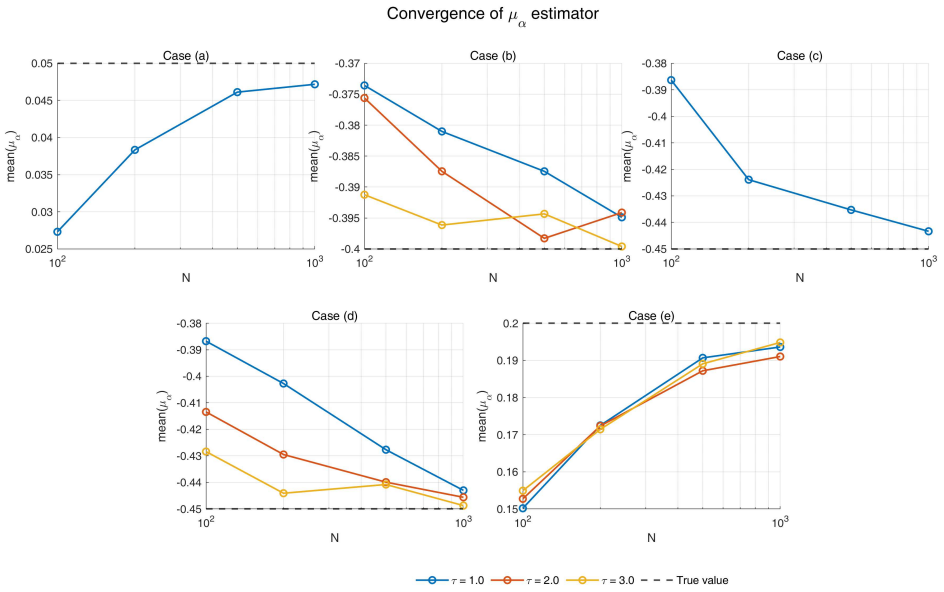


Fig. 2: Convergence of $\hat{\mu}_\alpha$ for Cases (a)-(e). Each panel displays the average estimate over 500 replications versus $\log N$. The dashed line marks the true μ_α. Colours distinguish different values of τ.

To examine the shape of the sampling distribution beyond the summary statistics in Tables 1-5, Figure 3 displays the empirical distributions of $\hat{\mu}_\alpha$ for the five cases. Three features are noteworthy.

First, the spread of the distributions steadily shrinks as N increases. For all cases, the interquartile range at $N = 1000$ is substantially narrower than at $N = 100$, and the number of outliers decreases accordingly. This is consistent with the $O(1/\sqrt{N})$ reduction in standard deviation expected from the asymptotic theory, and complements the convergence patterns shown in Figure 2.

Second, the skewness of the distributions differs systematically between positive and negative μ_α cases. For the negative μ_α cases (b)-(d), the distributions are right-skewed at small N,

with several positive outliers, reflecting the positive finite-sample bias reported in Tables 1-5. For instance, in Case (d) the median at $N = 100$ is around -0.35, whereas at $N = 1000$ it is close to -0.45. For the positive μ_α cases (a) and (e), the distributions are left-skewed at small N, mirroring the negative bias observed in the tables. As N grows, both skewness and outliers diminish, and the distributions become approximately symmetric around the true value.

Third, the presence of outliers at small N highlights the finite-sample instability of the Yule-Walker estimator under strong mean reversion. This is particularly relevant for Case (c), where frequent regime switching reduces the number of consecutive same-state pairs $n_{k,k}$, leading to a wider distribution and more outliers. These findings provide further empirical support for the finite-sample caveats stated in Remark 4.2.

