

Inference for multiple change-points in generalized integer-valued autoregressive model

Danshu Sheng¹, Dehui Wang^{1*}

Abstract In this paper, we propose a computationally valid and theoretically justified methods, the likelihood ratio scan method (LRSM), for estimating multiple change-points in a piecewise stationary generalized conditional integer-valued autoregressive process. LRSM with the usual window parameter h is more satisfied to be used in long-time series with few and even change-points vs. LRSM with the multiple window parameter h_{mix} performs well in short-time series with large and dense change-points. The computational complexity of LRSM can be efficiently performed with order $O((\log n)^3 n)$. Moreover, two bootstrap procedures, namely parametric and block bootstrap, are developed for constructing confidence intervals (CIs) for each of the change-points. Simulation experiments and real data analysis show that the LRSM and bootstrap procedures have excellent performance and are consistent with the theoretical analysis.

Keywords : Piecewise stationary GCINAR process · Multiple change-points estimation · Likelihood ratio · Confidence interval · Count time series

1 Introduction

Modeling and analysis of non-stationary count time series have attracted a lot of attention over the past years. Although complex non-stationary models have been developed in different fields, they are often difficult to explain. The concept of piecewise stationary models has become a popular method by dividing nonstationary data into several stationary parts. Among the different types of piecewise stationary models, the so-called multiple change-points (MCP) models have received special attention. Studies of change-points models date back to [Page \(1954, 1955\)](#). Since then, this topic has been of interest to statisticians and researchers in many other fields. So far, many excellent articles, such as [Lee et al. \(2003\)](#), [Davis et al. \(2006\)](#), [Chan et al. \(2014\)](#), [Chen et al. \(2021\)](#), [Aue and Horváth \(2013\)](#), [Niu et al. \(2016\)](#), [Casini and Perron \(2018\)](#), [Truong et al. \(2020\)](#), just to name a few, have studied and reviewed methodological issues related to estimation, detection and computation for continuous time series models involving structural

¹School of Mathematics and Statistics, Liaoning University, Shenyang, China

*Corresponding author, E-mail: wangdehui@lnu.edu.cn

changes. In contrast, the research of count time series models involving change-points is still mainly focused on the change-points detection, and there is little research on change-points estimation and computation.

One common and useful model to describe simple stationary count time series data is integer-valued autoregressive (INAR) time series model. Since [Al-Osh and Alzaid \(1987\)](#) proposed the INAR(1) model based on binomial thinning operator ([Steutel and Van Harn, 1979](#)), modeling INAR-type models based on different thinning operators have become a common approach and have been widely used in many fields like epidemiology, social sciences, economics, life sciences and others. To make the integer-valued models more flexible for practical purposes, many scholars have extended the INAR (1) model by changing thinning operators, innovation. For example, just to name a few, the INAR(p) model based on generalized thinning operator proposed by [Latour \(1998\)](#); the mixture INAR(1) model based on the mixture of Pegram and thinning operators studied by [Khoo et al. \(2017\)](#); the bounded binomial autoregressive (BAR) model proposed by [McKenzie \(1985\)](#); the nonlinear INAR(1) model, self-exciting threshold autoregressive process introduced by [Monteiro et al. \(2006\)](#). Actually, most of these INAR-type models can be written as the following p -order causal and stationary generalized conditional integer-valued autoregressive (GCINAR(p)) process.

Definition 1 *Considering a \mathbb{N}_0 -valued ($\mathbb{N}_0 = \mathbb{N} \cup \{0\}$) GCINAR(p) process $\{X_t\}_{t \in \mathbb{Z}}$, where the conditional mean is defined by the following recursion*

$$\mathbb{E}(X_t | \mathcal{F}_{t-1}) = \sum_{k=1}^p \beta_k X_{t-k} + \beta_0. \quad (1.1)$$

$\mathcal{F}_t = \sigma(X_s, s \leq t)$ is the σ -field generated by the whole information up to time t . The parameter vector $\boldsymbol{\theta} = (\beta_0, \dots, \beta_p)^\top$ satisfied $\sum_{i=1}^p \beta_i < 1$, $\beta_i \geq 0$ for $i = 1, \dots, p$ and $\beta_0 > 0$.

Remark 1 *Clearly, GCINAR can be called a conditional linear AR (CLAR, [Grunwald et al. \(2000\)](#)) model in term of the form of the model. However, the change-point of the continuous-type CLAR models has been studied by some scholars, see [Yau and Zhao \(2016\)](#), [Ng et al. \(2022\)](#). Therefore, the focus of this paper is on the model defined on integer values and the conditional expectation is linear. For example, the INAR-type model based on thinning operators, the INARCH models ([Wei, 2010](#)), the BAR models ([McKenzie, 1985](#)) and among others.*

Then, a simple and useful idea to model piecewise stationary count time series is to construct the following GCINAR model with multiple change-points, that is, the so-called multiple change-points generalized conditional integer-valued autoregressive model (MCP-GCINAR).

Definition 2 *The MCP-GCINAR process with m change-points $\{X_t\}_{t=1}^n$ is defined by the recursion:*

$$\mathbb{E}(X_t | \mathcal{F}_{t-1}) = \begin{cases} \beta_{1,1} X_{t-1,1} + \dots + \beta_{p_1,1} X_{t-p_1,1} + \beta_{0,1}, & 0 < t \leq \tau_1, \\ \vdots \\ \beta_{1,j} X_{t-\tau_{j-1}-1,j} + \dots + \beta_{p_j,j} X_{t-\tau_{j-1}-p_j,j} + \beta_{0,j}, & \tau_{j-1} < t \leq \tau_j, \\ \vdots \\ \beta_{1,m} X_{t-\tau_m-1,m} + \dots + \beta_{p_m,m} X_{t-\tau_m-p_m,m} + \beta_{0,m}, & \tau_m < t \leq n, \end{cases} \quad (1.2)$$

where $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_m)$ denotes the vector of unknown locations of change-points, $\tau_0 = 0$ and $\tau_{m+1} = n$. Each change-point location τ_j is an integer between 1 and $n - 1$ inclusive, and the change-points are ordered such that $\tau_{j_1} < \tau_{j_2}$ if, and only if, $j_1 < j_2$. The time series vector (X_1, X_2, \dots, X_n) can be written as

$$(X_{1,1}, \dots, X_{n_1,1}, \dots, X_{\tau_{j-1}+1,j}, \dots, X_{\tau_j+n_j,j}, \dots, X_{\tau_{m+1},m+1}, \dots, X_{\tau_m+n_{m+1},m+1}), \quad (1.3)$$

where $n_j = \tau_j - \tau_{j-1}$ for $j = 1, \dots, m + 1$ and $n = n_1 + n_2 + \dots + n_{m+1}$. In particular, the j th piece $\{X_{t,j}\}_{t=1}^{n_j}$ of the series is modeled as a stationary GCINAR process (1.1) with order p_j ,

$$E(X_{t,j}|\mathcal{F}_{t-1}) = \beta_{1,j}X_{t-1,j} + \dots + \beta_{p_j,j}X_{t-p_j,j} + \beta_{0,j},$$

$\boldsymbol{\theta}_j = (\beta_{0,j}, \beta_{1,j}, \dots, \beta_{p_j,j})^\top$ is the parameter vector corresponding to the j th segment GCINAR(p_j) process, which is assumed to be an interior point of the compact space $\Theta_j(p_j) = [\delta, 1/\delta] \times [0, 1 - \delta]^{p_j} \cap \mathbb{M}_j$, where $\mathbb{M}_j = \{0 \leq \sum_{k=1}^{p_j} \beta_{k,j} \leq 1 - \delta < 1\}$, $\delta \in (0, 1)$ is a constant.

So far, most research has focused on the change-points detection of the INAR-type models. For example, among others, Kang and Lee (2009), Yu and Kim (2020) and Lee and Jo (2022) considered the problem of testing for a parameter change in different types of INAR(1) models by taking advantage of the cumulative sum (CUSUM) test. Chattopadhyay et al. (2021) considered the problem of change-point analysis for the INAR(1) model with time-varying covariates. Yu et al. (2022) applied the empirical likelihood ratio (ELR) test to uncover a structural change in INAR processes. Weiß (2007, 2009a,b), Weiß and Testik (2009) studied the Shewhart, combined jumps, exponentially weighted moving average and cumulative sum charts for controlling the Poisson counts process.

However, few research has studied the estimation of change-points, especially optimization. Diop and Kengne (2021b) derive a data-driven procedure based on the slope heuristic to calibrate the penalty term of the contrast to achieve general integer-valued time series change-points estimation, and optimized by dynamic programming (DP) algorithm; Sheng and Wang (2023) studied the change-points analysis of the MCP-GCINAR model based on minimum description length (MDL) principle, and optimized by genetic algorithm (GA). Although these two algorithms have been suggested for implementing the inference of change-points in the MCP-GCINAR model, optimization can be computationally very expensive because the number of possible change-points combinations grows exponentially as the sample size grows. Specifically, the GA involves various tuning parameters, and the DP algorithm exhibits a computational order $O(n^2)$. In fact, if the computational complexity of optimal methods is too great for the application at hand, we can resort to an ‘‘approximate methods’’. That is, we can roughly judge a potential change-points set which is far smaller than the sample size by some methods, and then the optimization problem based on information criteria is realized by selecting the best subset of this potential change-points set. Of course, it is necessary to ensure that the number of elements in the potential change-points is higher than the true number of change-points, and there are some subsets in a neighborhood of all true change-points. This ‘‘approximate methods’’ approach is also often considered in some continuous-type models, such as Yau and Zhao (2016), Ng et al. (2022), Chan et al. (2014), Safikhani and Shojaie (2022). In addition, it should be noted that the generalized likelihood ratio

scan method (GLRSM) proposed by [Ng et al. \(2022\)](#) requires the likelihood function to be a measurable and continuous function with respect to $\{X_t\}$. And the proof for the GLRSM relies on the application of the continuous mapping theorem. Consequently, it can not be simply applied to count time series models.

Inspired by the above review/discussion, we propose two computationally valid and theoretically justified methods for off-line change-points inference in the MCP-GCINAR processes. The main contributions in this paper are as follows:

- We propose a computationally efficient and theoretically sound procedure LRSM for the inference of the MCP-GCINAR model in the case of long-time series with few and even change-points. Performing LRSM involves three steps: first, a likelihood ratio scan statistic is used to obtain a potential change-points set; second, a model selection procedure based on the MDL principle is employed to give a set of consistent change-points estimates; and finally, an exhaustive method is used to determine the final change-points in an extended local window and its convergence and asymptotic distribution are given. Also, we demonstrated that the computational order of the LRSM can be as low as $O((\log n)^3 n)$. After simulation results and real data analysis, LRSM performs well in samples of long-time series with few and even change-points.
- We construct the CIs based on two bootstrap procedures, parametric and block bootstrap, for each estimated change-point. In addition, the validity of the two bootstrap procedures are discussed.

Although these methods are similar to the problem of change-points estimation in continuous-type models, there are many differences in the derivation of asymptotic properties, model assumptions, and algorithm implementation. For example, the techniques and assumptions used to develop the asymptotic properties in LRSM, the techniques used to develop the validity of the two bootstrap procedures.

The rest contents of this article are organized as follows. In Section 2, we state the details of the three-step LRSM and construct the confidence intervals by two bootstrap procedures. Section 3 discusses the tuning parameters, computational complexity and some implementation issues. Extensive simulation studies and real data applications are given in Sections 4 and 5 respectively. The article ends with a conclusion section. All proofs, some implementation algorithms and additional simulation results are given in the Appendix.

2 MCP-GCINAR model change-points inference based on the three-step LRSM

In this section, we state the three-step LRSM to implement the inference of the MCP-GCINAR model. Following Section 2.1 in [Sheng and Wang \(2023\)](#), considering the complexity of proof and computation, Poisson quasi-maximum likelihood (PQML) estimation as the cost function for the MDL criterion is reasonable and convenient. Thus, we first review the Poisson quasi-maximum

log-likelihood for the GCINAR process and provide some notations and assumptions used in the LRSM.

The conditional Poisson quasi-maximum log-likelihood for the data set is obtained by

$$\begin{aligned} L_n(k, \boldsymbol{\theta}, p) &= \sum_{t=k+1}^{n+k} [X_t \log \xi_t(\boldsymbol{\theta}, p | \mathbf{X}_{t-1}) - \xi_t(\boldsymbol{\theta}, p | \mathbf{X}_{t-1})] \\ &= \sum_{t=k+1}^{n+k} \ell_t(\boldsymbol{\theta}, p), \end{aligned} \quad (2.1)$$

where $\{X_t\}_{t=k+1-p}^{n+k}$ is generated from the GCINAR process (1.1), $k \in \mathbb{Z}$ and $k \geq p$, $\xi_t(\boldsymbol{\theta}, p | \mathbf{X}_{t-1}) = \sum_{i=1}^p \beta_i X_{t-i} + \beta_0 \triangleq \mathbf{X}_{t-1}^\top \boldsymbol{\theta}$ with $\mathbf{X}_{t-1} = (1, X_{t-1}, X_{t-2}, \dots, X_{t-p})^\top$ and $\boldsymbol{\theta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$. Then the PQML estimate of $\boldsymbol{\theta}$ is defined by

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta(p)} L_n(k, \boldsymbol{\theta}, p).$$

Assuming that $0 < \sum_{i=1}^p \beta_i < 1$ and $E(X_t) < \infty$, then $\hat{\boldsymbol{\theta}}$ satisfy the following asymptotic normality

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{d} N(0, \boldsymbol{\Sigma}) \quad \text{as } n \rightarrow \infty,$$

where $\boldsymbol{\theta}^0$ denote the true parameter value of $\boldsymbol{\theta}$. The asymptotic variance matrix of the PQML estimate can be consistently estimated by $\hat{\boldsymbol{\Sigma}}_n = \hat{\mathbf{J}}_n^{-1}(k, \hat{\boldsymbol{\theta}}) \hat{\mathbf{I}}_n(k, \hat{\boldsymbol{\theta}}) \hat{\mathbf{J}}_n^{-1}(k, \hat{\boldsymbol{\theta}})$ with

$$\hat{\mathbf{J}}_n(k, \hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{t=k+1}^{n+k} \frac{1}{\xi_t(\boldsymbol{\theta}, p | \mathbf{X}_{t-1})} \frac{\partial \xi_t(\boldsymbol{\theta}, p | \mathbf{X}_{t-1})}{\partial \boldsymbol{\theta}} \frac{\partial \xi_t(\boldsymbol{\theta}, p | \mathbf{X}_{t-1})}{\partial \boldsymbol{\theta}^\top} \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (2.2)$$

$$\hat{\mathbf{I}}_n(k, \hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{t=k+1}^{n+k} \left(\frac{X_t}{\xi_t(\boldsymbol{\theta}, p | \mathbf{X}_{t-1})} - 1 \right)^2 \frac{\partial \xi_t(\boldsymbol{\theta}, p | \mathbf{X}_{t-1})}{\partial \boldsymbol{\theta}} \frac{\partial \xi_t(\boldsymbol{\theta}, p | \mathbf{X}_{t-1})}{\partial \boldsymbol{\theta}^\top} \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \quad (2.3)$$

Notations:

- Denote this whole class of the MCP-GCINAR model by \mathcal{M} and any model from this class by $\mathcal{F} \in \mathcal{M}$.
- Upper bounds for the GCINAR orders p are represented by p_{\max} . Setting $\mathbf{p} = (p_1, \dots, p_{m+1})$ belongs to the parameter domain $\mathcal{P} = (0, p_{\max}]^{m+1} \cap \mathbb{Z}^{m+1}$, $\boldsymbol{\theta}(m) = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{m+1})$ belongs to the parameter domain $\Theta = \prod_{j=1}^{m+1} \Theta_j(p_j)$.
- Denote the true number of change-points by m_0 , the set of true change-points by $\mathcal{J}_0 = \{\tau_1^0, \tau_2^0, \dots, \tau_{m_0}^0\}$, the set of true order by $\mathbf{p}_0 = \{p_1^0, \dots, p_{m_0+1}^0\}$.
- Let $|\mathcal{J}|$ be the cardinality of the set \mathcal{J} .

Assumptions:

H. 1 For each segment, assume that $\boldsymbol{\theta}_j$ is an interior point of the compact space $\Theta_j(p_j)$ and satisfies $\sum_{i=1}^{p_j} \beta_{i,j} < 1$, $\beta_i \geq 0$ for $i = 1, \dots, p_j$.

H. 2 Assume that there exists a $\epsilon_\theta > 0$ such that $\min_{1 \leq j \leq m_0} \|\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_j\| > \epsilon_\theta$.

H. 3 Assume that there exists a $\epsilon_\tau > 0$ such that $\min_{0 \leq j \leq m_0} |\tau_{j+1} - \tau_j| > n\epsilon_\tau$.

H. 4 For each segment, assume that there exist a constant $\epsilon_K > 0$ such that $\mathbb{E}[e^{l_t(\boldsymbol{\theta}_j) - \mathbb{E}[l_t(\boldsymbol{\theta}_j)]}] < \epsilon_K$ for all $\boldsymbol{\theta}_j \in \Theta_j(p_j)$.

H. 5 For each segment, assume that $\mathbb{E}|X_t|^{4+\epsilon_X} < \infty$ for some $\epsilon_X > 0$.

Under the assumption H.1, the j th segment process $\{X_{t,j}\}$ is ergodic and has a strictly stationary solution, additionally, $\{X_{t,j}\}$ is strong mixing with geometric rate. Assumption H.2 imposes restrictions on the parametric differences between each segment, which is essential for the existence of change-points. Assumption H.3 imposes restrictions on the distances between change-points. To accurately estimate the specified GCINAR parameter values, the segments must have a sufficient number of observations. If not, the estimation is over-determined and the likelihood has an infinite value. This assumption is common in likelihood-based model selection, such as Davis et al. (2006), Davis and Yau (2013), Yau and Zhao (2016). Assumption H.5 is proposed to guarantee the consistency of the change-points estimate in Theorem 2. Assumption H.4 provides the condition for the establishment of the large deviation conclusion in Theorem 1. It's worth mentioning that, if the arbitrary moments of subsegment model are bounded, assumption H.5 and H.4 are unnecessary. For example, if the MCP-GCINAR model represent the BAR model (McKenzie, 1985) with multiple change-points, it is no longer necessary to assume H.5 and H.4.