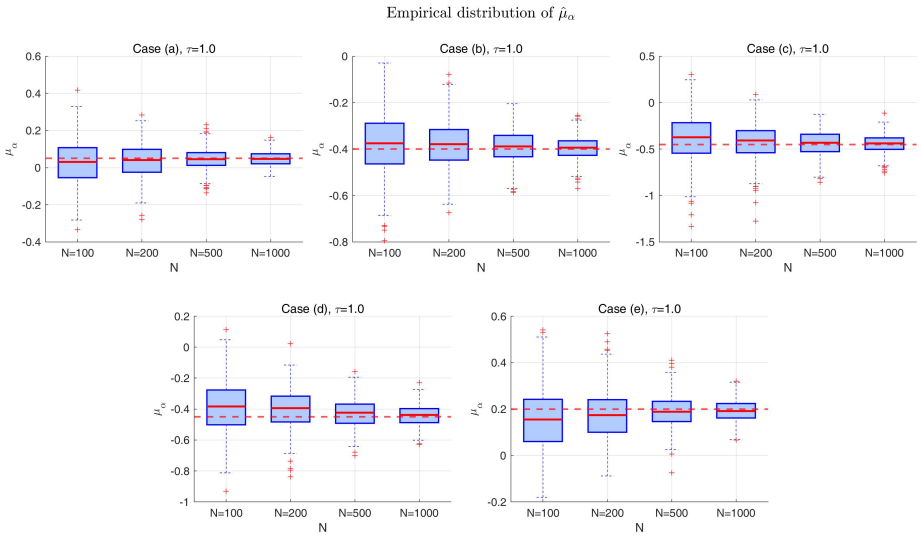


Fig. 3: Box plots of the empirical distributions of $\hat{\mu}_\alpha$ over 500 replications for Cases (a)-(e) at sample sizes $N = 100, 200, 500, 1000$. The red dashed line indicates the true value of μ_α. Boxes show the interquartile range, whiskers extend to $1.5 \times$ IQR, and red crosses denote outliers.

Figure 4 displays the sensitivity of the MSE of $\hat{\mu}_\alpha$ to the noise scaling parameter τ for Cases (b), (d), and (e). For the negative μ_α cases (b) and (d), the MSE decreases as τ increases. For example, in Case (d) at $N = 100$, the MSE drops from 0.0291 at $\tau = 1.0$ to 0.0196 at $\tau = 3.0$; at $N = 200$, it falls from 0.0171 to 0.0101. A similar pattern is observed in Case (b), where the MSE at $N = 100$ decreases from 0.0184 to 0.0140. This pattern is driven by the finite-sample bias: as shown in Tables 1-5, the positive bias of $\hat{\mu}_\alpha$ under negative autocorrelation is substantially reduced when τ is larger. Since MSE is dominated by the squared bias in this setting, the reduction in bias more than offsets the increase in variance, leading to a lower overall MSE.

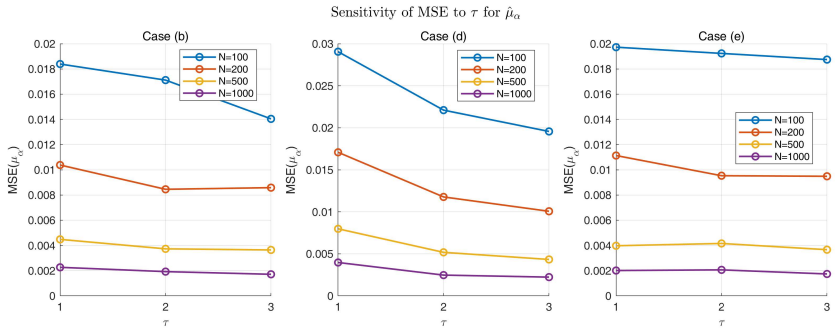


Fig. 4: Sensitivity of the MSE of $\hat{\mu}_\alpha$ to the noise scaling parameter τ for Cases (b), (d), and (e). Each panel corresponds to one case, and the horizontal axis shows $\tau \in \{1.0, 2.0, 3.0\}$. Different colours represent different sample sizes $N = 100, 200, 500, 1000$.

In contrast, for the positive μ_α case (e), the MSE is nearly insensitive to τ. At $N = 100$, it changes only slightly from 0.0197 at $\tau = 1.0$ to 0.0187 at $\tau = 3.0$; at $N = 1000$, it remains around 0.0020 for all three values of τ. This indicates that when the autocorrelation is positive, the estimator is already well-behaved and the noise level has little effect on its finite-sample performance. Across all three panels, the MSE decreases substantially as N increases, which is consistent with the strong consistency and asymptotic normality established in Theorem 4.2 and Theorem 4.3.

To assess the benefit of incorporating the random environment, we compare the proposed $r = 3$ model with a pooled $r = 1$ estimator that ignores the regime structure. Table 6 reports the results for Case (d), which shares the negative mean reversion and heterogeneous state means with the empirical application. The pooled estimator exhibits a substantially larger bias and MSE across all sample sizes. At $N = 100$, the bias of the pooled estimator is approximately 0.14, compared with 0.057 for the $r = 3$ model. At $N = 1000$, the bias of the pooled estimator remains around 0.14, whereas that of the $r = 3$ model drops to 0.025. This contrast is an important finding: ignoring the latent environmental states induces a persistent, non-vanishing bias that cannot be eliminated by increasing the sample size alone.

Table 6: Comparison of the $r = 3$ RERCGINARS(1) model and the $r = 1$ pooled Yule-Walker estimator applied to data generated from Case (d) with $\tau = 1.0$. True value: $\mu_\alpha = -0.45$.