2.1 First Step: obtain potential change-points based on LR scan statistics

For $t = h, \dots, n - h$, define the scanning window at t and the corresponding observations as

$$W_t(h) = \{t - h + 1, \dots, t + h\} \quad \text{and} \quad X_{W_t(h)} = (X_{t-h+1}, \dots, X_{t+h}),$$

respectively, where h is called the window radius. To establish asymptotic theory, we assume that $h = h(n)$ depends on the sample size n . Then the likelihood ratio scan statistic for the scanning window $W_t(h)$ by

$$S_h(t) = \frac{1}{h} L_h(t - h, \hat{\boldsymbol{\theta}}_1, p_1) + \frac{1}{h} L_h(t, \hat{\boldsymbol{\theta}}_2, p_2) - \frac{1}{h} L_{2h}(t - h, \hat{\boldsymbol{\theta}}, p)$$

where $L_h(t - h, \hat{\boldsymbol{\theta}}_1, p_1)$, $L_h(t, \hat{\boldsymbol{\theta}}_2, p_2)$, $L_{2h}(t - h, \hat{\boldsymbol{\theta}}, p)$, defined by (2.1), are the Poisson quasi-likelihoods formed by the observations $\{X_s\}_{s=t-h+1}^t$, $\{X_s\}_{s=t+1}^{t+h}$, $\{X_s\}_{s=t-h+1}^{t+h}$, evaluated at the PQML estimates $\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \hat{\boldsymbol{\theta}}$, respectively. Similar to AR model, each GCINAR(p_j) model can be

regarded as an GCINAR(p_{\max}) model (with the last few coefficients equal to 0), we can regard each GCINAR model as p_{\max} order when deducing the theoretical results in this step.

Next, $S_h(t)$ scans the observed time series to obtain a sequence of likelihood ratio scan statistics $(S_h(h), S_h(h+1), \dots, S_h(n-h))$. As discussed by [Yau and Zhao \(2016\)](#), after this construction, $S_h(t)$ tends to be larger if t is change-point. In particular, if h is selected as $2h < n\epsilon_\tau$ and $h > p_{\max}$, there is at most one change-point in each scanning window. Thus, we can derive a set of potential change points from local change point estimates, given by

$$\hat{\mathcal{J}}^{(1)}(h) = \left\{ \tau \in \{h, h+1, \dots, n-h\} : S_h(\tau) = \max_{t \in (\tau-h, \tau+h]} S_h(t) \right\},$$

where $S_h(t) \triangleq 0$, for $t < h$ and $t > n-h$. Clearly, if $S_h(\tau)$ is the maximum on the window $[\tau-h+1, \tau+h]$ centered on τ , then τ is a local point of change estimate.

Denote the number of elements in $\hat{\mathcal{J}}^{(1)}(h)$ by $\hat{m}^{(1)}$, the local change-points estimates set by $\hat{\mathcal{J}}^{(1)}(h) = \{\tau_1^{(1)}, \tau_2^{(1)}, \dots, \tau_{\hat{m}^{(1)}}^{(1)}\}$. [Theorem 1](#) states that all change-points can be identified in an h -neighborhood of $\hat{\mathcal{J}}^{(1)}(h)$ obtained in this step.

Theorem 1 *Suppose Assumptions H.1-H.4 hold, $2h < n\epsilon_\tau$ and $\epsilon_\tau > c$ for some $c > 0$, then there exists some $d > 0$ such that, for $h \geq d(\log n)^3$,*

$$\mathbb{P}\left(\max_{\tau \in \mathcal{J}_0} \min_{s=1, \dots, \hat{m}^{(1)}} |\tau - \hat{\tau}_s^{(1)}| < h\right) \rightarrow 1.$$

For the first step, make the following supplementaries through [Remark 2](#).

Remark 2

- *To enhance practical performance, when we evaluate $S_h(t)$, the order p for the three estimates $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}$ can be individually selected by using the information criteria: Akaike information criterion (AIC) or Bayesian information criterion (BIC).*
- *Usually, $\hat{m}^{(1)}$ is much larger than the true number of change-points, especially if h is small. In order to enhance computational efficiency, we can select the first m_{\max} elements in $\hat{\mathcal{J}}^{(1)}$ that have the largest $S_h(t)$ s as the final $\hat{\mathcal{J}}^{(1)}$, where m_{\max} is a reasonably large number upper bound of the number of change points, such as 20 or 50.*

Note that these two operations are only intended to enhance practical performance and computational efficiency, and are not necessary, especially the value of m_{\max} has little effect on the estimation results.

2.2 Second Step: consistency estimation based on MDL principle

The set of potential change-points $\hat{\mathcal{J}}^{(1)}$ can be viewed as a rough estimate of the change-points, which usually overestimates the true set of change-points. To detect the true change-points, the best subset of $\hat{\mathcal{J}}^{(1)}$ can be selected according to some prescribed information criterion (IC), such as BIC ([Yao, 1988](#)), MDL ([Davis et al., 2006](#)). Here, MDL is selected as IC to select the most concise model as the optimal model.

Given a set of change-points $\mathcal{J} = \{\tau_1, \dots, \tau_m\}$, the MDL criterion is defined as

$$\begin{aligned} \text{MDL}(m, \mathcal{J}, \mathbf{p}) = & \log(m) + (m+1) \log n + \sum_{j=1}^{m+1} \log(p_j) + \sum_{j=1}^{m+1} \frac{p_j + 1}{2} \log(n_j) \\ & - \sum_{j=1}^{m+1} L_{n_j}(\tau_{j-1}, \boldsymbol{\theta}_j, p_j; \mathbf{X}_j). \end{aligned} \quad (2.4)$$

where $\mathbf{X}_j = \{X_{\tau_{j-1}+1}, \dots, X_{\tau_j}\}$. Given the potential change-points estimates $\hat{\mathcal{J}}^{(1)}$, the change-points and corresponding GCINAR orders can be estimated by

$$(\hat{m}^{(2)}, \hat{\mathcal{J}}^{(2)}, \hat{\mathbf{p}}^{(2)}) = \arg \min_{m=|\mathcal{J}|, |\mathcal{J}| \subseteq \hat{\mathcal{J}}^{(1)}, \mathbf{p} \in \mathcal{P}} \text{MDL}(m, \mathcal{J}, \mathbf{p}), \quad (2.5)$$

where $\hat{m}^{(2)} = |\hat{\mathcal{J}}^{(2)}|$, $\hat{\mathcal{J}}^{(2)} = \{\tau_1^{(2)}, \tau_2^{(2)}, \dots, \tau_{\hat{m}^{(2)}}^{(2)}\}$, $\hat{\mathbf{p}}^{(2)} = (\hat{p}_1^{(2)}, \dots, \hat{p}_{\hat{m}^{(2)}}^{(2)})$. Minimization MDL equation (2.5) can be achieved by optimal partitioning (OP) algorithm (Jackson et al., 2005). The implementation details of OP algorithm are given in Appendix. Following Sheng and Wang (2023), which proved the consistency of estimates based on the MDL principle, we have the following Theorem 2.

Theorem 2 *Under the setting in Theorem 1 and suppose Assumption 5 hold, there is $\hat{m}^{(2)} \xrightarrow{p} m_0$. In addition, given that $\hat{m}^{(2)} = m_0$, we have*

$$\mathbb{P}\left(\max_{j=1, \dots, m_0} |\hat{\tau}_j^{(2)} - \tau_j^0| < h\right) \rightarrow 1, \quad \text{and} \quad \max_{j=1, \dots, m_0} |\hat{p}_j^{(2)} - p_j^0| \xrightarrow{p} 0.$$

2.3 Third Step: final change-points estimates

Although the consistent estimate of m can be achieved based on the MDL criterion, it should be noted that $\hat{\mathcal{J}}^{(2)}$ is a subset of $\hat{\mathcal{J}}^{(1)}$, Theorem 2 only implies that $\max_{j=1, \dots, m_0} |\hat{\tau}_j^{(2)} - \tau_j^0| = O_p(h)$, which is not optimal compared with the typical rate of $O_p(1)$. Nevertheless, the Theorem 2 also guarantees that the true change-point τ_j^0 is within $(\hat{\tau}_j^{(2)} - h, \hat{\tau}_j^{(2)} + h]$ with probability approaching 1 for all $j = 1, \dots, \hat{m}^{(2)}$. A simple idea is to use exhaustive methods to determine the final change-points in an extended local window $(\hat{\tau}_j^{(2)} - 2h, \hat{\tau}_j^{(2)} + 2h]$ and corresponding observations. Yau and Zhao (2016) discusses in detail that this is possible, and if $3h < n\epsilon_\tau$, there is only one change point inside the extended local window.

Define the extended local window and corresponding observations for the j th estimated change-point $\hat{\tau}_j^{(2)}$ by

$$EW_j(h) = \{\hat{\tau}_j^{(2)} - 2h + 1, \dots, \hat{\tau}_j^{(2)} + 2h\} \quad \text{and} \quad X_{EW_j(h)} = \{X_{\hat{\tau}_j^{(2)} - 2h + 1}, \dots, X_{\hat{\tau}_j^{(2)} + 2h}\}.$$

For $j = 1, \dots, \hat{m}^{(2)}$, Let

$$L_j(\tau, \boldsymbol{\theta}_j, \boldsymbol{\theta}_{j+1}) = L_{\tau - \hat{\tau}_j^{(2)} + 2h}(\hat{\tau}_j^{(2)} - 2h, \boldsymbol{\theta}_j, \hat{p}_j^{(2)}) + L_{\hat{\tau}_j^{(2)} + 2h - \tau}(\tau, \boldsymbol{\theta}_{j+1}, \hat{p}_{j+1}^{(2)}),$$

where $L_{\tau-\hat{\tau}_j^{(2)}+2h}(\hat{\tau}_j^{(2)}-2h, \boldsymbol{\theta}_j, \hat{p}_j^{(2)})$ and $L_{\hat{\tau}_j^{(2)}+2h-\tau}(\tau, \boldsymbol{\theta}_{j+1}, \hat{p}_{j+1}^{(2)})$ are defined in (2.1), i.e.,

$$L_{\tau-\hat{\tau}_j^{(2)}+2h}(\hat{\tau}_j^{(2)}-2h, \boldsymbol{\theta}_j, \hat{p}_j^{(2)}) = \sum_{t=\hat{\tau}_j^{(2)}-2h+1}^{\tau} \ell_t(\boldsymbol{\theta}_j, \hat{p}_j^{(2)}),$$

$$L_{\hat{\tau}_j^{(2)}+2h-\tau}(\tau, \boldsymbol{\theta}_{j+1}, \hat{p}_{j+1}^{(2)}) = \sum_{t=\tau+1}^{\hat{\tau}_j^{(2)}+2h} \ell_t(\boldsymbol{\theta}_{j+1}, \hat{p}_{j+1}^{(2)}).$$

Define the final estimate as

$$\hat{\tau}_j^{(3)} = \arg \max_{\tau \in (\hat{\tau}_j^{(2)}-h, \hat{\tau}_j^{(2)}+h]} L_j(\tau, \hat{\boldsymbol{\theta}}_j, \hat{\boldsymbol{\theta}}_{j+1}),$$

where $\hat{\boldsymbol{\theta}}_j$ and $\hat{\boldsymbol{\theta}}_{j+1}$ are the maximizers of $L_{\tau-\hat{\tau}_j^{(2)}+2h}(\hat{\tau}_j^{(2)}-2h, \boldsymbol{\theta}_j, \hat{p}_j^{(2)})$ and $L_{\hat{\tau}_j^{(2)}+2h-\tau}(\tau, \boldsymbol{\theta}_{j+1}, \hat{p}_{j+1}^{(2)})$ on $\boldsymbol{\Theta}_j(\hat{p}_j)$ and $\boldsymbol{\Theta}_j(\hat{p}_{j+1})$, respectively. Following Theorem 1 and 2 in Cui et al. (2021), we have the convergence and asymptotic distribution of the final estimates.

Theorem 3 *Under the setting in Theorem 2 and $3h < n\epsilon_\tau$, Then, we have*

$$\hat{\tau}_j^{(3)} - \tau_j^0 \xrightarrow{d} \arg \max_{\tau} W_{j,\tau},$$

where

$$W_{j,\tau} = \begin{cases} \sum_{t=\tau_j^0+1}^{\tau_j^0+\tau} [\ell_t(\boldsymbol{\theta}_j^0, p_j^0) - \ell_t(\boldsymbol{\theta}_{j+1}^0, p_{j+1}^0)] & \tau > 0, \\ 0 & \tau = 0, \\ \sum_{t=\tau_j^0+\tau}^{\tau_j^0-1} [\ell_t(\boldsymbol{\theta}_{j+1}^0, p_{j+1}^0) - \ell_t(\boldsymbol{\theta}_j^0, p_j^0)] & \tau < 0 \end{cases} \quad (2.6)$$

is a double-sided random walk. In particular, $\hat{\tau}_j^{(3)} = \tau_j^0 + O_p(1)$.

2.4 CIs' approximation for the final change-points estimates.

Although Theorem 3 deduces the asymptotic distribution $\arg \max_{\tau} W_{j,\tau}$ of $(\hat{\tau}_j^{(3)} - \tau_j^0)$, it is difficult to use in practice. According to the arguments of Cui et al. (2021) about the approximation distribution of $\arg \max_{\tau} W_{j,\tau}$, we have the following theorem when change is small.

Theorem 4 *Let $\mathbf{d}_j = \boldsymbol{\theta}_j^0 - \boldsymbol{\theta}_{j+1}^0$, under the setting of Theorem 3, if $\|\mathbf{d}_j\| \rightarrow 0$ as $n \rightarrow 0$, then*

$$\hat{\Delta}_j^{-1}(\hat{\tau}_j^{(3)} - \tau_j^0) \xrightarrow{d} \arg \max_{r \in \mathbb{R}} \mathcal{B}(r) - \frac{1}{2}|r|, \quad (2.7)$$

where

$$\hat{\Delta}_j = (\hat{\mathbf{d}}_j^\top \hat{\mathbf{J}}_j \hat{\mathbf{d}}_j)^{-2} (\hat{\mathbf{d}}_j^\top \hat{\mathbf{I}}_j \hat{\mathbf{d}}_j) \quad \text{and} \quad \hat{\mathbf{d}}_j = \hat{\boldsymbol{\theta}}_j - \hat{\boldsymbol{\theta}}_{j+1}.$$

With $\hat{\mathbf{J}}_j = \hat{\mathbf{J}}_{4h}(\hat{\tau}_j^{(3)} - 2h, \hat{\boldsymbol{\theta}}_{j+1}^{(3)})$ and $\hat{\mathbf{I}}_j = \hat{\mathbf{I}}_{4h}(\hat{\tau}_j^{(3)} - 2h, \hat{\boldsymbol{\theta}}_{j+1}^{(3)})$ are defined in (2.2) and (2.3). $\mathcal{B}(r)$ is the two-sided standard Brownian motion in \mathbb{R} .

Remark 3 If the orders of GCINAR in two consecutive segments, denoted by \hat{p}_j and \hat{p}_{j+1} , are not equal, then the GCINAR parameter vectors $\hat{\theta}_j$ and $\hat{\theta}_{j+1}$ will be interpreted as being in the order of $\max(\hat{p}_j, \hat{p}_{j+1})$.

To simplify notation, denote $V = \arg \max_{r \in \mathbb{R}} \mathcal{B}(r) - \frac{1}{2}|r|$. The distribution of V has been studied by Yao (1987), who demonstrates that V has a symmetric distribution, and for $a > 0$, the distribution function is given by

$$P(V \leq a) = 1 + \sqrt{\frac{a}{2\pi}} \exp\left(-\frac{a}{8}\right) + \frac{3}{2} \exp(a) \Phi\left(-\frac{3}{2}\sqrt{a}\right) - \frac{1}{2}(a+5) \Phi\left(-\frac{1}{2}\sqrt{a}\right),$$

where $\Phi(\cdot)$ represents the standard normal distribution function. Let $F_{\alpha/2}$ be the $(1 - \alpha/2)$ th quantile of V , that is $P(V \leq F_{\alpha/2}) = 1 - \alpha/2$, and $F_{0.05} = 7.6873$. Then an approximate $100(1 - \alpha)\%$ confidence interval (CI) for τ_j^0 can be constructed by

$$CI = \left[\hat{\tau}_j^{(3)} - \lfloor \hat{\Delta}_j F_{\alpha/2} \rfloor - 1, \hat{\tau}_j^{(3)} + \lfloor \hat{\Delta}_j F_{\alpha/2} \rfloor + 1 \right], \quad (2.8)$$

where $\lfloor a \rfloor$ is the largest integer not greater than a . Since the minimum distance between change-points is much larger than the maximum window radius h , i.e. $n\epsilon_\tau/h \rightarrow \infty$, the distance between the extended local windows $EW_j(h)$ s diverge to ∞ . Under H.1, the GCINAR(p) process is strongly mixing, the CIs constructed are asymptotically independent. According to a Bonferroni-type argument, one can construct an asymptotically correct $1 - \alpha$ simultaneous CI covering all $\hat{m}^{(2)}$ change-points by using a collection of $(1 - \alpha)^{1/\hat{m}^{(2)}}$ CIs for each of a set of $\hat{m}^{(2)}$ change-points.

2.5 Bootstrap approximation

Due to the restriction of condition $\|\mathbf{d}_j\| \rightarrow 0$ in Theorem 4, some scholars have confirmed the fact that the asymptotic theory of change-points estimates given by Theorem 3 in Cui et al. (2021) provides a poor approximation to the actual multimodal finite sample distribution under small parameter changes, thus the asymptotic theory of change-points estimates given by Theorem 4 also has this problem. Furthermore, the pivotal approximations in CI (2.8) work unsatisfactorily under medium and large parameter changes. Considering these issues, inspired by Ng et al. (2022), we propose the following two bootstrap procedures, parametric bootstrap and block bootstrap, to construct the CI for the change-point τ_j^0 .

Parametric Bootstrap Algorithm (PBA). The basic idea of parametric bootstrap is to first use the model based on the PQML estimate to simulate replicated samples of the series before and after the change-point. Then the bootstrap samples are used to approximate the asymptotic distribution of $\arg \max_{\tau} W_{j,\tau}$ in Theorem 3. The detail of PBA for τ_j^0 is given introduction in the following.

Parametric Bootstrap Algorithm:

Input: The significance level α , resample size n_p and the sampling number B .

The order estimates $\hat{p}_j^{(2)}$ and $\hat{p}_{j+1}^{(2)}$ from “Second step”.

The parameter estimates $\hat{\theta}_j$ and $\hat{\theta}_{j+1}$, change-point estimate $\hat{\tau}_j^{(3)}$ from “Third step”.

Simulate replicated samples:

for $s = 1 : B$

simulate time series sample $\{\tilde{X}_t^{(s)}\}_{t=1}^{n_p+1}$ follow the special GCINAR($\hat{\theta}_j, \hat{p}_j^{(2)}$) model,

simulate time series sample $\{\tilde{X}_t^{(s)}\}_{t=n_p+2}^{2n_p+1}$ follow the special GCINAR($\hat{\theta}_{j+1}, \hat{p}_{j+1}^{(2)}$) model,

join them to form the resampled process $\{\tilde{X}_t^{(s)}\}_{t=1}^{2n_p+1}$.

end

Approximate the asymptotic distribution of $\arg \max_{\tau} W_{j,\tau}$:

for $s = 1 : B$

obtain $\{\tilde{\ell}_t^{(s)}(\hat{\theta}_j, \hat{p}_j^{(2)})\}_{t=1}^{2n_p+1}$ and $\{\tilde{\ell}_t^{(s)}(\hat{\theta}_{j+1}, \hat{p}_{j+1}^{(2)})\}_{t=1}^{2n_p+1}$ from (2.1) based on the sample $\{\tilde{X}_t^{(s)}\}_{t=1}^{2n_p+1}$,
compute the double-sided random walk,

$$\tilde{W}_{j,\tau}^{(s)} = \begin{cases} \sum_{t=n_p+2}^{n_p+1+\tau} [\tilde{\ell}_t^{(s)}(\hat{\theta}_j, \hat{p}_j^{(2)}) - \tilde{\ell}_t^{(s)}(\hat{\theta}_{j+1}, \hat{p}_{j+1}^{(2)})] & \tau > 0, \\ 0 & \tau = 0, \\ \sum_{t=n_p+1+\tau}^{n_p} [\tilde{\ell}_t^{(s)}(\hat{\theta}_{j+1}, \hat{p}_{j+1}^{(2)}) - \tilde{\ell}_t^{(s)}(\hat{\theta}_j, \hat{p}_j^{(2)})] & \tau < 0, \end{cases} \quad (2.9)$$

compute $\tilde{\tau}_{j,n_p}^{(s)} = \arg \max_{\tau \in \{-n_p, \dots, n_p\}} \tilde{W}_{j,\tau}^{(s)}$.

end

Obtain the final CI_j^{PBA} :

compute the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ percentiles of the sample $\{\tilde{\tau}_{j,n_p}^{(1)}, \dots, \tilde{\tau}_{j,n_p}^{(B)}\}$, denoted by \tilde{l} and \tilde{u} ,

obtain the parametric bootstrap $100(1 - \alpha)\%$ CI for the change-point τ_j^0 : $CI_j^{PBA} = [\hat{\tau}_j^{(3)} - \tilde{u}, \hat{\tau}_j^{(3)} - \tilde{l}]$.

Block Bootstrap Algorithm (BBA). Unlike PBA, block bootstrap obtain replicate samples from joining block subsamples of the observations before and after the change-point. Then similar to PBA to get the final CI_j^{BBB} . The detail of BBA for τ_j^0 is given introduction in the following.

Block Bootstrap Algorithm:

Input: The significance level α , resample size n_b and the sampling number B .

The data set: $\{X_t\}_{t=\hat{\tau}_j^{(3)}+1}^{\hat{\tau}_{j+1}^{(3)}}$.

The order estimates $\hat{p}_j^{(2)}$ and $\hat{p}_{j+1}^{(2)}$ from ‘‘Second step’’.

The parameter estimates $\hat{\theta}_j$ and $\hat{\theta}_{j+1}$, change-point estimate $\hat{\tau}_j^{(3)}$ from ‘‘Third step’’.

Simulate replicate samples:

for $s = 1 : B$

sample the block of observations $\{X_t^{*(s)}\}_{t=1}^{n_b+1}$ from the original data set $\{X_t\}_{t=\hat{\tau}_j^{(3)}+1}^{\hat{\tau}_{j+1}^{(3)}}$,

sample the block of observations $\{X_t^{*(s)}\}_{t=n_b+1}^{2n_b+1}$ from the original data set $\{X_t\}_{t=\hat{\tau}_j^{(3)}+1}^{\hat{\tau}_{j+1}^{(3)}}$,

join them to form the resampled process $\{X_t^{*(s)}\}_{t=1}^{2n_b+1}$.

end

Approximate the asymptotic distribution of $\arg \max_{\tau} W_{j,\tau}$:

for $s = 1 : B$

obtain $\{\ell_t^{*(s)}(\hat{\theta}_j, \hat{p}_j^{(2)})\}_{t=1}^{2n_b+1}$ and $\{\ell_t^{*(s)}(\hat{\theta}_{j+1}, \hat{p}_{j+1}^{(2)})\}_{t=1}^{2n_b+1}$ from (2.1) based on the sample $\{X_t^{*(s)}\}_{t=1}^{2n_b+1}$,
compute the double-sided random walk,

$$W_{j,\tau}^{*(s)} = \begin{cases} \sum_{t=n_b+2}^{n_b+1+\tau} \left[\ell_t^{*(s)}(\hat{\theta}_j, \hat{p}_j^{(2)}) - \ell_t^{*(s)}(\hat{\theta}_{j+1}, \hat{p}_{j+1}^{(2)}) \right] & \tau > 0, \\ 0 & \tau = 0, \\ \sum_{t=n_b+1+\tau}^{n_b} \left[\ell_t^{*(s)}(\hat{\theta}_{j+1}, \hat{p}_{j+1}^{(2)}) - \ell_t^{*(s)}(\hat{\theta}_j, \hat{p}_j^{(2)}) \right] & \tau < 0, \end{cases} \quad (2.10)$$

compute $\tau_{j,n_b}^{*(s)} = \arg \max_{\tau \in \{-n_b, \dots, n_b\}} W_{j,\tau}^{*(s)}$.

end

Obtain the final CI_j^{BBA} :

compute the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ percentiles of the sample $\{\tau_{j,n_b}^{*(1)}, \dots, \tau_{j,n_b}^{*(B)}\}$, denoted by l^* and u^* ,

obtain the block bootstrap $100(1 - \alpha)\%$ CI for the change-point τ_j^0 : $CI_j^{BBA} = [\hat{\tau}_j^{(3)} - u^*, \hat{\tau}_j^{(3)} - l^*]$.

Next, we consider the validity of the parametric and block bootstrap procedures. Assume that, in the parametric bootstrap procedure, the j th segment GCINAR model in the MCP-GCINAR model is composed of a specific thinning operator (‘‘ \diamond ’’) plus a innovation Z_t , that is $X_{t,j} = \sum_{k=1}^p \beta_{k,j} \diamond X_{t-k,j} + Z_{t,j}$. And the form of the thinning operator as well as the distribution of the innovation Z_t are known, then under the Assumption H.6, we have the following Theorem 5. For notational simplicity, we omit the superscript (s) that indicates the s th realization of the B bootstrap simulations.

H. 6 Denote $\mathcal{L}_{Z_{t,j},\beta}$ as the distribution of the j th segment innovation process $Z_{t,j}$ with the parameter $\beta \in \mathbb{R}^+$ (\mathbb{R}^+ is the set of positive real numbers). Assume that $\mathbb{E}(Z_{t,j})^{\epsilon_Z} < \infty$ for some $\epsilon_Z \in \mathbb{N}$ and $0 < \mathcal{L}_{Z_{t,j},\beta}(0) < 1$ holds, where $\mathcal{L}_{Z_{t,j},\beta}(k) = P(Z_{t,j} = k), k \in \mathbb{N}$. For all $\epsilon > 0$, there exist a $c > 0$ such that for all $\beta' \in \mathbb{R}^+$ with $|\beta' - \beta| < c$, there is $\sum_{k=0}^{\infty} |\mathcal{L}_{Z_{t,j},\beta}(k) - \mathcal{L}_{Z_{t,j},\beta'}(k)| < \epsilon$. Furthermore, assume that there exists a neighborhood $M_{\beta^0,c} = \{\beta \mid |\beta - \beta^0| < c, \beta \in \mathbb{R}^+\}$ of β^0 such that $\sum_{k=0}^{\infty} k^{\epsilon_Z} \mathcal{L}_{Z_{t,j},\beta}(k) < \infty$ holds uniformly on $M_{\beta^0,c}$.

Theorem 5 Under the setting in Theorem 3, and for all segment, H.6 holds, then for any fixed reasonably large $n_p \in \mathbb{N}$, $n_b \in \mathbb{N}$, and $n_b < \min(\hat{\tau}_{j+1}^{(3)} - \hat{\tau}_j^{(3)}, \hat{\tau}_j^{(3)} - \hat{\tau}_{j-1}^{(3)})$, we have

$$\sup_{x \in \mathbb{R}} \left| \tilde{\mathbb{P}}(\tilde{\tau}_{j,n_p} \leq x) - \mathbb{P}(\arg \max_{\tau} W_{j,\tau} \leq x) \right| \xrightarrow{p} 0, \quad (2.11)$$

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}^*(\tilde{\tau}_{j,n_b} \leq x) - \mathbb{P}(\arg \max_{\tau} W_{j,\tau} \leq x) \right| \xrightarrow{p} 0, \quad (2.12)$$

where $W_{j,\tau}$ is defined in (2.6), and $\tilde{\mathbb{P}}$ and \mathbb{P}^* denote probability measures, respectively, under the two bootstrap schemes conditional on the original data set $\{X_t\}_{t=\hat{\tau}_{j-1}^{(3)}+1}^{\hat{\tau}_{j+1}^{(3)}}$.

Note that a lot of model-formal assumptions, such as the form of the thinning operator and the distribution of $Z_{t,j}$, restrict the application of PBA because we usually do not have this information a priori. To task this situation, we assume that the inference of change-points is obtained from the multiple change-points Poisson INAR (MCP-PINAR) model, where the j -th segment in Model (1.2) is defined as follows,

$$X_{t,j} = \sum_{k=1}^{p_j} \beta_{k,j} \bullet X_{t-k,j} + Z_{t,j}, \quad (2.13)$$

where “ \bullet ” is the Poisson thinning operator, defined as: $\beta \circ X = \sum_{i=1}^X B_i(\beta)$, the counting series $\{B_i(\beta)\}$ is an independent and identically distributed (i.i.d.) Poisson random sequence with mean β . And $\{Z_{t,j}\}$ is a sequence of i.i.d. Poisson random variables with mean $\beta_{0,j}$. Then, we do the “simulate replicated samples” step based on the MCP-PINAR model, and the $100(1 - \alpha)\%$ CI for the change-point τ_j^0 can be obtained after implementing PBA.

Remark 4 This setup is reasonable and convenient. On the one hand, this is a continuation of the PQML idea, which assumes that the conditional distribution is a Poisson distribution. So since there is no prior distribution information, we might as well continue with this idea. Moreover, the subsequent simulation results also verify that the performance of confidence interval constructed by the PBA is still satisfactory when the wrong model is specified. Another advantage of this is that Theorem 5 can be held without assumption H.6, which is a complex condition after all.

3 Computational issues

In this section, the tuning parameters and computational complexity are discussed.

3.1 Tuning parameters setting in the LRSM

The tuning parameters involved in the LRSM are p_{\max} , m_{\max} and h . As discussed in Section 2.1, as long as p_{\max} , m_{\max} are large enough, their values have little effect on the estimation results. In contrast, the LRSM estimates are mainly affected by the window parameter h . Smaller windows are more sensitive to changes occurring in short time durations, and are more likely to guarantee the condition, $2h < n\epsilon_\tau$, while larger windows are more sensitive to small changes. It is theoretically crucial to choose an $h > d(\log n)^3$ for Theorem 1 to hold, where d is an unknown window coefficient constant. Based on the experience of scholars as well as our simulation studies, setting $h = \max(n/20, (\log n)^4/25)$ will generally yield satisfactory results with different models and sample sizes.

However, such a setting may not guarantee the condition $2h < n\epsilon_\tau$ hold, when the sample size is small and the change-points are dense. To solve this, we can combine multiple h in the first step of LRSM to get potential change-points. The detailed operation is explained as follows,

Setting window coefficient set \tilde{d} (such as $\tilde{d} = \{1, 2\}$), and $h_{mix} = \{h_i\}_{i=1}^{|\tilde{d}|}$, where $h_i = \tilde{d}_i(\log n)^4/25$ (such as $h_1 = (\log n)^4/25, h_2 = 2(\log n)^4/25$). Then start from $i = 1$ to $i = |\tilde{d}|$, get the potential change-points set $\hat{\mathcal{J}}^{(1)}(h_i)$, combine them and get the final $\hat{\mathcal{J}}^{(1)}(h)$.

Similar aggregation procedures can be found in Ng et al. (2022). Empirical evidence suggests that, setting $\tilde{d} = \{0.2, 0.4, 0.6, 0.8, 1, 1.2\}$ for small sample size and $\tilde{d} = \{1, 2, 3, 4, 5, 6\}$ for large sample size usually results in more satisfactory performance.

3.2 Tuning parameters setting in Bootstrap procedure

The tuning parameters involved in the two bootstrap methods are the sampling number B , re-sample size n_p and n_b . For the sampling number B , large enough to guarantee reliable results, such as $B = 1000$. For the parametric bootstrap, we set $n_p = n/2$ to ensure that n_p is sufficiently large to include the maxima of the random walk. For the block bootstrap, the following iterative procedure is used to find a data-driven or adaptive block bandwidth n_b :

- **Step 1:** Compute a $(1 - \alpha)\%$ CI using CIs' approximation or the PBA method. Denote the width of the CI by l . Set initial block bandwidth $n_b = 2l$.

- **Step 2:** Perform the BBA to obtain $\{\tau_{j,n_b}^{*(s)}\}, s = 1, 2, \dots, B$. If the proportion of the sample $\{\tau_{j,n_b}^{*(s)}\}$ lying in either of the regions $[\hat{\tau}_j^{(3)} - n_b, \hat{\tau}_j^{(3)} - (1 - \alpha)n_b]$ or $[\hat{\tau}_j^{(3)} + (1 - \alpha)n_b, \hat{\tau}_j^{(3)} + n_b]$ is greater than $(\alpha/2)\%$, increase n_b by l .

- **Step 3:** Repeat **Step 2:** until a final block bandwidth n_b is found. Set this n_b as the adaptive bandwidth.

3.3 Computational complexity

The computational complexity of the three-step LRSM has been discussed by many scholars, such as Yau and Zhao (2016), Ng et al. (2022). Since the size of $W_t(h)$ is $2h$, the computational complexity of $S_h(t)$ for each t is order $O(h)$, and the computational complexity of the first step in LRSM is $O(nh)$. The computational complexity of the second step, the OP Algorithm, is $O((\hat{m}^{(1)})^2n)$ and the computational complexity of the third step for evaluating extended local

windows is $O(\hat{m}^{(2)}h^2)$. Therefore, the total computational complexity in the three-step LRSM procedure is $O(hn)$.

4 Simulation

To evaluate the finite-sample performance of the proposed LRSM and Bootstrap procedures, we conducted extensive simulation studies and split the simulation studies into the following five parts. In the first part, we consider the sensitivity of the LRSM to its tuning parameter, the windows parameter h . In the second simulation, we compare the performance of LRSM with GA and penQLIK. Furthermore, the performance of LRSM to deal with a large number of change-points under different sample size is compared in the third part. Finally, the performance of the CIs constructed by the parametric and block bootstrap approximations proposed in Section 2.5 is examined, and compared to the approximation method of CIs, proposed in Theorem 4.

We first introduce the following notations and evaluation metrics that are primarily used in simulations.

Notations:

- Denote “ \circ ” as the binomial thinning operator, which is proposed by [Steutel and Van Harn \(1979\)](#) and defined as: $\beta \circ X = \sum_{i=1}^X B_i(\beta)$, the counting series $\{B_i(\beta)\}$ is an independent and identically distributed (i.i.d.) Bernoulli random sequence with mean β .

- Denote “ $*$ ” as the negative binomial thinning operator of [Ristić et al. \(2009\)](#), it is defined as $\beta * X = \sum_{i=1}^X B_i(\beta/(1+\beta))$, the counting series $\{B_i(\beta/(1+\beta))\}$ is a sequence of i.i.d. geometric random variables with mean β ,

- Denote the Poisson distribution with mean β by $\text{Poi}(\beta)$, the Geometric distribution with mean β by $\text{Geo}(\beta/(1+\beta))$.

- Let $\mathbf{v}_n = (v_1, \dots, v_m)$, $0 < v_1 < \dots < v_m < 1$, satisfy $\tau_j = [v_j n]$, where $[x]$ is the greatest integer that is less than or equal to x .

Evaluation Metrics:

- Denote the true positive rate of m by $\text{TPR}(m)$, i.e., the proportion of informative points are correctly identified.

- Define the following two type evaluation metrics to measure the under-segmentation error and the over-segmentation error of the change-points location estimate $\hat{\mathbf{v}}_n$, respectively.

$$\zeta_u(\mathbf{v}^0 | \hat{\mathbf{v}}_n) = \sup_{b \in \hat{\mathbf{v}}_n} \inf_{a \in \mathbf{v}^0} |a - b|, \quad \zeta_o(\hat{\mathbf{v}}_n | \mathbf{v}^0) = \sup_{b \in \mathbf{v}^0} \inf_{a \in \hat{\mathbf{v}}_n} |a - b| \quad (\text{Boysen et al., 2009}).$$

A desirable estimate should be able to balance both quantities.

- Define the following metric to measure the location accuracy of estimated change-points.

$$\zeta_d(\hat{\mathbf{v}}_n, \mathbf{v}^0) = \frac{1}{|\mathbf{v}^0|} \sum_{v_k^0 \in \mathbf{v}^0} \min_{\hat{v}_j \in \hat{\mathbf{v}}_n} |\hat{v}_j - v_k^0| \quad (\text{Chen et al., 2021}),$$

which is the distance from the estimated set $\hat{\mathbf{v}}_n$ and the true change-points set \mathbf{v}^0 .

All simulations are carried out using the MATLAB software. The empirical results displayed in the tables are computed over 1000 replications.