N	Model	Mean	Bias	MSE
100	$r = 3$	-0.3929	0.0571	0.0290
	$r = 1$	-0.3079	0.1421	0.0473
200	$r = 3$	-0.4131	0.0369	0.0151
	$r = 1$	-0.2930	0.1570	0.0416
500	$r = 3$	-0.4306	0.0194	0.0080
	$r = 1$	-0.3106	0.1394	0.0246
1000	$r = 3$	-0.4246	0.0254	0.0058
	$r = 1$	-0.3104	0.1396	0.0225

Table 7 reports the sensitivity of $\hat{\mu}_\alpha$ to the burn-in length and the initial value. The results suggest that $B = 500$ is sufficient and that the estimator is robust to the choice of initial value.

Table 7: Sensitivity of $\hat{\mu}_\alpha$ to burn-in length and initial value for Case (d) with $\tau = 1.0$, $N = 500$, and $R = 100$. True value: $\mu_\alpha = -0.45$.

Setting	Mean	Bias	MSE
Burn-in length ($X_0 = 0$)			
$B = 200$	-0.4223	0.0277	0.0068
$B = 500$	-0.4328	0.0172	0.0078
$B = 1000$	-0.4229	0.0271	0.0085
Initial value ($B = 500$)			
$X_0 = 0$	-0.4252	0.0248	0.0092
$X_0 \sim$ stationary	-0.4327	0.0173	0.0069

6 Real Data Analysis

In this section, we apply the proposed model to a real-world $\mathbb{Z}$-valued time series arising from a complex social system, which displays both positive and negative observations together with regime-switching dynamics. The data consist of the monthly differences in motor vehicle thefts reported to two police beats (numbered 1608 and 2811) in Pittsburgh, Pennsylvania, USA, from January 1990 to December 2001. The dataset was previously analyzed by [21]. We compare the proposed model with the single-regime generalized random coefficient INAR(1) (GRCINARS(1)) benchmark of [18].

6.1 Data Description and Preliminary Analysis

Let $y_n^{(1)}$ and $y_n^{(2)}$ denote the monthly numbers of motor vehicle thefts reported to police beats 1608 and 2811, respectively, covering the period from January 1990 to December 2001. The difference series is defined as

$$X_n = y_n^{(1)} - y_n^{(2)}, \qquad n = 1, 2, \ldots, N,$$

with $N = 144$ observations. The difference series is integer-valued and takes both positive and negative values, which motivates the use of our $\mathbb{Z}$-valued model. Descriptive statistics are reported in Table 8. The difference series has a mean close to zero (0.0417), a standard deviation of 3.1752, and contains 26 zero values (18.06%) and 52 negative values (36.11%). This empirical feature motivates the use of the generalized signed thinning operator, which, unlike standard non-negative INAR models, does not restrict the state space to non-negative integers.

The sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of $\{X_n\}$ are presented in Figure 5 (b) and (c). The ACF exhibits a significant positive spike at lag 1 (approximately 0.51), followed by a gradual decay, and the PACF shows a sharp

Table 8: Descriptive statistics of the motor vehicle theft difference series.

Series	N	Mean	Std. dev.	Median	Min.	Max.	Skewness
Difference $\{X_n\}$	144	0.0417	3.1752	0	-13	12	-0.2473

cutoff after lag 1. This pattern is consistent with a first-order autoregressive structure. The presence of additional autocorrelations at higher lags is consistent with the regime-dependent AR(1) dynamics induced by random environment shifts. We therefore adopt the proposed RERCGINARS(1) model, which captures the regime-switching nature of the series through its state-dependent innovation distribution.

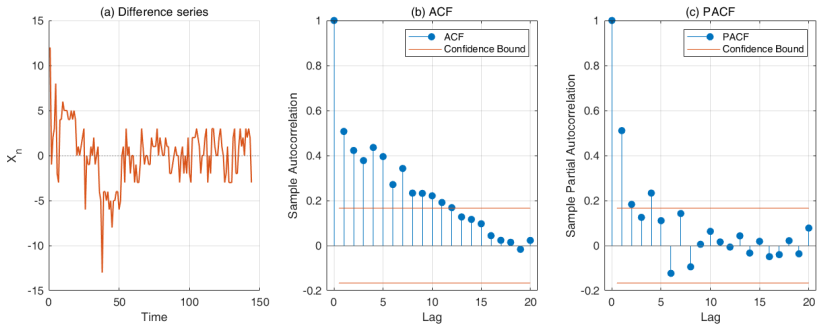


Fig. 5: Motor vehicle theft difference series with ACF and PACF plots.

6.2 Environmental State Identification

Since the environmental states are unobserved and only the difference series is used, we apply K-means clustering directly to X_n to identify homogeneous regimes. The K-means clustering on X_n is an approximation to the latent environmental states $\{Z_n\}$, and the resulting state misclassification bias is quantified in Section 6.4 via the closed-loop Monte Carlo simulation.

We apply K-means with $r = 2$ and $r = 3$. For $r = 2$, the two clusters correspond to periods when beat 1608 has fewer thefts than beat 2811 ($n_1 = 52$, $n_{1,1} = 29$, X_n mean -3.1923) and periods when beat 1608 exceeds beat 2811 ($n_2 = 92$, $n_{2,2} = 69$, X_n mean 1.8696), with a silhouette coefficient of 0.7132. The high value indicates a clear separation between the two states. For $r = 3$, the third cluster contains only 28 observations with $n_{k,k} = 13$, which is insufficient for stable Yule-Walker estimation of the state-specific autocovariance $\gamma_1(3)$. Small $n_{k,k}$ yields highly noisy estimates of $\hat{\gamma}_1^{(k)}$. We therefore adopt $r = 2$ as the primary specification. To provide a more formal basis for this choice, Table 9 compares the predictive performance for $r = 1, 2, 3$. The results show that $r = 2$ provides the best trade-off between model complexity and predictive accuracy: while $r = 1$ (the single-regime GRCINARS(1) benchmark) achieves a slightly lower root mean squared error (RMSE), it fails to capture the regime-switching dynamics, as reflected in the lower Ljung-Box p-value (0.0018 vs. 0.0051);

$r = 3$ yields a higher RMSE and mean absolute error (MAE), and its third state has only $n_{k,k} = 13$ consecutive same-state pairs with a negative state-specific ratio (-0.0267), indicating that $r = 3$ may be over-parameterized for this dataset. The clustering results are displayed in Figure 6.

Table 9: Model selection for the number of environmental states r.

r	RMSE	MAE	LB p-value
1	2.5441	1.9594	0.0018
2	2.5669	1.9809	0.0051
3	2.6471	2.0830	0.0000

Note: For $r = 1$, the model reduces to the single-regime GRCINARS(1) benchmark. For $r = 2$ and $r = 3$, the RERCGINARS(1) model is estimated via Yule-Walker with K-means clustering. Ljung-Box p-values are computed from the standardized residuals.

After clustering, both states have sufficient consecutive same state pairs for reliable estimation. Based on the estimated state sequence $\{\hat{Z}_n\}$, the estimated transition matrix is

$$\hat{P} = \begin{pmatrix} 0.5673 & 0.4327 \\ 0.2527 & 0.7473 \end{pmatrix},$$

which shows that the first state has an expected duration of about $1/(1-0.5673) \approx 2.31$ months, while the second state lasts on average $1/(1-0.7473) \approx 3.96$ months. The transition probabilities are estimated from the empirical frequencies of transitions in the estimated state sequence $\{\hat{Z}_n\}$, with a Laplace smoothing pseudo-count of 0.5 to avoid zero transition probabilities:

$$\hat{p}_{ij}^{\text{smooth}} = \frac{\#\{n : \hat{Z}_n = i,\ \hat{Z}_{n+1} = j\} + 0.5}{\sum_{j'=1}^{r} \#\{n : \hat{Z}_n = i,\ \hat{Z}_{n+1} = j'\} + 0.5r}, \qquad i, j \in \{1, \ldots, r\},$$

where $r = 2$ in our application.

6.3 Parameter Estimation and Model Comparison

Once the environmental states are identified, we estimate the model parameters using the Yule-Walker method developed in Section 4. For each state $k \in \{1,2\}$, we compute the sample mean $\hat{\mu}_k$, the sample variance $\hat{\gamma}_0^{(k)}$ from all observations belonging to state k, and the first-order sample autocovariance $\hat{\gamma}_1^{(k)}$ from consecutive same-state pairs. The estimates are reported in Table 10. We obtain $\hat{\mu}_\alpha = 0.5218$, indicating a positive autocorrelation structure in the difference series. The state-specific ratios are $\hat{\gamma}_1^{(1)}/\hat{\gamma}_0^{(1)} = 0.6588$ and $\hat{\gamma}_1^{(2)}/\hat{\gamma}_0^{(2)} = 0.4444$, both positive.

To quantify the finite-sample uncertainty of $\hat{\mu}_\alpha$, we conducted a parametric bootstrap with $B = 500$ replications. In each replication, we generated a synthetic series of length $N = 144$ from the fitted RERCGINARS(1) model using the estimated parameters $(\hat{\mu}_\alpha, \hat{\mu}_k, \hat{P})$ and the state-specific innovation distributions, and then re-applied the two-step estimation procedure.