Table 1: Sensitivity analysis for the window h in the LRSM based on Model (A1).

d	$h = \max(\frac{n}{20}, d\frac{(\log n)^4}{25})$	Sample size (n)	TRP(m)	$\zeta_u(\mathbf{v}^0 \hat{\mathbf{v}}_n)$	$\zeta_o(\hat{\mathbf{v}}_n \mathbf{v}^0)$	$\zeta_d(\hat{\mathbf{v}}_n, \mathbf{v}^0)$	Average time (s)
0.5	29	500	0.8070	0.0410	0.0973	0.1645	0.7439
0.5	50	1000	0.9930	0.0284	0.0287	0.0484	1.5489
0.5	100	2000	0.9980	0.0107	0.0105	0.0178	3.6764
1	59	500	0.7110	0.0370	0.1219	0.2059	0.5506
1	91	1000	0.9760	0.0250	0.0316	0.0518	1.5935
1	133	2000	0.9890	0.0110	0.0102	0.0172	4.3840
1.5	89	500	0.6710	0.0274	0.1237	0.2071	0.7409
1.5	136	1000	0.9690	0.0163	0.0248	0.0414	2.2516
1.5	200	2000	0.9950	0.0063	0.0060	0.0108	8.5836
2	119	500	0.6400	0.0234	0.1290	0.2111	1.6022
2	182	1000	0.9580	0.0113	0.0238	0.0401	4.9702
2	267	2000	0.9990	0.0055	0.0054	0.0100	13.1729
2.5	149	500	0.3660	0.0383	0.2214	0.3714	1.6629
2.5	227	1000	0.9030	0.0118	0.0398	0.0647	5.8175
2.5	333	2000	0.9900	0.0054	0.0078	0.0135	14.9413
3	178	500	0.0000	0.0892	0.3831	0.7040	0.7435
3	273	1000	0.8010	0.0132	0.0716	0.1127	3.8218
3	400	2000	0.9880	0.0053	0.0088	0.0152	11.0657

d_{mix}	$h_{mix} = d_{mix}\frac{(\log n)^4}{25}$	Sample size (n)	TRP(m)	$\zeta_u(\mathbf{v}^0 \hat{\mathbf{v}}_n)$	$\zeta_o(\hat{\mathbf{v}}_n \mathbf{v}^0)$	$\zeta_d(\hat{\mathbf{v}}_n, \mathbf{v}^0)$	Average time (s)
{0.2, 0.4, 0.6, 0.8, 1, 1.2}	{12, 24, 36, 48, 59, 71}	500	0.9130	0.0276	0.0529	0.0909	3.7002
{0.2, 0.4, 0.6, 0.8, 1, 1.2}	{19, 37, 55, 73, 91, 110}	1000	1.0000	0.0166	0.0166	0.0301	9.9580
{0.2, 0.4, 0.6, 0.8, 1, 1.2}	{27, 54, 80, 107, 133, 160}	2000	0.9990	0.0099	0.0096	0.0166	31.0050

4.1 Sensitivity analysis for the tuning parameter h in the LRSM

To study the choice of the tuning parameter h in the three-step LRSM, we conducted a sensitivity analysis by using data generated from Model (A1) with the sample sizes $n = 500, 1000, 2000$. $h = \max(n/20, d(\log n)^4/25)$ was considered using different values of d . Furthermore, the combine multiple h , introduced in Section 3.1, was also studied with $d_{mix} = \{0.2, 0.4, 0.6, 0.8, 1, 1.2\}$. The results are summarized in Table 1.

Model (A1) is as follows:

$$X_t = \begin{cases} 0.5 \circ X_{t-1,1} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(0.5) \quad 0 < t \leq \tau_1^0, \\ 0.126 \circ X_{t-\tau_1^0-1,2} + 0.254 \circ X_{t-\tau_1^0-2,2} + 0.297 \circ X_{t-\tau_1^0-3,2} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(1) \quad \tau_1^0 < t \leq \tau_2^0, \\ 0.4 \circ X_{t-\tau_2^0-1,3} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(2) \quad \tau_2^0 < t \leq n. \end{cases}$$

where $(\tau_1^0, \tau_2^0) = ([0.3n], [0.6n])$.

The following conclusions can be drawn from Table 1. In terms of the calculation time (Average time) and the metric TRP, smaller windows reduce computing cost, and are more sensitive to changes occurring in short time durations. In terms of the metrics $\zeta_u(\mathbf{v}^0|\hat{\mathbf{v}}_n)$, $\zeta_o(\hat{\mathbf{v}}_n|\mathbf{v}^0)$, and $\zeta_d(\hat{\mathbf{v}}_n, \mathbf{v}^0)$, larger windows balance the under-segmentation error and the over-segmentation error, improving the accuracy of estimates. Such as, $n = 2000$, $h = 200, 267$ have the optimal accuracy 0.0108 and 0.01 in the entire result table. In particular, the mixture window h_{mix} has excellent performance, of course, this advantage is at the cost of increasing the calculation cost. In addition, the LRSM is robust given a mild violation of $2h < n\epsilon_\tau$. Such as, with the sample size $n =$

1000, $\tau_1^0 = 300$ and $\tau_2^0 = 600$, the minimum distance being 300, the LRSM requires $h < 300/2 = 150$ for consistent estimate. However, the estimate remains satisfactory when $h = 273$. This may be related to the findings that the likelihood ratio test statistic can consistently estimate a single change-point in the presence of multiple changes. Thus, although the choice of h will affect the accuracy of estimation results, it is not sensitive for estimation purposes.

4.2 Comparing LRSM and existing methods

In this subsection, we compare the performance of GA (Sheng and Wang, 2023), penalized contrast QLIK (penQLIK, Diop and Kengne (2021b)), and the LRSM in implementing change-point inference. We consider Model (A1) with sample sizes of $n = 500, 1000, 2000$. Additionally, we include the penQLIK method proposed by Diop and Kengne (2021b) for comparison with the LRSM. The penQLIK is a data-driven method that utilizes the slope heuristic procedure (Baudry et al., 2012) to calibrate the penalty term. The penQLIK criterion is minimized using a dynamic programming algorithm. It is important to note that penQLIK does not consider the change of orders p . Therefore, we implement penQLIK by considering a sufficiently large $p = 5$, $q = 5$ (denoted as penQLIK(5,5)).

Table 2: Comparison results of the penQLIK, GA and LRSM based on Model (A1).

Method	Sample size(n)	TRP(m)	$\zeta_u(\mathbf{v}^0 \hat{\mathbf{v}}_n)$	$\zeta_o(\hat{\mathbf{v}}_n \mathbf{v}^0)$	$\zeta_d(\hat{\mathbf{v}}_n, \mathbf{v}^0)$	Average time (s)
penQLIK(5,5)	500	0.4190	0.1250	0.0440	0.0772	129.9832
	1000	0.7140	0.0916	0.0172	0.0298	935.9727
	2000	0.8290	0.1087	0.0073	0.0127	5697.9496
GA	500	0.9090	0.0174	0.0437	0.0721	1052.2157
	1000	0.9990	0.0087	0.0090	0.0162	3217.8543
	2000	1.0000	0.0043	0.0043	0.0078	3688.3210
LRSM	500	0.8070	0.0410	0.0973	0.1645	0.7439
	1000	0.9930	0.0284	0.0287	0.0484	1.5489
	2000	0.9980	0.0107	0.0105	0.0178	3.6764

As you can see from Table 2, GA is the top performer in almost all evaluation metrics. It has a smaller $\zeta_d(\hat{\boldsymbol{\tau}}, \tilde{\boldsymbol{\tau}}^0)$, the highest TRP(m), and balances $\zeta_u(\tilde{\boldsymbol{\tau}}^0|\hat{\boldsymbol{\tau}})$ and $\zeta_o(\hat{\boldsymbol{\tau}}|\tilde{\boldsymbol{\tau}}^0)$. However, the results of the LRSM are also acceptable, especially in Model (A1) with sample size $n = 2000$. It balances $\zeta_u(\tilde{\boldsymbol{\tau}}^0|\hat{\boldsymbol{\tau}})$ and $\zeta_o(\hat{\boldsymbol{\tau}}|\tilde{\boldsymbol{\tau}}^0)$ and the value of TRP(m) is closed to 1. In addition, although penQLIK can handle this sequence-changing model, it produces a large number of under-segmentation errors. In other words, penQLIK does not compress the model any better than the MDL criterion. Taking into account all simulation results, LRSM has good and stable performance. Especially considering the calculation time, the advantage is more obvious. In summary, GA achieves the global optimization of MDL, so the optimal results are given in terms of estimation accuracy. However, when the sample size is large and the number of change-points is relatively small, GA may

require computational cost, and the LRSM can serve as a reliable and computationally efficient alternative.

4.3 Performance of the LRSM based on difference scenarios

In this subsection, the performance of the LRSM is studied based on the scenario of a large number of dense change-points with a finite-sample (sample size $n = 2000$) and in the scenario of a long time series (sample size $n = 10000$). Two types of the MCP-GCINAR model are considered, B-type models have a Poisson distribution innovation and a binomial thinning operator, while C-type models have a negative binomial thinning operator and a Geometric distribution innovation. Models (B1) - (B9) represent the case of B-type models under $m_0 = 1$ to $m_0 = 9$, and Models (C1) - (C9) represent the case of C-type models under change-points numbers $m_0 = 1$ to $m_0 = 9$.

While sample size $n = 2000$, the LRSM is implemented by the $h = (\log n)^4/25 = 133$ and $h_{mix} = \{27, 54, 80, 107, 133, 160\}$, and while in the scenario of a long-time series, sample size $n = 10000$, the LRSM is implemented by the $h = (\log n)^4/25 = 287$. All results of Models (B1) - (B9) are summarized in Table 3 and Table 4. The specific forms of Models (B1) - (C9) are in the Appendix to save space.

Table 3 reveals that the application of h_{mix} can improve LRSM's performance, because it significantly improves the metric TRP, reduces the metrics $\zeta_u(\mathbf{v}^0|\hat{\mathbf{v}}_n)$, $\zeta_o(\hat{\mathbf{v}}_n|\mathbf{v}^0)$, and $\zeta_a(\hat{\mathbf{v}}_n, \mathbf{v}^0)$, and balances the under-segmentation error and the over-segmentation error. But this apparent advantage also comes at the cost of increasing the computational burden. Therefore, in the scenario of large sample size ($n = 10000$), Table 4 only summarizes the results of applying LRSM with $h = 287$. After all, considering h_{mix} , a lot of time and computing resources are needed. However, it can be seen that in such a situation, the general h can handle the change-point problem of the MCP-GCINAR models well.

4.4 Constructing confidence intervals

In this subsection, we examine the coverage accuracy of the CIs for the change-points estimates based on the correct estimates of Models (B1) - (C2) in the previous subsection. For each of the cases, CIs using CIs approximation (2.7), parametric bootstrap and block bootstrap with adaptive block bandwidth are generated.

Table 5 provides a summary of the CIs results based on the LRSM procedure, including the median and mean of the estimate $\hat{\tau}_j^{(3)}$, the range of the middle 90% of the final estimates (90% Range) average of the end-points (Mean of 90% CI), and coverage probability (Coverage Prob.) of the CIs.

From the results in Table 5, it can be seen that the CIs of both PBA and BBA methods have quite accurate coverage probability, which are close to the nominal level of 90%. Particularly, the CIs constructed from the two bootstrap procedures have more accurate coverage probability than that obtained from the approximation method (2.7). This is consistent with the previous discussion, the pivotal approximations in CI (2.8) work unsatisfactorily under medium and large

Table 3: Performance of the three-step LRSM based on Models (B1) - (C9) with sample size $n = 2000$, $h = 133$ and $h_{mix} = \{27, 54, 80, 107, 133, 160\}$.

Model m_0	window	Average					Model m_0	window	Average						
		TRP(m)	$\zeta_u(\mathbf{v}^0 \hat{\mathbf{v}}_n)$	$\zeta_o(\hat{\mathbf{v}}_n \mathbf{v}^0)$	$\zeta_d(\hat{\mathbf{v}}_n, \mathbf{v}^0)$	time (s)			TRP(m)	$\zeta_u(\mathbf{v}^0 \hat{\mathbf{v}}_n)$	$\zeta_o(\hat{\mathbf{v}}_n \mathbf{v}^0)$	$\zeta_d(\hat{\mathbf{v}}_n, \mathbf{v}^0)$	time (s)		
(B1)	1	h	0.9890	0.0034	0.0026	0.0052	6.9152	(C1)	1	h	0.9590	0.0085	0.0037	0.0073	6.9470
		h_{mix}	0.9990	0.0026	0.0023	0.0047	49.3130			h_{mix}	0.9780	0.0078	0.0031	0.0062	50.2625
(B2)	2	h	0.9860	0.0066	0.0050	0.0084	7.6469	(C2)	2	h	0.9540	0.0124	0.0062	0.0108	7.6078
		h_{mix}	0.9980	0.0028	0.0027	0.0049	53.1847			h_{mix}	0.9710	0.0084	0.0044	0.0081	51.9576
(B3)	3	h	0.9860	0.0052	0.0065	0.0089	8.0786	(C3)	3	h	0.9700	0.0075	0.0064	0.0095	8.0287
		h_{mix}	1.0000	0.0034	0.0034	0.0056	43.0814			h_{mix}	0.9800	0.0064	0.0048	0.0074	43.2224
(B4)	4	h	0.9740	0.0048	0.0099	0.0127	8.8324	(C4)	4	h	0.9710	0.0073	0.0079	0.0109	8.8331
		h_{mix}	0.9970	0.0041	0.0040	0.0066	46.3998			h_{mix}	0.9560	0.0085	0.0055	0.0085	46.1871
(B5)	5	h	0.7260	0.0073	0.0687	0.0683	9.3995	(C5)	5	h	0.9500	0.0086	0.0136	0.0152	9.4700
		h_{mix}	0.9580	0.0070	0.0135	0.0150	48.0895			h_{mix}	0.8810	0.0135	0.0074	0.0101	48.1504
(B6)	6	h	0.0110	0.0113	0.1640	0.2514	8.7535	(C6)	6	h	0.0860	0.0144	0.1127	0.1484	9.0956
		h_{mix}	0.3690	0.0193	0.0868	0.0868	41.7372			h_{mix}	0.8960	0.0229	0.0463	0.0497	42.9014
(B7)	7	h	0.0430	0.0180	0.4030	0.8972	8.0163	(C7)	7	h	0.2140	0.0353	0.1097	0.1374	10.3091
		h_{mix}	0.7630	0.0254	0.0604	0.0857	41.4308			h_{mix}	0.6180	0.0456	0.0384	0.0436	45.9520
(B8)	8	h	0.0090	0.0258	0.6544	1.6095	7.3036	(C8)	8	h	0.1220	0.0231	0.1261	0.1748	10.6836
		h_{mix}	0.6390	0.0264	0.1274	0.2370	42.7765			h_{mix}	0.7900	0.0300	0.0407	0.0442	45.5573
(B9)	9	h	0.0010	0.0147	0.4927	1.2206	7.7363	(C9)	9	h	0.0310	0.0196	0.1323	0.2141	7.0678
		h_{mix}	0.2180	0.0216	0.2059	0.3833	46.0448			h_{mix}	0.8770	0.0298	0.0528	0.0550	42.8664

Table 4: Performance of the three-step LRSM based on Models (B1) - (C9) with sample size $n = 10000$ and $h = 287$.

Model m_0	TRP(m)	$\zeta_u(\mathbf{v}^0 \hat{\mathbf{v}}_n)$	$\zeta_o(\hat{\mathbf{v}}_n \mathbf{v}^0)$	$\zeta_d(\hat{\mathbf{v}}_n, \mathbf{v}^0)$	Average time (s)	Model m_0	TRP(m)	$\zeta_u(\mathbf{v}^0 \hat{\mathbf{v}}_n)$	$\zeta_o(\hat{\mathbf{v}}_n \mathbf{v}^0)$	$\zeta_d(\hat{\mathbf{v}}_n, \mathbf{v}^0)$	Average time (s)		
(B1)	1	0.99	0.0007	0.0005	0.0010	69.4290	(C1)	1	1.00	0.0008	0.0008	0.0015	69.01765
(B2)	2	0.94	0.0040	0.0006	0.0010	70.5583	(C2)	2	0.93	0.0046	0.0009	0.0016	71.3397
(B3)	3	0.98	0.0012	0.0008	0.0013	69.3625	(C3)	3	0.89	0.0081	0.0013	0.0019	70.1191
(B4)	4	0.97	0.0031	0.0008	0.0014	73.3520	(C4)	4	0.84	0.0111	0.0010	0.0016	73.67172
(B5)	5	0.96	0.0020	0.0010	0.0015	75.4675	(C5)	5	0.87	0.0096	0.0014	0.0019	76.34772
(B6)	6	0.82	0.0026	0.0210	0.0169	74.3734	(C6)	6	0.92	0.0065	0.0030	0.0036	75.81408
(B7)	7	0.99	0.0149	0.0130	0.0105	78.2058	(C7)	7	0.73	0.0268	0.0033	0.0037	80.13946
(B8)	8	0.94	0.0110	0.0113	0.0087	79.4421	(C8)	8	0.94	0.0054	0.0026	0.0031	81.14015
(B9)	9	0.80	0.0041	0.0239	0.0166	81.5739	(C9)	9	0.98	0.0034	0.0026	0.0032	81.07215

Table 5: Confidence intervals based on the LRSM change-points correct estimates under the Models (B1) to (C2).

Model	τ_j^0	Median of $\hat{\tau}_j^{(3)}$	Mean of $\hat{\tau}_j^{(3)}$	Construct CIs Method	90% range	Mean of 90% CI	Coverage Prob.
Model (B1)	1000	998	996.2961	CIs' approximation	[993,1003]	[990.83, 1001.76]	74.32 %
	1000			Parametric bootstrap	[956,1006]	[952.64, 1004.14]	89.02 %
	1000			Block bootstrap($n_b = 50$)	[971,1005]	[968.61, 1002.91]	85.50 %
Model (C1)	1000	999	999.8216	CIs' approximation	[993,1006]	[993.17, 1006.48]	70.85 %
	1000			Parametric bootstrap	[959,1007]	[957.22, 1007.80]	89.50 %
	1000			Block bootstrap($n_b = 50$)	[965,1006]	[964.98, 1007.02]	86.65 %
Model (B2)	600	598	596.1302	CIs' approximation	[593,604]	[590.68, 601.58]	71.44 %
	600			Parametric bootstrap	[557,606]	[552.56, 604.24]	90.01 %
	600			Block bootstrap($n_b = 42$)	[572,605]	[570.06, 603.07]	85.07 %
	1200	1200	1203.3673	CIs' approximation	[1196,1203]	[1199.80, 1206.94]	87.08 %
	1200			Parametric bootstrap	[1193,1204]	[1196.12, 1207.71]	92.43 %
	1200			Block bootstrap($n_b = 18$)	[1195,1204]	[1198.34, 1207.50]	90.21 %
Model (C2)	600	599	598.7670	CIs' approximation	[592,606]	[591.94, 605.59]	68.45 %
	600			Parametric bootstrap	[559,607]	[556.34, 606.83]	87.63 %
	600			Block bootstrap($n_b = 42$)	[567,606]	[567.43, 606.36]	84.64 %
	1200	1200	1200.9876	CIs' approximation	[1196,1204]	[1196.65, 1205.33]	76.80 %
	1200			Parametric bootstrap	[1193,1204]	[1193.90, 1205.53]	81.03 %
	1200			Block bootstrap($n_b = 18$)	[1192,1205]	[1192.98, 1206.61]	83.20 %

parameter changes.