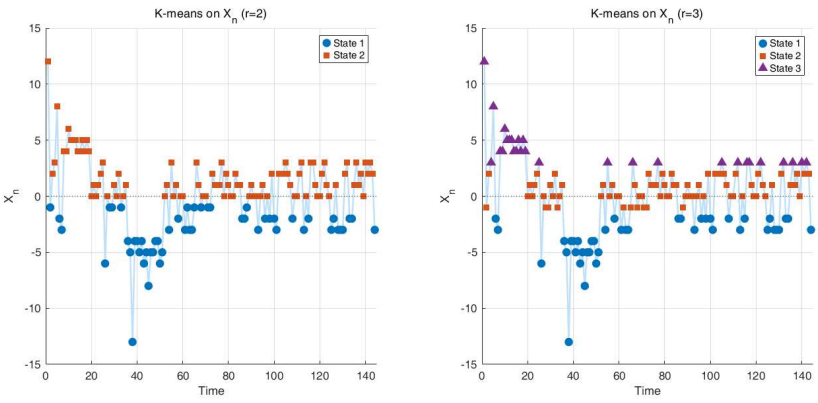


Fig. 6: K-means clustering results on the difference series $\{X_n\}$, for $r = 2$ (left) and $r = 3$ (right).

Table 10: Parameter estimates for the RERCGINARS(1) model applied to the motor vehicle theft differences ($r = 2$).

State	n_k	$n_{k,k}$	$\hat{\mu}_k$	$\hat{\gamma}_0^{(k)}$	$\hat{\gamma}_1^{(k)}$	$\hat{\gamma}_1^{(k)}/\hat{\gamma}_0^{(k)}$
1	52	29	-3.1923	4.4630	2.9402	0.6588
2	92	69	1.8696	3.8960	1.7316	0.4444
$\hat{\mu}_\alpha = \sum_k (n_k/N) \cdot \hat{\gamma}_1^{(k)}/\hat{\gamma}_0^{(k)} = 0.5218$						

Note: The innovation means are $\hat{\mu}_{\varepsilon(1,1)} = -1.5264$, $\hat{\mu}_{\varepsilon(1,2)} = 3.5355$, $\hat{\mu}_{\varepsilon(2,1)} = -4.1679$, and $\hat{\mu}_{\varepsilon(2,2)} = 0.8939$, obtained from $\hat{\mu}_{\varepsilon(i,j)} = \hat{\mu}_j - \hat{\mu}_\alpha \hat{\mu}_i$.

The bootstrap standard error is 0.1228, and the 95% confidence interval is $[0.1037, 0.5756]$. Since this interval excludes zero, the autoregressive coefficient μ_α is significantly positive at the 5% level according to the bootstrap confidence interval. The two-step estimation procedure introduces finite-sample bias due to state misclassification, and the resulting confidence interval should therefore be interpreted with this limitation in mind.

We now consider the predictive performance of the fitted model. We use one-step-ahead forecasts based on the Hamilton filter. For $n = 1, \ldots, N-1$, the predicted value is

$$\hat{X}_{n+1} = \hat{\mu}_\alpha X_n + \sum_{i=1}^{r} \xi_n(i) \sum_{j=1}^{r} \hat{p}_{ij}\, \hat{\mu}_{\varepsilon(i,j)},$$

where $\xi_n(i) = P(Z_n = i \mid \mathcal{F}_n)$ is the filtered state probability obtained from the Hamilton filter with an empirical state-conditional density, and $\hat{P} = (\hat{p}_{ij})$ is the estimated transition matrix. This formulation accounts for the uncertainty in the estimated state labels by replacing the hard cluster assignment with the filtered state probability. The resulting RMSE is 2.5669, and MAE is 1.9809 over the 143 one-step-ahead forecasts.

For comparison, we compute the single-regime GRCINARS(1) benchmark of [18] on the same difference series using the Yule-Walker method. The comparison is summarized in Table 11. The single-regime model yields $\hat{\mu}_\alpha = 0.5107$, whereas the two-regime model yields a slightly larger estimate of 0.5218. The state-specific ratios are both positive, which provides empirical support for the model assumption that a common autoregressive coefficient μ_α governs the first-order dynamics within each regime. In terms of point forecasts, the two-regime model is within approximately 0.9% of the single-regime benchmark, as reported in Table 11. The two-regime specification provides a more interpretable decomposition of the dynamics across regimes.

Table 11: Comparison of the single-regime GRCINARS(1) benchmark and the proposed RERCGINARS(1) model ($r = 2$) applied to the motor vehicle theft difference series.

Model	$\hat{\mu}_\alpha$	RMSE	MAE
GRCINARS(1) [18]	0.5107	2.5441	1.9594
RERCGINARS(1), $r = 2$	0.5218	2.5669	1.9809

Note: Both models are estimated on the same difference series using the Yule-Walker method. For the RERCGINARS(1) model, one-step-ahead predictions are computed using the Hamilton filter with an empirical state-conditional density. The number of valid predictions is $n = 143$ for both models.

6.4 Model Diagnostics

We examine the state-standardized residuals

$$r_n = \frac{X_n - \hat{X}_n}{\hat{s}_{Z_{n-1}}},$$

where $\hat{s}_{Z_{n-1}}$ is the empirical standard deviation of the raw residuals within the estimated state $\hat{Z}_{n-1}$. This state-specific standardization removes the heteroskedasticity induced by regime switching. The empirical standard deviations are $\hat{s}_1 = 2.8828$ and $\hat{s}_2 = 2.3975$. The Ljung-Box test applied to the standardized residuals at lag 10 (with 7 degrees of freedom, accounting for the three estimated parameters) yields a p-value of 0.0051, indicating that some autocorrelation remains. This suggests that some temporal dependence remains in the difference series, although the residual autocorrelation is substantially weaker than in the single-regime benchmark.

The autocorrelation and partial autocorrelation functions of the standardized residuals are displayed in Figure 7. The ACF shows a moderate spike at lag 5 (approximately 0.27), and the PACF exhibits a similar spike at lag 5 (approximately 0.29). This is consistent with the Ljung-Box test result.

Beyond residual autocorrelation, another limitation is the two-step estimation procedure itself. Since the environmental states are estimated via clustering rather than observed, the two-step estimation procedure is not directly covered by the asymptotic consistency results of Theorem 4.2, which assume that the state sequence $\{Z_n\}$ is observed. To quantify the additional bias introduced by the clustering step, we conducted a closed-loop Monte Carlo

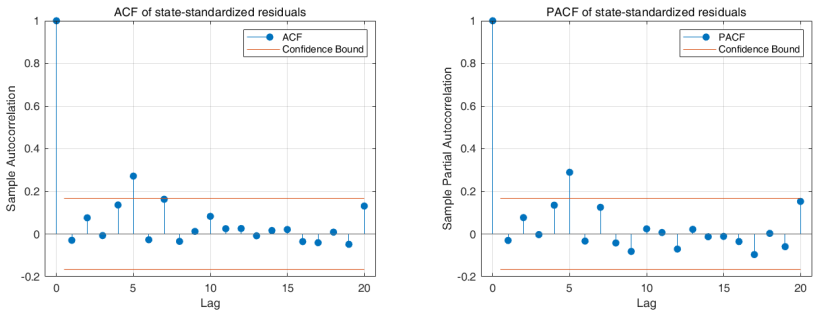


Fig. 7: ACF and PACF of the state-standardized residuals for the RERCGINARS(1) model with $r = 2$ applied to the motor vehicle theft differences.

simulation with $B_{\text{sim}} = 500$ replications. In each replication, we generated $\{X_n, Z_n\}$ from the fitted RERCGINARS(1) model with the same sample length as the observed difference series ($N = 144$), using the estimated parameters as the true values, treating the true state sequence Z_n as unobserved, and applying the same two-step estimation procedure.

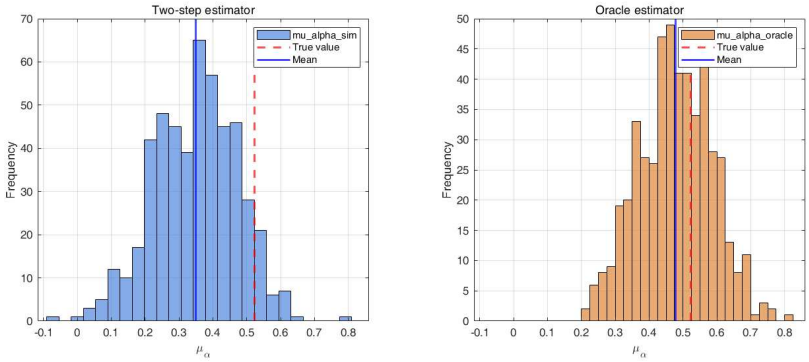


Fig. 8: Empirical distributions of $\hat{\mu}_\alpha$ over 500 closed-loop Monte Carlo replications. Left panel: two-step estimator using K-means clustering on X_n followed by Yule-Walker estimation. Right panel: oracle estimator using the true latent states.