5 Real data

In this section, the LRSM was used to detect the change-points in the daily stock volume data of four airline groups. The aim is to investigate whether any of the detected change-points were influenced by recent sudden international events, particularly the global COVID-19 over the past few years. Four airline data sets were originally downloaded it on-line at the Yahoo Finance web site (<https://hk.finance.yahoo.com/>), including:

- Singapore Airlines Limited (C6L.SI): the series has 5,927 observations from January 4, 2000, to May 19, 2023.
- Air China Limited (0753.HK): the series has 4,550 observations from December 16, 2004, to May 19, 2023.
- Deutsche Lufthansa AG (LHA.DE): the series has 6,774 observations from December 17, 1996, to May 19, 2023.
- United Airlines Holdings, Inc. (UAL): the series has 4,338 observations from February 27, 2006, to May 19, 2023.

As the data is relatively large, considering the convenience of calculation, we analyze the data by the unit of '1e+06' trading volume. Figure 1 shows sample path figures for six data sets. It can be roughly judged from the figure that there are obvious change-points in the company's trading volume data.

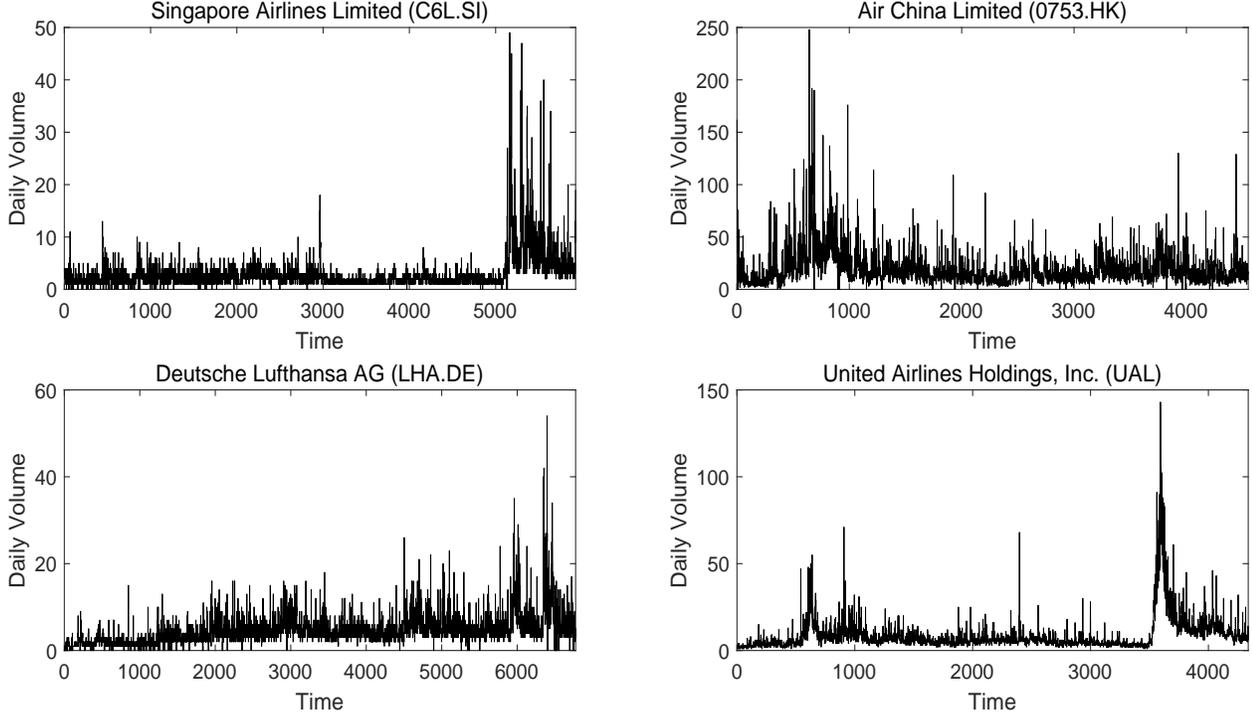


Figure 1: Sample paths for six airline group daily trading volume data sets, including C6L.SI, 0753.HK, LHA.DE, UAL.

We applied the three-step LRSM with $h = (\log n)^4/25$, $m_{max} = 30$, $p_{max} = 7$ to implement the change-points inference. Furthermore, we evaluated the goodness of fit of the model using root mean square of differences between observations and forecasts (RMS) and Pearson residuals (Pr). These evaluation metrics are commonly employed. The equations defining RMS and Pr are provided below.

$$\text{RMS} = \sqrt{\frac{1}{n-1} \sum_{t=2}^n (X_t - \mathbb{E}(X_t | \mathcal{F}_{t-1}))^2}, \quad \text{Pr}_t = \frac{X_t - \mathbb{E}(X_t | \mathcal{F}_{t-1})}{\sqrt{\text{Var}(X_t | \mathcal{F}_{t-1})}}.$$

Table 6 summarizes all change-points inference results for four airline group data sets, including the estimate of the number of change-points (\hat{m}), the estimate of change-points locations ($\hat{\tau}$) and their corresponding time (Date), the MDL value, RMS and the mean and variance of Pr_t (Pr_m and Pr_{Var}). In addition, two bootstrap methods, PBA and BBA, are applied to obtain CIs of change-points estimates, where the adaptive block bandwidth n_b in BBA is in parentheses. In order to be intuitive, the change-points of each model and the CIs constructed by two bootstrap methods are shown in the figures. However, due to space constraints, only one group (C6L.SI) is listed (Model 5.1 and Figure 3); the others are in the Appendix.

From Table 6, it can be seen that the analysis results also show that the stock volumes of four airline groups have been relatively stable in the past 20 years or so, because a maximum of six change-points are detected in the large amount of data in each group. As for the estimated model, the SE of estimated parameters in most estimated models is significant, except “0753.HK”

estimated model. For this data set, some other models, such as integer-valued autoregressive conditional heteroskedastic (INGARCH) models, should be considered to fit this data in future studies. Furthermore, it is discernible that the RMS outcome for 0753.HK is large, and its Pr_{Var} value is relatively large. This could be attributed to the presence of additional types of outliers in the segmented data of 0753.HK, such as additive outliers, thereby affecting the analytical results. The diagnostic checking plots in Figure 2 display the Pearson residuals of Model MCP-GCINAR for C6L.SI dataset. These plots indicate that most lagged values of the ACF and PACF fall within the boundaries of the blue lines, suggesting a white noise characteristic for the residuals. To further support this claim, we conduct the Ljung-Box (LB) test on the series of Pearson residuals. This test is performed with delay orders ranging from 1 to 10. Interestingly, all corresponding p -values are found to be greater than 0.05, suggesting that the series of Pearson residuals exhibits characteristics of white noise.

Next, the following international event that may have an impact on the airline groups is analyzed for the common change-points obtained.

- $\hat{\tau} = 5103$ in C6L.SI data (February 10, 2020); $\hat{\tau} = 3701$ in 0753.HK data (December 6, 2019); $\hat{\tau} = 5886$ in LHA data (November 21, 2019); $\hat{\tau} = 3480$ in UAL data (December 20, 2019). Since the outbreak of COVID-19 at the end of 2019, it has brought a great impact on various industries around the world, especially on the airline industry, which can be said to be unprecedented, which is also clearly reflected in the stock trading volume data. From the data analysis, it can be seen that before the outbreak of COVID-19, the stock trading volume of airlines had remained at a relatively stable level. However, since the beginning of COVID-19, the airline stock trading volume has almost all seen the most significant change-point since going public. To figure out why, first of all, the epidemic led many governments to implement strict travel restrictions. Most flights were canceled and passenger demand dropped significantly, which led to the decline of airline stock prices. A lack of confidence in the airline's future has led many investors to sell these shares, which is the main reason for the sharp increase in trading volume. Second, airlines are under considerable financial pressure due to the rising costs of prolonged grounding, epidemic prevention and control and flight resumption, which also affect the performance of their stocks. Finally, COVID-19's impact on the global economy has had a negative impact. As a result, airlines have been severely hit and may face the risk of bankruptcy, affecting investor confidence in the future of airlines and stock trading behavior.

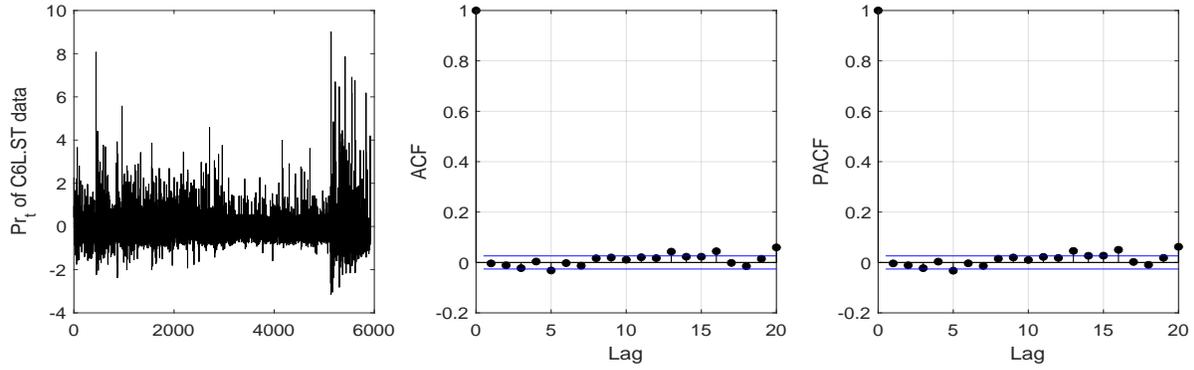


Figure 2: Diagnostic checking plots of the fitted MCP-GCINAR model: the traces, ACF and PACF plots of Pearson residual.

Table 6: Summary of the change-points estimate results implemented by the LRSM based on four airline groups data set.

Group	LRSM							CIs	
	\hat{m}	$\hat{\tau}$	Date	MDL	RMS	Pr_m	Pr_{Var}	CI-PBA	CI-BBA(n_b)
C6L.SI	1	5103	2020/2/10	-4438.7116	2.2851	0.0015	0.8932	[5082,5118]	[5074,5115]($n_b=60$)
0753.HK	6	420	2006/8/25	-180550.5027	12.2339	0.0007	1.6302	[416,425]	[407,451.5]($n_b=108$)
		994	2008/12/22					[989,1000]	[971,1036]($n_b=108$)
		1934	2012/10/8					[1914,1954]	[1843,2158]($n_b=432$)
		3175	2017/10/19					[3160,3198]	[3167,3315]($n_b=288$)
		3701	2019/12/6					[3704,3757]	[3697,3755]($n_b=328$)
		4030	2021/4/12					[3985,4027]	[3879,4015]($n_b=168$)
LHA.DE	3	1159	2001/5/25	-18682.851	6.1740	-0.0004	1.2861	[1015,1145]	[1008,1149]($n_b=204$)
		5886	2019/11/21					[5799,5933]	[5701,6066]($n_b=186$)
		6495	2022/4/20					[6500,6560]	[6501,6555]($n_b=60$)
UAL	1	3480	2019/12/20	-54833.6158	7.3444	-0.0002	1.5487	[3370,3552]	[3486,3660]($n_b=362$)

The corresponding estimated model based on 'C6L.SI' data set is as follows:

$$\mathbf{E}(X_t | \mathcal{F}_{t-1}) = \begin{cases} 0.3343X_{t-1,1} + 0.1566X_{t-2,1} + 0.0575X_{t-3,1} + 0.0486X_{t-4,1} + 0.0419X_{t-5,1} + 0.0482X_{t-6,1} + 0.6690 & 0 < t \leq 5103, \\ (0.0192) \quad (0.0217) \quad (0.0175) \quad (0.0169) \quad (0.0168) \quad (0.0157) \quad (0.0395) \\ 0.4335X_{t-5104,2} + 0.0861X_{t-5105,2} + 0.0266X_{t-5106,2} + 0.1095X_{t-5107,2} + 0.0988X_{t-5108,2} + 1.8154 & 5103 < t \leq 5927. \\ (0.0553) \quad (0.0426) \quad (0.0385) \quad (0.0399) \quad (0.0396) \quad (0.2669) \end{cases} \quad (5.1)$$

where in parentheses are the standard errors (SE) of the estimators obtained from the square roots of the elements in the diagonal of the matrix $\hat{\Sigma}_j = \hat{\mathbf{J}}_j^{-1} \hat{\mathbf{I}}_j \hat{\mathbf{J}}_j^{-1}$ computed on each segment j , with $\hat{\mathbf{J}}_j$ and $\hat{\mathbf{I}}_j$ are given by (2.2) - (2.3).

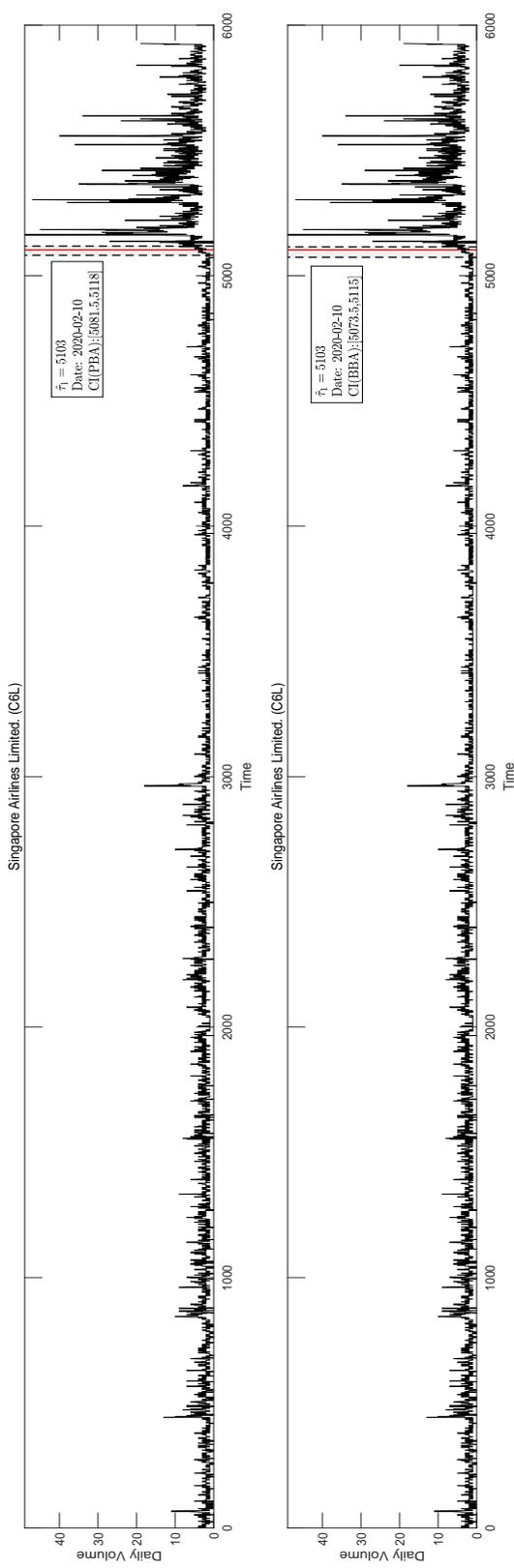


Figure 3: Sample path of the C6L.SI daily trading volume and estimate results given by the LRSM. The red line is the LRSM change-points estimate and the shaded regions are the corresponding PBA and BBA CIs.

6 Conclusion and discussion

In this paper, we propose a three-step LRSM, which provides a computationally valid and theoretically justified methods for change-points inference in the MCP-GCINAR process. We infer that the computational complexity of the LRSM is $O((\log n)^3 n)$. Simulation results and real data analysis results show that, the LRSM with usual window parameter h performs well in samples of long-time series with few and even change-points, and is robust when the assumption of window h is violated. In contrast, the LRSM with the multiple window parameter h_{mix} performs well in short-time series with large and dense change-points, at the cost of high computational cost. Furthermore, we demonstrated the asymptotic distribution of the change-points estimates. The approximation distribution and two bootstrap procedures, parametric bootstrap and block bootstrap, to approximate the finite sample distribution of change-point estimates, and hence construct the CIs for the change-points. Nevertheless, with proper modifications to the assumptions of the LRSM, it is possible to extend it to more generalized integer-value time series models, such as INGARCH models, integer-value moving average (INMA) models. We will leave the above issues as our future work.

Acknowledgements

This work is supported by National Natural Science Foundation of China (No.12271231).

Appendix

We first state the following lemmas and their proofs, which are required to prove theorems in the LRSM. Note that $\boldsymbol{\theta}$ corresponds one-to-one to the order p , to lighten notation, denote $\ell_t(\boldsymbol{\theta}) = \ell_t(\boldsymbol{\theta}, p)$ and $\xi_t(\boldsymbol{\theta}) = \xi_t(\boldsymbol{\theta}, p | \mathbf{X}_{t-1})$.