The results are summarized in Figure 8. The two-step estimator exhibits a bias of -0.1733 and an RMSE of 0.2123. In comparison, the oracle estimator using the true latent states has a much smaller bias of -0.0438 and an RMSE of 0.1167. The oracle estimator can be viewed as a theoretical upper bound on the performance achievable when the states are observed. The difference in bias between the two estimators is -0.1295, which accounts for approximately 74.7% of the total finite-sample bias and quantifies the additional error introduced by state misclassification in the clustering step. The 74.7% figure is specific to the present dataset with $N = 144$ and the estimated parameter values; in larger samples or with a higher signal-to-noise ratio, the proportion of bias attributable to state misclassification may be different. These simulation results nevertheless indicate that the two-step procedure is subject to

non-negligible finite-sample bias. Consequently, the point estimate $\hat{\mu}_\alpha = 0.5218$ reported in Section 6.3 should be interpreted with caution due to this bias.

7 Conclusion

In this paper, we proposed a novel random coefficient INAR model based on a generalized signed thinning operator, with the innovation term governed by a finite-state Markov chain. We established the Markovian property of the proposed RERCGINARS(1) process and derived its conditional moments and autocovariance structure. Parameter estimation is carried out via the Yule-Walker method, and the strong consistency and asymptotic normality of the resulting estimators under mild regularity conditions are proved.

The numerical simulation experiments are conducted across two and three distinct environmental states, with the noise scaling parameter $\tau \in \{1.0, 2.0, 3.0\}$ controlling the variance of the Poisson-difference innovations. The results show that the Yule-Walker estimators are consistent, with the finite-sample bias diminishing as the sample size increases. For the state-specific means μ_k, larger τ inflates the sampling variance and MSE, whereas for the autoregressive coefficient μ_α under negative mean reversion, the finite-sample MSE decreases as τ increases due to the reduction of the positive finite-sample bias.

The empirical analysis on the monthly motor vehicle theft differences between two police beats in Pittsburgh shows the applicability of the proposed model. The difference series exhibits a clear regime-switching behavior, with two states identified through a data-driven clustering approach. The Yule-Walker estimates reveal a positive autocorrelation structure ($\hat{\mu}_\alpha = 0.5218$ for $r = 2$), indicating a persistent tendency in the difference series. The bootstrap 95% confidence interval for μ_α is $[0.1037, 0.5756]$, which excludes zero, indicating the significance of the autoregressive coefficient. The two-state model produces positive state-specific autocorrelation ratios, consistent with the model assumption of a common autoregressive coefficient. The point forecast accuracy of the two-regime model is comparable to that of the benchmark GRCINARS(1) model, while the two-regime specification offers a more interpretable decomposition of the dynamics and the ability to capture state-dependent innovation distributions. Nevertheless, the residual diagnostics indicate that some serial dependence remains in the difference series, which motivates higher-order or more flexible extensions.

The proposed RERCGINARS(1) model accommodates negative values and regime-switching dynamics induced by unobserved environmental fluctuations, making it suitable for non-stationary integer-valued time series in criminology and other fields where count data exhibit heterogeneous dynamics driven by latent external factors. Future work may explore higher-order extensions, alternative thinning distributions, or more efficient algorithms for simultaneous state identification and parameter estimation.

Statements and Declarations

- Availability of data and materials
  Data will be made available on request.
- Competing interests
  The authors declare that there are no competing interests.
- Funding

This work was supported by the National Natural Science Foundation of China under Grant numbers 11461032 and 11401267, Humanities and Social Sciences Research Project of Jiangxi Province under Grant number TJ24103, Scientific Research Startup Foundation of Jiangxi University of Science and Technology under Grant number JXUST-102, and Key Laboratory of Low Dimensional Quantum Materials and Sensor Devices of Jiangxi Education Institutes under Grant number GanJiaoKeZi-20241301.

- Author contribution

  Yunyan Wang: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Software; Resources; Validation; Visualization; Writing – original draft; Writing – review & editing.

  Mingtian Tang: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Project administration; Supervision; Validation; Writing – review & editing.

References

[1] Al-Osh, M.A., Alzaid, A.A.: First-order integer-valued autoregressive (INAR(1)) process. J. Time. Ser. Anal. **8**(3), 261–275 (1987) https://doi.org/10.1111/j.1467-9892.1987.tb00438.x

[2] Barreto-Souza, W.: Zero-modified geometric INAR(1) process for modeling count time series with deflation or inflation of zeros. J. Time. Ser. Anal. **36**(6), 839–852 (2015) https://doi.org/10.1111/jtsa.12131

[3] Khoo, W.C., Ong, S.H., Biswas, A.: Modeling time series of counts with a new class of INAR(1) model. Stat. Pap. **58**, 393–416 (2017) https://doi.org/10.1007/s00362-015-0704-0

[4] Yang, K., Wang, D.H., Jia, B.T., Li, H.: An integer-valued threshold autoregressive process based on negative binomial thinning. Stat. Pap. **59**, 1131–1160 (2018) https://doi.org/10.1007/s00362-016-0808-1

[5] Barreto-Souza, W., Ndreca, S., Silva, R.B., Silva, R.W.: Non-linear INAR (1) processes under an alternative geometric thinning operator. Test. **32**(2), 695–725 (2023) https://doi.org/10.1007/s11749-023-00849-y

[6] Scotto, M.G., Weiß, C.H., Gouveia, S.: Thinning-based models in the analysis of integer-valued time series: a review. Stat. Model. **15**(6), 590–618 (2015) https://doi.org/10.1177/1471082X15584701

[7] Serrao, V., Poojari, S., Kamath, A.: Recent advancements in integer-valued autoregressive models for count data time series: A comprehensive review. MethodsX **16**, 103805 (2026) https://doi.org/10.1016/j.mex.2026.103805

[8] Zheng, H.T., Basawa, I.V., Datta, S.: First-order random coefficient integer-valued autoregressive processes. J. Stat. Plan. Infer. **137**(1), 212–229 (2007) https://doi.org/10.

1016/j.jspi.2005.12.003

[9] Lee, S., Jo, M.: Bivariate random coefficient integer-valued autoregressive models: parameter estimation and change point test. J. Time. Ser. Anal. **44**(5-6), 644–666 (2023) https://doi.org/10.1111/jtsa.12662

[10] Liu, X., Jiang, H., Wang, D.: Estimation of parameters in the MDDRCINAR(p) model. J. Stat. Comput. Sim. **93**(6), 983–1010 (2023) https://doi.org/10.1080/00949655.2021.1970163

[11] Tang, M.T., Wang, Y.Y.: Asymptotic behavior of random coefficient INAR model under random environment defined by difference equation. Adv. Differ. Equ. **2014**, 1–9 (2014) https://doi.org/10.1186/1687-1847-2014-99

[12] Nastić, A.S., Laketa, P.N., Ristić, M.M.: Random environment integer-valued autoregressive process. J. Time. Ser. Anal. **37**(2), 267–287 (2016) https://doi.org/10.1111/jtsa.12161

[13] Laketa, P.N., Nastić, A.S., Ristić, M.M.: Generalized random environment INAR models of higher order. Mediterr. J. Math. **15**(9), 1–22 (2018) https://doi.org/10.1007/s00009-017-1054-z

[14] Cui, Y., Wang, Y.Y.: Estimation for random coefficient integer-valued autoregressive model under random environment. Adv. Differ. Equ. **2019**, 1–16 (2019) https://doi.org/10.1186/s13662-019-2436-2

[15] Popović, P.M., Bakouch, H.S., Ristić, M.M.: A non-linear random environment INAR(1) model. J. Comput. Appl. Math. **390**, 113408 (2021) https://doi.org/10.1016/j.cam.2021.113408

[16] Jia, Y.N., Zhu, F.K., Wang, T., Li, Q.: A new time-varying-threshold integer-valued autoregressive model. Commun. Nonlinear Sci. Numer. Simulat. **163**(1), 110445 (2026) https://doi.org/10.1016/j.cnsns.2026.110445

[17] Kim, H.Y., Park, Y.: A non-stationary integer-valued autoregressive model. Stat. Pap. **49**, 485–502 (2008) https://doi.org/10.1007/s00362-006-0028-1

[18] Zhang, H.X., Wang, D.H., Zhu, F.K.: Generalized RCINAR(1) process with signed thinning operator. Commun. Stat-Theor M. **41**(10), 1750–1770 (2012) https://doi.org/10.1080/03610926.2010.551452

[19] Kachour, M., Bakouch, H.S., Mohammadi, Z.: A new INAR(1) model for $\mathbb{Z}$-valued time series using the relative binomial thinning operator. J. Econ. Stat. **243**(2), 125–152 (2023) https://doi.org/10.1515/jbnst-2022-0059

[20] Meyn, S.P., Tweedie, R.L.: Markov Chains and Stochastic Stability. Communications and Control Engineering Series. Springer, London (1993). https://doi.org/10.1017/CBO9780511626630

[21] Pirković, B.A., Ristić, M.M., Nastić, A.S.: Random environment integer-valued autoregressive process with discrete Laplace marginal distributions. Revstat Stat. J. **21**(4), 469–490 (2023) https://doi.org/10.57805/revstat.v21i4.430