Lemma 1 *For the j th change-point τ_j^0 , the scan statistic $S_h(\tau_j^0)$*

$$\begin{aligned}
 S_h(\tau_j^0) &= \frac{1}{h} L_h(\tau_j^0 - h, \hat{\boldsymbol{\theta}}_j, p_j) + \frac{1}{h} L_h(\tau_j^0, \hat{\boldsymbol{\theta}}_{j+1}, p_{j+1}) - \frac{1}{h} L_{2h}(\tau_j^0 - h, \hat{\boldsymbol{\theta}}_{j,j+1}, p_{j,j+1}) \\
 &= \frac{1}{h} \sum_{t=\tau_j^0-h+1}^{\tau_j^0} \ell_{t, \boldsymbol{\theta}_j^0}(\hat{\boldsymbol{\theta}}_j) + \frac{1}{h} \sum_{t=\tau_j^0+1}^{\tau_j^0+h} \ell_{t, \boldsymbol{\theta}_{j+1}^0}(\hat{\boldsymbol{\theta}}_{j+1}) - \frac{1}{h} \left[\sum_{t=\tau_j^0-h+1}^{\tau_j^0} \ell_{t, \boldsymbol{\theta}_j^0}(\hat{\boldsymbol{\theta}}_{j,j+1}) + \sum_{t=\tau_j^0+1}^{\tau_j^0+h} \ell_{t, \boldsymbol{\theta}_{j+1}^0}(\hat{\boldsymbol{\theta}}_{j,j+1}) \right] \\
 &\xrightarrow{p} \sum_{i=0}^1 \mathbb{E} \left\{ \ell_{t, \boldsymbol{\theta}_{j+i}^0}(\boldsymbol{\theta}_{j+i}^0) - \ell_{t, \boldsymbol{\theta}_{j+i}^0}(\boldsymbol{\theta}_{j,j+1}) \right\} \\
 &= \sum_{i=0}^1 \mathbb{E} \left\{ X_t \log \frac{\xi_t(\boldsymbol{\theta}_{j+i}^0)}{\xi_t(\boldsymbol{\theta}_{j,j+1})} - [\xi_t(\boldsymbol{\theta}_{j+i}^0) - \xi_t(\boldsymbol{\theta}_{j,j+1})] \right\} \triangleq g_j > 0.
 \end{aligned}$$

where

$$\hat{\boldsymbol{\theta}}_{j,j+1} = \arg \max_{\boldsymbol{\theta} \in \Theta} \frac{1}{h} \left[\sum_{t=\tau_j^0-h+1}^{\tau_j^0} \ell_{t, \boldsymbol{\theta}_j^0}(\boldsymbol{\theta}) + \sum_{t=\tau_j^0+1}^{\tau_j^0+h} \ell_{t, \boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta}) \right],$$

$$\boldsymbol{\theta}_{j,j+1} = \arg \max_{\boldsymbol{\theta} \in \Theta} \left[\mathbb{E}(\ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta})) + \mathbb{E}(\ell_{t,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta})) \right].$$

and $\ell_{t,\boldsymbol{\theta}_s^0}(\boldsymbol{\theta})$ represents $\ell_t(\boldsymbol{\theta}, p)$ with the true parameter $\boldsymbol{\theta}_s^0$.

Proof of Lemma 1. We first have

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{j,j+1} &= \arg \max_{\boldsymbol{\theta} \in \Theta} \frac{1}{h} \left[\sum_{t=\tau_j^0-h+1}^{\tau_j^0} \ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta}) + \sum_{t=\tau_{j+1}^0+1}^{\tau_{j+1}^0+h} \ell_{t,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta}) \right], \\ &= \arg \max_{\boldsymbol{\theta} \in \Theta} \left[\mathbb{E}(\ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta})) + \mathbb{E}(\ell_{t,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta})) + o_p(1) \right] \xrightarrow{p} \boldsymbol{\theta}_{j,j+1}, \quad \text{as } h \rightarrow \infty. \end{aligned}$$

Combine $\hat{\boldsymbol{\theta}}_j \xrightarrow{p} \boldsymbol{\theta}_j^0$ and $\hat{\boldsymbol{\theta}}_{j+1} \xrightarrow{p} \boldsymbol{\theta}_{j+1}^0$, after some simple transformations

$$\begin{aligned} S_h(\tau_j^0) &= \frac{1}{h} L_h(\tau_j^0 - h, \hat{\boldsymbol{\theta}}_j, p_j) + \frac{1}{h} L_h(\tau_j^0, \hat{\boldsymbol{\theta}}_{j+1}, p_{j+1}) - \frac{1}{h} L_{2h}(\tau_j^0 - h, \hat{\boldsymbol{\theta}}_{j,j+1}, p_{j,j+1}) \\ &\xrightarrow{p} \sum_{i=0}^1 \mathbb{E} \left\{ \ell_{t,\boldsymbol{\theta}_{j+i}^0}(\boldsymbol{\theta}_{j+i}^0) - \ell_{t,\boldsymbol{\theta}_{j+i}^0}(\boldsymbol{\theta}_{j,j+1}) \right\} \\ &= \sum_{i=0}^1 \mathbb{E} \left\{ X_t \log \frac{\xi_t(\boldsymbol{\theta}_{j+i}^0)}{\xi_t(\boldsymbol{\theta}_{j,j+1})} - [\xi_t(\boldsymbol{\theta}_{j+i}^0) - \xi_t(\boldsymbol{\theta}_{j,j+1})] \right\} = g_j. \end{aligned} \tag{6.1}$$

Next, to show $g_j > 0$. By the definition of maximum likelihood estimate, for the first part ($i = 0$) in eq.(6.1),

$$\mathbb{E}[\ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta}_j^0) - \ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta}_{j,j+1})] \geq 0,$$

the equality sign is true if and only if $\boldsymbol{\theta}_{j,j+1} = \boldsymbol{\theta}_j^0$. Similarly, for the second part ($i = 1$),

$$\mathbb{E}[\ell_{t,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta}_{j+1}^0) - \ell_{t,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta}_{j,j+1})] \geq 0,$$

the equality sign is true if and only if $\boldsymbol{\theta}_{j,j+1} = \boldsymbol{\theta}_{j+1}^0$. Note that $\boldsymbol{\theta}_j^0 \neq \boldsymbol{\theta}_{j+1}^0$, which means that both parts cannot be equal at the same time, that is $g_j > 0$, and the proof is completed.

Lemma 2 Let $\{X_t\}$ be a piecewise stationary MCP-GCINAR process defined in (1.2), and $Y_t(\boldsymbol{\theta}) = \ell_t(\boldsymbol{\theta}) - \mathbb{E}(\ell_t(\boldsymbol{\theta}))$. For all $\boldsymbol{\theta} \in \Theta$, if $\mathbb{E}(X_t)^{2k+2\epsilon_X} < \infty$ for some $\epsilon_X > 0$, then $\sup_{\boldsymbol{\theta} \in \Theta} \mathbb{E}|\ell_t(\boldsymbol{\theta})|^{k+\epsilon_X} < \infty$.

Proof of Lemma 2. By the definition of $\ell_t(\boldsymbol{\theta}_j)$, we have

$$\begin{aligned} \mathbb{E}|\ell_t(\boldsymbol{\theta}_j)|^{k+\epsilon_X} &= \mathbb{E}|X_{t,j} \log \xi_t(\boldsymbol{\theta}_j) - \xi_t(\boldsymbol{\theta}_j)|^{k+\epsilon_X} \\ &\leq \mathbb{E}|X_t \log \xi_t(\boldsymbol{\theta}_j) + \xi_t(\boldsymbol{\theta}_j)|^{k+\epsilon_X}. \end{aligned}$$

Clearly, there exist a suitable constant $0 < C_1 < \infty$ satisfy

$$\mathbb{E}|\ell_t(\boldsymbol{\theta}_j, p_j)|^{k+\epsilon_X} \leq C_1 \mathbb{E}|X_t^2|^{k+\epsilon_X}.$$

That is, if $E(X_t)^{2k+2\epsilon_X} < \infty$, there is

$$\sup_{\boldsymbol{\theta} \in \Theta} E|\ell_t(\boldsymbol{\theta})|^{k+\epsilon_X} < \infty \text{ and } \sup_{\boldsymbol{\theta} \in \Theta} E|Y_t(\boldsymbol{\theta})|^{k+\epsilon_X} < \infty.$$

The proof of Lemma 2 is completed.

Lemma 3 *For any $\epsilon > 0$, there exists a positive integer C_2 such that for any $h > C_2$,*

$$P\left(|S_h(t)| > \epsilon\right) \leq 6 \exp\left(-\frac{1}{4}h^{1/3}\epsilon^{2/3}\right),$$

for all t such that $W_t(h)$ does not contain any change-point.

Proof of Lemma 3. Since the scanning window $W_t(h)$ has no change point, we can assume that all data in the window comes from the segment specified by $\boldsymbol{\theta}^0$. Hence, $S_h(t)$ can be written as

$$\begin{aligned} S_h(t) &= \frac{1}{h} \sum_{s=t-h+1}^t [\ell_s(\hat{\boldsymbol{\theta}}_1) - \ell_s(\boldsymbol{\theta}^0)] + \frac{1}{h} \sum_{s=t+1}^{t+h} [\ell_s(\hat{\boldsymbol{\theta}}_2) - \ell_s(\boldsymbol{\theta}^0)] - \frac{1}{h} \sum_{s=t-h+1}^{t+h} [\ell_s(\hat{\boldsymbol{\theta}}) - \ell_s(\boldsymbol{\theta}^0)] \\ &= \text{I} + \text{II} + \text{III}, \end{aligned} \tag{6.2}$$

where $\hat{\boldsymbol{\theta}}_1$, $\hat{\boldsymbol{\theta}}_2$ and $\hat{\boldsymbol{\theta}}$ and the PQML estimates of the parameter $\boldsymbol{\theta}$ in the left half, right half, and the entire scanning window, respectively. For the third part (III) in eq.(6.2),

$$\begin{aligned} \frac{1}{h} \sum_{s=t-h+1}^{t+h} [\ell_s(\hat{\boldsymbol{\theta}}) - \ell_s(\boldsymbol{\theta}^0)] &= \frac{1}{h} \sum_{s=t-h+1}^{t+h} [\ell_s(\hat{\boldsymbol{\theta}}) - E(\ell_s(\hat{\boldsymbol{\theta}}))] - \frac{1}{h} \sum_{s=t-h+1}^{t+h} [\ell_s(\boldsymbol{\theta}^0) - E(\ell_s(\boldsymbol{\theta}^0))] \\ &\quad + 2[E(\ell_s(\hat{\boldsymbol{\theta}})) - E(\ell_s(\boldsymbol{\theta}^0))]. \end{aligned}$$

To simplify notation, let $Y_s(\boldsymbol{\theta}, p) = \ell_s(\boldsymbol{\theta}) - E(\ell_s(\boldsymbol{\theta}))$. Within any segment, $\ell_s(\boldsymbol{\theta})$ is a measurable and Lipschitz continuous function with respect to $\{X_t\}$. under the assumption H.4, there exists a $C_2^{(1)}$ such that $E(e^{|Y_s(\boldsymbol{\theta})|}) \leq C_2^{(1)}$, and $Y_s(\boldsymbol{\theta})$ is obviously a finite sequence of martingale differences for $s = t-h+1, \dots, t+h$ and $\boldsymbol{\theta} \in \Theta$. Then, following the Theorem 3.2 of [Lesigne and Voln \(2011\)](#), for any $\epsilon > 0$, there exists a positive integer $C_2^{(2)}$ depending only on $C_2^{(1)}$ and ϵ , such that, for any $h > C_2^{(2)}$ and $\boldsymbol{\theta} \in \Theta$,

$$P\left(\left|\frac{1}{h} \sum_{s=t-h+1}^{t+h} [\ell_s(\boldsymbol{\theta}) - E(\ell_s(\boldsymbol{\theta}))]\right| > \frac{\epsilon}{3}\right) \leq \exp\left(-\frac{1}{4}h^{1/3}\epsilon^{2/3}\right). \tag{6.3}$$

Moreover, $\ell_s(\boldsymbol{\theta})$ is the uniform integrability for all $\boldsymbol{\theta} \in \Theta$, combine $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}^0$, we have that, for any $\epsilon > 0$, there exists a constant $C_2^{(3)} > 0$ such that, for any $h > C_2^{(3)}$,

$$2\left|E(\ell_s(\hat{\boldsymbol{\theta}})) - E(\ell_s(\boldsymbol{\theta}^0))\right| < \frac{\epsilon}{3} \tag{6.4}$$

Thus, applying (6.3) to $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^0$, there exists a constant $C_2^{(4)} > 0$ such that, for any $h > C_2^{(4)}$,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{h} \sum_{s=t-h+1}^{t+h} [\ell_s(\hat{\boldsymbol{\theta}}) - \mathbb{E}(\ell_s(\hat{\boldsymbol{\theta}}))]\right| > \frac{\epsilon}{3}\right) &\leq \exp(-\frac{1}{4}h^{1/3}\epsilon^{2/3}), \\ \mathbb{P}\left(\left|\frac{1}{h} \sum_{s=t-h+1}^{t+h} [\ell_s(\boldsymbol{\theta}^0) - \mathbb{E}(\ell_s(\boldsymbol{\theta}^0))]\right| > \frac{\epsilon}{3}\right) &\leq \exp(-\frac{1}{4}h^{1/3}\epsilon^{2/3}). \end{aligned}$$

Then, together with (6.4), we obtain that, for any $\epsilon > 0$, there exists a constant $C_2^{(5)} = \max\{C_2^{(3)}, C_2^{(4)}\}$ such that, for any $h > C_2^{(5)}$,

$$\mathbb{P}\left(\left|\frac{1}{h} \sum_{s=t-h+1}^{t+h} [\ell_s(\hat{\boldsymbol{\theta}}) - \ell_s(\boldsymbol{\theta}^0)]\right| > \epsilon\right) \leq 2 \exp(-\frac{1}{4}h^{1/3}\epsilon^{2/3}). \quad (6.5)$$

Similarly exponential inequalities hold for the other two parts (I) and (II) in eq.(6.2), and the proof of Lemma 3 is completed.

Lemma 4 *For any $\epsilon > 0$, there exists a positive integer C_3 such that for any $h > C_3$,*

$$\mathbb{P}\left(|S_h(\tau_j^0) - g_j| > \epsilon\right) \leq 22 \exp(-\frac{1}{4}h^{1/3}\epsilon^{2/3}),$$

for all $j = 1, \dots, m_0$.

Proof of Lemma 4. Using the notations in Lemma 1, there is

$$\begin{aligned} &S_h(\tau_j^0) - g_j \\ &= \left[\frac{1}{h} \sum_{t=\tau_j^0-h+1}^{\tau_j^0} \ell_{t,\boldsymbol{\theta}_j^0}(\hat{\boldsymbol{\theta}}_j) - \mathbb{E}\ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta}_j^0)\right] + \left[\frac{1}{h} \sum_{t=\tau_j^0+1}^{\tau_j^0+h} \ell_{t,\boldsymbol{\theta}_{j+1}^0}(\hat{\boldsymbol{\theta}}_{j+1}) - \mathbb{E}\ell_{t,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta}_{j+1}^0)\right] \\ &\quad - \left[\frac{1}{h} \sum_{t=\tau_j^0-h+1}^{\tau_j^0} \left(\ell_{t,\boldsymbol{\theta}_j^0}(\hat{\boldsymbol{\theta}}_{j,j+1}) + \ell_{t+h,\boldsymbol{\theta}_{j+1}^0}(\hat{\boldsymbol{\theta}}_{j,j+1})\right) - \mathbb{E}\left(\ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta}_{j,j+1}) + \ell_{t,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta}_{j,j+1})\right)\right] \\ &= \left[\frac{1}{h} \sum_{t=\tau_j^0-h+1}^{\tau_j^0} \left(\ell_{t,\boldsymbol{\theta}_j^0}(\hat{\boldsymbol{\theta}}_j) - \ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta}_j^0)\right)\right] + \left[\frac{1}{h} \sum_{t=\tau_j^0-h+1}^{\tau_j^0} \ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta}_j^0) - \mathbb{E}\ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta}_j^0)\right] \\ &\quad + \left[\frac{1}{h} \sum_{t=\tau_j^0+1}^{\tau_j^0+h} \left(\ell_{t,\boldsymbol{\theta}_{j+1}^0}(\hat{\boldsymbol{\theta}}_{j+1}) - \ell_{t,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta}_{j+1}^0)\right)\right] + \left[\frac{1}{h} \sum_{t=\tau_j^0+1}^{\tau_j^0+h} \ell_{t,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta}_{j+1}^0) - \mathbb{E}\ell_{t,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta}_{j+1}^0)\right] \\ &\quad + \frac{1}{h} \sum_{t=\tau_j^0-h+1}^{\tau_j^0} \left(\left(\ell_{t,\boldsymbol{\theta}_j^0}(\hat{\boldsymbol{\theta}}_{j,j+1}) + \ell_{t+h,\boldsymbol{\theta}_{j+1}^0}(\hat{\boldsymbol{\theta}}_{j,j+1})\right) - \left(\ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta}_{j,j+1}) + \ell_{t+h,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta}_{j,j+1})\right)\right) \\ &\quad + \left[\frac{1}{h} \sum_{t=\tau_j^0-h+1}^{\tau_j^0} \ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta}_{j,j+1}) - \mathbb{E}\ell_{t,\boldsymbol{\theta}_j^0}(\boldsymbol{\theta}_{j,j+1})\right] + \left[\frac{1}{h} \sum_{t=\tau_j^0+1}^{\tau_j^0+h} \ell_{t,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta}_{j,j+1}) - \mathbb{E}\ell_{t,\boldsymbol{\theta}_{j+1}^0}(\boldsymbol{\theta}_{j,j+1})\right] \end{aligned}$$

=I + II + III + IV

For the parts I and II, according to equations (6.3) and (6.5) in Lemma 3, we have that for any $\epsilon > 0$, there exists a constant $C_3^{(1)}$ such that, for any $h > C_3^{(1)}$,

$$\begin{aligned} & \mathbb{P}\left(\left|\frac{1}{h} \sum_{t=\tau_j^0-h+1}^{\tau_j^0} [\ell_{t,\theta_j^0}(\hat{\boldsymbol{\theta}}_j) - \ell_{t,\theta_j^0}(\boldsymbol{\theta}_j^0)]\right| > \epsilon\right) + \mathbb{P}\left(\left|\frac{1}{h} \sum_{t=\tau_j^0-h+1}^{\tau_j^0} \ell_{t,\theta_j^0}(\boldsymbol{\theta}_j^0) - \mathbb{E}\ell_{t,\theta_j^0}(\boldsymbol{\theta}_j^0)\right| > \epsilon\right) \\ & \leq 5 \exp\left(-\frac{1}{4}h^{1/3}\epsilon^{2/3}\right). \end{aligned}$$

and

$$\begin{aligned} & \mathbb{P}\left(\left|\frac{1}{h} \sum_{t=\tau_j^0+1}^{\tau_j^0+h} [\ell_{t,\theta_{j+1}^0}(\hat{\boldsymbol{\theta}}_{j+1}) - \ell_{t,\theta_{j+1}^0}(\boldsymbol{\theta}_{j+1}^0)]\right| > \epsilon\right) + \mathbb{P}\left(\left|\frac{1}{h} \sum_{t=\tau_j^0+1}^{\tau_j^0+h} \ell_{t,\theta_{j+1}^0}(\boldsymbol{\theta}_{j+1}^0) - \mathbb{E}\ell_{t,\theta_{j+1}^0}(\boldsymbol{\theta}_{j+1}^0)\right| > \epsilon\right) \\ & \leq 5 \exp\left(-\frac{1}{4}h^{1/3}\epsilon^{2/3}\right). \end{aligned} \tag{6.6}$$

By the definition of $\boldsymbol{\theta}_{j,j+1}$, we have

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathbb{E}(\ell_{t,\theta_j^0}(\boldsymbol{\theta}) + \ell_{t+h,\theta_{j+1}^0}(\boldsymbol{\theta})) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{j,j+1}} = 0.$$

Obviously, $\ell_{t,\boldsymbol{\theta}}(\boldsymbol{\theta})$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$ almost surely, thus

$$\mathbb{E} \frac{\partial}{\partial \boldsymbol{\theta}} (\ell_{t,\theta_j^0}(\boldsymbol{\theta}_{j,j+1}) + \ell_{t+h,\theta_{j+1}^0}(\boldsymbol{\theta}_{j,j+1})) = 0.$$

That is, $\left\{ \frac{\partial}{\partial \boldsymbol{\theta}} (\ell_{t,\theta_j^0}(\boldsymbol{\theta}_{j,j+1}) + \ell_{t+h,\theta_{j+1}^0}(\boldsymbol{\theta}_{j,j+1})) \right\}$ is a martingale difference sequence. Similarly to the proof of Lemma 3, we have that for any $\epsilon > 0$, there exists a constant $C_3^{(2)}$ such that, for any $h > C_3^{(2)}$,

$$\begin{aligned} & \mathbb{P}\left(\left|\frac{1}{h} \sum_{t=\tau_j^0-h+1}^{\tau_j^0} [(\ell_{t,\theta_j^0}(\hat{\boldsymbol{\theta}}_{j,j+1}) + \ell_{t+h,\theta_{j+1}^0}(\hat{\boldsymbol{\theta}}_{j,j+1})) - (\ell_{t,\theta_j^0}(\boldsymbol{\theta}_{j,j+1}) + \ell_{t+h,\theta_{j+1}^0}(\boldsymbol{\theta}_{j,j+1}))]\right| > \epsilon\right) \\ & \leq 6 \exp\left(-\frac{1}{4}h^{1/3}\epsilon^{2/3}\right). \end{aligned} \tag{6.7}$$

Also, applying (6.3) to $\boldsymbol{\theta}_{j,j+1}$, for any $\epsilon > 0$, there exists a constant $C_3^{(3)}$ such that, for any $h > C_3^{(3)}$,

$$\mathbb{P}\left(\left|\frac{1}{h} \sum_{t=\tau_j^0-h+1}^{\tau_j^0} \ell_{t,\theta_j^0}(\boldsymbol{\theta}_{j,j+1}) - \mathbb{E}\ell_{t,\theta_j^0}(\boldsymbol{\theta}_{j,j+1})\right| > \epsilon\right) \leq 3 \exp\left(-\frac{1}{4}h^{1/3}\epsilon^{2/3}\right), \tag{6.8}$$

and

$$\mathbb{P}\left(\left|\frac{1}{h} \sum_{t=\tau_j^0+1}^{\tau_j^0+h} \ell_{t,\theta_{j+1}^0}(\boldsymbol{\theta}_{j,j+1}) - \mathbb{E}\ell_{t,\theta_{j+1}^0}(\boldsymbol{\theta}_{j,j+1})\right| > \epsilon\right) \leq 3 \exp\left(-\frac{1}{4}h^{1/3}\epsilon^{2/3}\right). \tag{6.9}$$

Combining formulas (6.6) - (6.9), and the proof of Lemma 4 is completed.

The proof of Theorem 1 is similar to the proof of Theorem 1 in Yau and Zhao (2016).

Proof of Theorem 1. Let $A_t = \{\text{some point in the } t\text{-th local-window is a local change-point estimate}\}$ and $A = \bigcap_{t \in \mathcal{J}_0} A_t$. If we can prove that $P(A) \rightarrow 1$ as $n \rightarrow \infty$, then the proof of Theorem 1 holds. Let $\mathbb{Z}_n = \{1, 2, \dots, n\}$. Define $\mathcal{E} = \mathbb{Z}_n \setminus \bigcap_{t \in \mathcal{J}_0} W_t(h)$ as the set of all points outside the h -neighbourhood of the true change-points. A sufficient condition for the event A to occur is that

$$\min_{t \in \mathcal{J}_0} S_h(t) > \max_{t \in \mathcal{E}} S_h(t). \quad (6.10)$$

Let $g = \frac{1}{2} \min_{j=1, \dots, m_0} (g_j)$ where g_j s are defined in Lemma 1. Note that from (6.10), there is

$$P(A) \geq P\left(\min_{t \in \mathcal{J}_0} S_h(t) > g > \max_{t \in \mathcal{E}} S_h(t)\right).$$

Hence, the proof of Theorem 1 is completed by proving the following two facts,

(i) $P\left(\min_{t \in \mathcal{J}_0} S_h(t) > g\right) \rightarrow 1,$

(ii) $P\left(g > \max_{t \in \mathcal{E}} S_h(t)\right) \rightarrow 1.$

For (i),

$$P\left(\min_{t \in \mathcal{J}_0} S_h(t) > g\right) = 1 - P\left(\bigcup_{t \in \mathcal{J}_0} \{S_h(t) \leq g\}\right) \geq 1 - \sum_{t \in \mathcal{J}_0} P(S_h(t) \leq g),$$

from Lemma 4 and the definition of g , it can be shown that $P(S_h(t) \leq g) \leq 22 \exp(-\frac{1}{4}h^{1/3}\epsilon^{2/3})$, for all $t \in \mathcal{J}_0$. Thus, set $h = d(\log n)^3$, for some $d > 0$ and $m_0 = O(1)$, we have

$$P\left(\min_{t \in \mathcal{J}_0} S_h(t) > g\right) \geq 1 - 22m_0 \exp(-\frac{1}{4}h^{1/3}\epsilon^{2/3}) \rightarrow 1.$$

For (ii), note that when $t \in \mathcal{E}$, all observations in $S_h(t)$ belong to one stationary piece. From Lemma 3, $P(S_h(t) \geq g) < 6 \exp(-\frac{1}{4}h^{1/3}\epsilon^{2/3})$ for all $t \in \mathcal{E}$. Thus, set $h = d(\log n)^3$, for some $d > 64/g^2$, we have

$$\begin{aligned} P\left(g > \max_{t \in \mathcal{E}} S_h(t)\right) &= 1 - P\left(\bigcup_{t \in \mathcal{E}} \{S_h(t_j) \geq g\}\right) \geq 1 - \sum_{j=1}^{m_0+1} (\tau_j^0 - \tau_{j-1}^0) P(S_h(t) \geq g), \\ &> 1 - 6 \exp(-\frac{1}{4}h^{1/3}\epsilon^{2/3}) \rightarrow 1 \end{aligned}$$

where $t_j \in (\tau_{j-1}^0 + h, \tau_j^0 - h)$. To sum up, when $h = d(\log n)^3$ for some $d > 64/g^2$, there is $P(A) \rightarrow 1$, and the proof of Theorem 1 is completed.

Proof of Theorem 2. In Sheng and Wang (2023), it shows that, if m_0 is known, under the assumption $E|X_{t,j}|^{2+\epsilon_X} < \infty$, estimates based on MDL principle are strongly consistent. If m_0 is unknown, under the assumption $E|X_{t,j}|^{4+\epsilon_X} < \infty$, estimates are weakly consistent, further, under the assumption $E|X_{t,j}|^{8+\epsilon_X} < \infty$, change-points estimates are strongly consistent.

Proof of Theorem 3. Let $G(\mathbf{X}_t) = |X_t \log \xi_t(\boldsymbol{\theta}_j) + \xi_t(\boldsymbol{\theta}_j)|$. Clearly, $G(\mathbf{X}_t)$ is an integrable function and $|\ell_t(\boldsymbol{\theta}_j)| \leq G(\mathbf{X}_t)$ for all $\boldsymbol{\theta}_j \in \Theta_j$. Furthermore, from Lemma 2, $E(G(\mathbf{X}_t)) < \infty$. By using the uniform law of large number in Jennrich (1969), we have as $h \rightarrow \infty$, $\frac{1}{h} \sum_{t=1}^h \ell_t(\boldsymbol{\theta}_j)$ converges uniformly to $E_{\theta_j}(\ell_t(\boldsymbol{\theta}_j))$, for any $\boldsymbol{\theta}_j \in \Theta_j$. The remaining part of the proof is the same as that of Theorem 3 in Yau and Zhao (2016) and we omit it.

Proof of Theorem 4. The proof of Theorem 4 is similar to the Theorem 3 of Cui et al. (2021), and we omit it.

Lemma 5 *Let*

$$\begin{aligned} g_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \mathbf{X}_t) &= \text{sgn}(t) (\ell_t(\boldsymbol{\theta}_1, \mathbf{X}_t) - \ell_t(\boldsymbol{\theta}_2, \mathbf{X}_t)), \\ &= \text{sgn}(t) \left\{ \left[X_t \log \frac{\sum_{i=1}^p \beta_{i,1} X_{t-i} + \beta_{0,1}}{\sum_{i=1}^p \beta_{i,2} X_{t-i} + \beta_{0,2}} - \left(\sum_{i=1}^p (\beta_{i,1} - \beta_{i,2}) X_{t-i} + \beta_{0,1} - \beta_{0,2} \right) \right] \right\} \end{aligned} \quad (6.11)$$

where $\boldsymbol{\theta}_1 = (\beta_{0,1}, \dots, \beta_{p,1})$ and $\boldsymbol{\theta}_2 = (\beta_{0,2}, \dots, \beta_{p,2})$ are the interior points of the compact space $\Theta(p) = [\delta, \tilde{\delta}] \times [0, 1 - \delta]^p \cap \mathbb{M}$, where $\mathbb{M} = \{0 \leq \sum_{k=1}^p \beta_k \leq 1 - \delta < 1\}$, δ and $\tilde{\delta}$ are finite positive constants with δ approaching 0 and $\tilde{\delta} < +\infty$, and $\text{sgn}(t) = 1$ with $t > 0$, $\text{sgn}(t) = 0$ with $t = 0$, $\text{sgn}(t) = -1$ with $t < 0$. Then the function $g_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \mathbf{X}_t)$ about \mathbf{X}_t has partial derivatives, which satisfy a Lipschitz condition, that is, the derivative

$$\frac{\partial^{p+1} g_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \mathbf{X}_t)}{\partial X_t \partial X_{t-1} \dots \partial X_{t-p}}$$

is Lipschitz.

Proof of Lemma 5. For any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta(p)$, $\mathbf{X}_t \in [0, +\infty)^{p+1}$, there exist a constant $0 < C_5^{(1)} < \infty$, such that

$$\begin{aligned} \left| \frac{\partial^{p+1} g_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \mathbf{X}_t)}{\partial X_t \partial X_{t-1} \dots \partial X_{t-p}} \right| &= \left| (-1)^p \left(\frac{\prod_{i=1}^p \beta_{i,1}}{\sum_{i=1}^p \beta_{i,1} X_{t-i} + \beta_{0,1}} - \frac{\prod_{i=1}^p \beta_{i,2}}{\sum_{i=1}^p \beta_{i,2} X_{t-i} + \beta_{0,2}} \right) \right| \\ &< \beta_{0,1}^{-1} + \beta_{0,2}^{-1} < C_5^{(1)}, \end{aligned}$$

and the Lemma 5 is clearly true.

Proof of Theorem 5. Without loss of generality, we consider the segment before and after the j th change-point, and the other change-points are similarly. Denote $\mathcal{L}(\mathbf{X})$, $\tilde{\mathcal{L}}(\mathbf{X})$, and $\mathcal{L}^*(\mathbf{X})$ as the distribution of random variable \mathbf{X} under probability measures P , \tilde{P} , and P^* , respectively. We first prove the assertion (2.11) in Theorem 5, that is, conditional on the sample $\{X_t\}_{t=\hat{\tau}_{j-1}^{(3)}+1}^{\hat{\tau}_{j+1}^{(3)}}$, for any finite $n_p \in \mathbb{N}$, there is

$$\tilde{\mathcal{L}}(\mathbf{W}_j) \triangleq \tilde{\mathcal{L}}(\tilde{W}_{j,-n_p}, \tilde{W}_{j,-n_p+1}, \dots, \tilde{W}_{j,n_p})$$

$$\xrightarrow{d} \mathcal{L}(W_{j,-n_p}, W_{j,-n_p+1}, \dots, W_{j,n_p}) \triangleq \mathcal{L}(\mathbf{W}_j), \quad \text{in probability.} \quad (6.12)$$

where $W_{j,s}, \widetilde{W}_{j,s}$ with $s = -n_p, \dots, n_p$ are defined by (2.6) and (2.9). It suffices to show that,

$$\begin{aligned} & \widetilde{\mathcal{L}} \left(\left(\widetilde{\ell}_{-n_p}(\widehat{\boldsymbol{\theta}}_{j+1}) - \widetilde{\ell}_{-n_p}(\widehat{\boldsymbol{\theta}}_j) \right), \dots, \left(\widetilde{\ell}_{n_p}(\widehat{\boldsymbol{\theta}}_j) - \widetilde{\ell}_{n_p}(\widehat{\boldsymbol{\theta}}_{j+1}) \right) \right) \quad \text{based on sample set } \{\widetilde{X}_t\}_{t=1}^{2n_p+1} \\ & \xrightarrow{d} \mathcal{L} \left(\left(\ell_{-n_p}(\boldsymbol{\theta}_{j+1}^0) - \ell_{-n_p}(\boldsymbol{\theta}_j^0) \right), \dots, \left(\ell_{n_p}(\boldsymbol{\theta}_j^0) - \ell_{n_p}(\boldsymbol{\theta}_{j+1}^0) \right) \right) \quad \text{based on sample set } \{X_t\}_{t=\widehat{\tau}_{j-1}^{(3)}+1}^{\widehat{\tau}_{j+1}^{(3)}} \\ & \quad \text{in probability.} \end{aligned} \quad (6.13)$$

Clearly, $g_t(\widehat{\boldsymbol{\theta}}_j, \widehat{\boldsymbol{\theta}}_{j+1}, \widetilde{\mathbf{X}}_t)$ defined in Lemma 5 is the extend function of $(\widetilde{\ell}_t(\widehat{\boldsymbol{\theta}}_j) - \widetilde{\ell}_t(\widehat{\boldsymbol{\theta}}_{j+1}))$,

$$g_t(\widehat{\boldsymbol{\theta}}_j, \widehat{\boldsymbol{\theta}}_{j+1}, \widetilde{\mathbf{X}}_t) = \text{sgn}(t) \left(\widetilde{\ell}_t(\widehat{\boldsymbol{\theta}}_j) - \widetilde{\ell}_t(\widehat{\boldsymbol{\theta}}_{j+1}) \right) \quad \text{for all } \widetilde{\mathbf{X}}_t \in \mathbb{N}^{p+1},$$

and $g_t(\boldsymbol{\theta}_j, \boldsymbol{\theta}_{j+1}, \mathbf{X}_t)$ fulfills the smoothness condition in Assumption 1 in Jentsch and Weiß (2019). Then combining the Assumption H.6 and following the Corollary 3.4 in Jentsch and Weiß (2019), there is

$$\begin{aligned} & \widetilde{\mathcal{L}} \left(\left(\widetilde{\ell}_{-n_p}(\widehat{\boldsymbol{\theta}}_{j+1}) - \widetilde{\ell}_{-n_p}(\widehat{\boldsymbol{\theta}}_j) \right), \dots, \left(\widetilde{\ell}_{n_p}(\widehat{\boldsymbol{\theta}}_j) - \widetilde{\ell}_{n_p}(\widehat{\boldsymbol{\theta}}_{j+1}) \right) \right) \quad \text{based on sample set } \{\widetilde{X}_t\}_{t=1}^{2n_p+1} \\ & \xrightarrow{d} \mathcal{L} \left(\left(\ell_{-n_p}(\widehat{\boldsymbol{\theta}}_{j+1}) - \ell_{-n_p}(\widehat{\boldsymbol{\theta}}_j) \right), \dots, \left(\ell_{n_p}(\widehat{\boldsymbol{\theta}}_j) - \ell_{n_p}(\widehat{\boldsymbol{\theta}}_{j+1}) \right) \right) \quad \text{based on sample set } \{X_t\}_{t=\widehat{\tau}_{j-1}^{(3)}+1}^{\widehat{\tau}_{j+1}^{(3)}} \\ & \quad \text{in probability.} \end{aligned} \quad (6.14)$$

Furthermore, since $\ell_t(\boldsymbol{\theta})$ is continuous with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_j^0$ and $\boldsymbol{\theta}_{j+1}^0$, and $\widehat{\boldsymbol{\theta}}_j \xrightarrow{P} \boldsymbol{\theta}_j^0$ and $\widehat{\boldsymbol{\theta}}_{j+1} \xrightarrow{P} \boldsymbol{\theta}_{j+1}^0$, we have

$$\begin{aligned} & \mathcal{L} \left(\left(\ell_{-n_p}(\widehat{\boldsymbol{\theta}}_{j+1}) - \ell_{-n_p}(\widehat{\boldsymbol{\theta}}_j) \right), \dots, \left(\ell_{n_p}(\widehat{\boldsymbol{\theta}}_j) - \ell_{n_p}(\widehat{\boldsymbol{\theta}}_{j+1}) \right) \right) \quad \text{based on sample set } \{X_t\}_{t=\widehat{\tau}_{j-1}^{(3)}+1}^{\widehat{\tau}_{j+1}^{(3)}} \\ & \xrightarrow{P} \mathcal{L} \left(\left(\ell_{-n_p}(\boldsymbol{\theta}_{j+1}^0) - \ell_{-n_p}(\boldsymbol{\theta}_j^0) \right), \dots, \left(\ell_{n_p}(\boldsymbol{\theta}_j^0) - \ell_{n_p}(\boldsymbol{\theta}_{j+1}^0) \right) \right) \quad \text{based on sample set } \{X_t\}_{t=\widehat{\tau}_{j-1}^{(3)}+1}^{\widehat{\tau}_{j+1}^{(3)}} \\ & \quad \text{in probability.} \end{aligned} \quad (6.15)$$

Combining (6.14) and (6.15) yields (6.13), which implies (6.12). Then, by the arg max continuous mapping theorem, for any $n_p \in \mathbb{N}$, there is

$$\widetilde{\tau}_{j,n_p} = \arg \max_{\tau \in \{-n_p, \dots, n_p\}} \widetilde{W}_{j,\tau} \xrightarrow{d} \arg \max_{\tau} W_{j,\tau} \quad \text{in probability,}$$

and the proof of the part (2.11) in Theorem 5 is completed.

Next, we prove the assertion (2.12) in Theorem 5. Note that the sample set $\{X_t^*\}_{t=1}^{2n_b+1}$ is a concatenation of subsegments $\{X_t\}_{t=\widehat{\tau}_{j-1}^{(3)}+1}^{\widehat{\tau}_j^{(3)}}$ and $\{X_t\}_{t=\widehat{\tau}_j^{(3)}+1}^{\widehat{\tau}_{j+1}^{(3)}}$, which belong to two stationary

process segments in MCP-GCINAR model. Furthermore, since $\{X_t^*\}_{t=1}^{n_b+1}$ and $\{X_t^*\}_{t=n_b+2}^{2n_b+1}$ are independently resampled conditional on the given $\{X_t\}_{t=\hat{\tau}_{j-1}^{(3)}+1}^{\hat{\tau}_{j+1}^{(3)}}$, that is, $\{X_t^*\}_{t=1}^{n_b+1}$ and $\{X_t^*\}_{t=n_b+2}^{2n_b+1}$ are independent. Therefore, there is

$$\mathcal{L}^*(X_1^*, \dots, X_{2n_b+1}^*) \xrightarrow{d} \mathcal{L}(X_{\hat{\tau}_{j-1}^{(3)}+1}, \dots, X_{\hat{\tau}_{j+1}^{(3)}}) \text{ in probability.}$$

Furthermore, similar to the proof of the assertion (2.11), we have

$$\begin{aligned} & \mathcal{L}\left(\left(\ell_{-n_b}^*(\hat{\boldsymbol{\theta}}_{j+1}) - \ell_{-n_b}^*(\hat{\boldsymbol{\theta}}_j)\right), \dots, \left(\ell_{n_b}^*(\hat{\boldsymbol{\theta}}_j) - \ell_{n_b}^*(\hat{\boldsymbol{\theta}}_{j+1})\right)\right) \text{ based on sample set } \{X_t^*\}_{t=1}^{2n_b+1} \\ & \xrightarrow{d} \mathcal{L}\left(\left(\ell_{-n_b}(\boldsymbol{\theta}_{j+1}^0) - \ell_{-n_b}(\boldsymbol{\theta}_j^0)\right), \dots, \left(\ell_{n_b}(\boldsymbol{\theta}_j^0) - \ell_{n_b}(\boldsymbol{\theta}_{j+1}^0)\right)\right) \text{ based on sample set } \{X_t\}_{t=\hat{\tau}_{j-1}^{(3)}+1}^{\hat{\tau}_{j+1}^{(3)}} \\ & \text{in probability.} \end{aligned}$$

By the arg max continuous mapping theorem, for any $n_b \in \mathbb{N}$ and $n_b < \min(\hat{\tau}_{j+1}^{(3)} - \hat{\tau}_j^{(3)}, \hat{\tau}_j^{(3)} - \hat{\tau}_{j-1}^{(3)})$, there is

$$\tau_{j,n_b}^* = \arg \max_{\tau \in \{-n_b, \dots, n_b\}} W_{j,\tau}^* \xrightarrow{d} \arg \max_{\tau} W_{j,\tau} \text{ in probability,}$$

and the proof of Theorem 5 is completed.

Optimal Partitioning Algorithm

We first define the notations used in the OP Algorithm.

Denote the cost function for the j th segment by

$$\mathcal{C}(X_{(\tau_{j-1}+1):\tau_j}) = -L_{n_j}(\tau_{j-1}, \boldsymbol{\theta}_j, p_j; X_{(\tau_{j-1}+1):\tau_j}) + \log(p_j) + \frac{p_j + 1}{2} \log(n_j).$$

where order p_j can be selected through AIC.

Optimal Partitioning Algorithm:

Input: The data set $\{X_t\}_{t=1}^n$, the penalty constant $c = \log(n)$.

The potential change-points set $\hat{\mathcal{J}}^{(1)}$ and $\hat{m}^{(1)}$ from ‘‘First step’’.

Initialise: $F(1) = -c$, $\tau^* = [0, \hat{\mathcal{J}}^{(1)}, n]$, $cp(1) = \text{Null}$.

Iterate: for $s_1 = 1, \dots, \hat{m}^{(1)}$

$$s'_1 = \arg \min_{s_2=1, \dots, s_1} [F(s_2) + \mathcal{C}(X_{(\tau_{s_2}+1):\tau_{s_2+1}}) + c].$$

$$F(s_1 + 1) = \min_{s_2=1, \dots, s_1} [F(s_2) + \mathcal{C}(X_{(\tau_{s_2}+1):\tau_{s_2+1}}) + c].$$

$$\tau' = \tau^*(s'_1).$$

$$cp(s_1 + 1) = \{cp(s'_1), \tau'\}.$$

end

Output: the change-points estimates set $\hat{\mathcal{J}}^{(2)} = cp(\hat{m}^{(1)} + 2)$.

Models (B1) - (B9) and (C1) - (C9)

Model (B):

$$X_t = \begin{cases} 0.5 \circ X_{t-1,1} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(0.5) & 0 < t \leq \tau_1^0, \\ 0.249 \circ X_{t-\tau_1^0-1,2} + 0.254 \circ X_{t-\tau_1^0-2,2} + 0.297 \circ X_{t-\tau_1^0-3,2} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(1) & \tau_1^0 < t \leq \tau_2^0, \\ 0.4 \circ X_{t-\tau_2^0-1,3} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(0.5) & \tau_2^0 < t \leq \tau_3^0, \\ 0.014 \circ X_{t-\tau_3^0-1,4} + 0.041 \circ X_{t-\tau_3^0-2,4} + 0.29 \circ X_{t-\tau_3^0-3,4} + 0.454 \circ X_{t-\tau_3^0-4,4} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(2) & \tau_3^0 < t \leq \tau_4^0, \\ 0.332 \circ X_{t-\tau_4^0-1,5} + 0.268 \circ X_{t-\tau_4^0-2,5} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(0.5) & \tau_4^0 < t \leq \tau_5^0, \\ 0.2 \circ X_{t-\tau_5^0-1,6} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(4) & \tau_5^0 < t \leq \tau_6^0, \\ 0.109 \circ X_{t-\tau_6^0-1,7} + 0.306 \circ X_{t-\tau_6^0-2,7} + 0.305 \circ X_{t-\tau_6^0-3,7} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(3) & \tau_6^0 < t \leq \tau_7^0, \\ 0.3 \circ X_{t-\tau_7^0-1,8} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(0.5) & \tau_7^0 < t \leq \tau_8^0, \\ 0.202 \circ X_{t-\tau_8^0-1,9} + 0.127 \circ X_{t-\tau_8^0-2,9} + 0.179 \circ X_{t-\tau_8^0-3,9} + 0.392 \circ X_{t-\tau_8^0-4,9} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(1) & \tau_8^0 < t \leq \tau_9^0, \\ 0.3 \circ X_{t-\tau_9^0-1,10} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(2) & \tau_9^0 < t \leq n. \end{cases}$$

Model (B1) consists of the first two segments of Model (B) and $\tau_1^0 = [0.5n]$. That is

$$X_t = \begin{cases} 0.5 \circ X_{t-1,1} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(0.5) & 0 < t \leq [0.5n] \\ 0.249 \circ X_{t-\tau_1^0-1,2} + 0.254 \circ X_{t-\tau_1^0-2,2} + 0.297 \circ X_{t-\tau_1^0-3,2} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Poi}(1) & [0.5n] < t \leq n \end{cases}$$

Similarly, we design (B2)-(B9) models as follows.

Model (B2) consists of the first three segments of Model (B) and $(\tau_1^0, \tau_2^0) = ([0.3n], [0.6n])$.

Model (B3) consists of the first four segments of Model (B) and $(\tau_1^0, \tau_2^0, \tau_3^0) = ([0.2n], [0.5n], [0.8n])$.

Model (B4) consists of the first five segments of Model (B) and

$$(\tau_1^0, \tau_2^0, \tau_3^0, \tau_4^0) = ([0.2n], [0.4n], [0.6n], [0.8n]).$$

Model (B5) consists of the first six segments of Model (B) and

$$(\tau_1^0, \tau_2^0, \tau_3^0, \tau_4^0, \tau_5^0) = ([0.1n], [0.3n], [0.6n], [0.7n], [0.9n]).$$

Model (B6) consists of the first seven segments of Model (B) and

$$(\tau_1^0, \tau_2^0, \tau_3^0, \tau_4^0, \tau_5^0, \tau_6^0) = ([0.1n], [0.2n], [0.3n], [0.5n], [0.8n], [0.9n]).$$

Model (B7) consists of the first eight segments of Model (B) and

$$(\tau_1^0, \tau_2^0, \tau_3^0, \tau_4^0, \tau_5^0, \tau_6^0, \tau_7^0) = ([0.1n], [0.2n], [0.3n], [0.4n], [0.5n], [0.8n], [0.9n]).$$

Model (B8) consists of the first nine segments of Model (B) and

$$(\tau_1^0, \tau_2^0, \tau_3^0, \tau_4^0, \tau_5^0, \tau_6^0, \tau_7^0, \tau_8^0) = ([0.1n], [0.2n], [0.3n], [0.4n], [0.5n], [0.7n], [0.8n], [0.9n]).$$

Model (B9) consists of the first ten segments of Model (B) and

$$(\tau_1^0, \tau_2^0, \tau_3^0, \tau_4^0, \tau_5^0, \tau_6^0, \tau_7^0, \tau_8^0, \tau_9^0) = ([0.1n], [0.2n], [0.3n], [0.4n], [0.5n], [0.6n], [0.7n], [0.8n], [0.9n]).$$

Similarly, we set (C1) - (C9) models based on the Model (C).

Model (C):

$$X_t = \begin{cases} 0.5 * X_{t-1,1} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Geo}(1/3) & 0 < t \leq \tau_1^0 \\ 0.249 * X_{t-\tau_1^0-1,2} + 0.254 * X_{t-\tau_1^0-2,2} + 0.297 * X_{t-\tau_1^0-3,2} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Geo}(1/2) & \tau_1^0 < t \leq \tau_2^0 \\ 0.4 * X_{t-\tau_2^0-1,3} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Geo}(1/3) & \tau_2^0 < t \leq \tau_3^0 \\ 0.014 * X_{t-\tau_3^0-1,4} + 0.041 * X_{t-\tau_3^0-2,4} + 0.29 * X_{t-\tau_3^0-3,4} + 0.454 * X_{t-\tau_3^0-4,4} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Geo}(2/3) & \tau_3^0 < t \leq \tau_4^0 \\ 0.332 * X_{t-\tau_4^0-1,5} + 0.268 * X_{t-\tau_4^0-2,5} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Geo}(1/3) & \tau_4^0 < t \leq \tau_5^0 \\ 0.2 * X_{t-\tau_5^0-1,6} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Geo}(4/5) & \tau_5^0 < t \leq \tau_6^0 \\ 0.109 * X_{t-\tau_6^0-1,7} + 0.306 * X_{t-\tau_6^0-2,7} + 0.305 * X_{t-\tau_6^0-3,7} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Geo}(3/4) & \tau_6^0 < t \leq \tau_7^0 \\ 0.3 * X_{t-\tau_7^0-1,8} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Geo}(1/3) & \tau_7^0 < t \leq \tau_8^0 \\ 0.202 * X_{t-\tau_8^0-1,9} + 0.127 * X_{t-\tau_8^0-2,9} + 0.179 * X_{t-\tau_8^0-3,9} + 0.392 * X_{t-\tau_8^0-4,9} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Geo}(1/2) & \tau_8^0 < t \leq \tau_9^0 \\ 0.3 * X_{t-\tau_9^0-1,10} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Geo}(2/3) & \tau_9^0 < t \leq n \end{cases}$$

Model (C1) consists of the first two segments of Model (B) and $\tau_1^0 = 0.5n$. That is

$$X_t = \begin{cases} 0.5 * X_{t-1,1} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Geo}(1/3) & 0 < t \leq [0.5n] \\ 0.249 * X_{t-\tau_1^0-1,2} + 0.254 * X_{t-\tau_1^0-2,2} + 0.297 * X_{t-\tau_1^0-3,2} + Z_t & Z_t \stackrel{i.i.d}{\sim} \text{Geo}(1/2) & [0.5n] < t \leq n \end{cases}$$

Similarly, we design (C2)-(C9) models as follows.

Model (C2) consists of the first three segments of Model (B) and $(\tau_1^0, \tau_2^0) = ([0.3n], [0.6n])$.

Model (C3) consists of the first four segments of Model (B) and $(\tau_1^0, \tau_2^0, \tau_3^0) = ([0.2n], [0.5n], [0.8n])$.

Model (C4) consists of the first five segments of Model (B) and

$$(\tau_1^0, \tau_2^0, \tau_3^0, \tau_4^0) = ([0.2n], [0.4n], [0.6n], [0.8n]).$$

Model (C5) consists of the first six segments of Model (B) and

$$(\tau_1^0, \tau_2^0, \tau_3^0, \tau_4^0, \tau_5^0) = ([0.1n], [0.3n], [0.6n], [0.7n], [0.9n]).$$

Model (C6) consists of the first seven segments of Model (B) and

$$(\tau_1^0, \tau_2^0, \tau_3^0, \tau_4^0, \tau_5^0, \tau_6^0) = ([0.1n], [0.2n], [0.3n], [0.5n], [0.8n], [0.9n]).$$

Model (C7) consists of the first eight segments of Model (B) and

$$(\tau_1^0, \tau_2^0, \tau_3^0, \tau_4^0, \tau_5^0, \tau_6^0, \tau_7^0) = ([0.1n], [0.2n], [0.3n], [0.4n], [0.5n], [0.8n], [0.9n]).$$

Model (C8) consists of the first nine segments of Model (B) and

$$(\tau_1^0, \tau_2^0, \tau_3^0, \tau_4^0, \tau_5^0, \tau_6^0, \tau_7^0, \tau_8^0) = ([0.1n], [0.2n], [0.3n], [0.4n], [0.5n], [0.7n], [0.8n], [0.9n]).$$

Model (C9) consists of the first ten segments of Model (B) and

$$(\tau_1^0, \tau_2^0, \tau_3^0, \tau_4^0, \tau_5^0, \tau_6^0, \tau_7^0, \tau_8^0, \tau_9^0) = ([0.1n], [0.2n], [0.3n], [0.4n], [0.5n], [0.6n], [0.7n], [0.8n], [0.9n]).$$

The corresponding estimated model based on '0753.HK' data set is as follows:

$$E(X_t | \mathcal{F}_{t-1}) = \begin{cases} 0.4220X_{t-1,1} + 0.0404X_{t-2,1} + 0.1558X_{t-3,1} + 5.1320 & 0 < t \leq 420, \\ (0.0737) & (0.0550) & (0.0682) & (1.9947) \\ 0.4392X_{t-4,2} + 0.1215X_{t-4,2,2} + 0.0945X_{t-4,2,2} + 13.3739 & 420 < t \leq 994, \\ (0.0763) & (0.0735) & (0.0556) & (1.6738) \\ 0.4462X_{t-9,3} + 0.0784X_{t-9,3} + 0.0194X_{t-9,3} + 8.6116 & 994 < t \leq 1934, \\ (0.0496) & (0.0386) & (0.0307) & (1.0422) \\ 0.4475X_{t-19,4} + 0.0872X_{t-19,4} + 0.0667X_{t-19,4} + 5.1398 & 1934 < t \leq 3175, \\ (0.0404) & (0.0406) & (0.0309) & (0.6315) \\ 0.3395X_{t-31,5} + 0.0882X_{t-31,5} + 0.1140X_{t-31,5} + 8.6316 & 3175 < t \leq 3701, \\ (0.0475) & (0.0456) & (0.0477) & (0.9677) \\ 0.3543X_{t-37,6} + 0.0408X_{t-37,6} + 13.6244 & 3701 < t \leq 4030, \\ (0.0565) & (0.0609) & (1.5216) \\ 0.3781X_{t-40,7} + 0.0816X_{t-40,7} + 0.1640X_{t-40,7} + 5.7080 & 4030 < t \leq 4550, \\ (0.0502) & (0.0534) & (0.0528) & (0.9627) \end{cases}$$

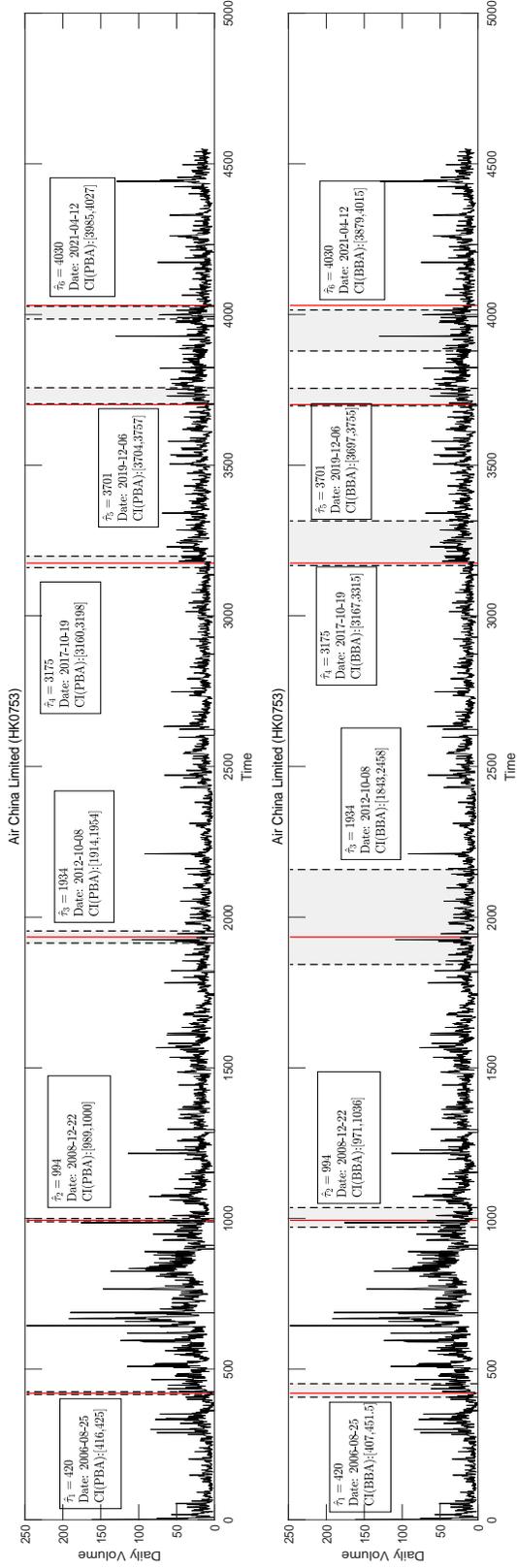


Figure 4: Sample path of the 0753.HK daily trades volume and estimate results given by the LRSM. The red lines are the LRSM change-points estimates and the shaded regions are the corresponding PBA and BBA CIs.

The corresponding estimated model based on ‘LHA.DE’ data set is as follows:

$$\mathbf{E}(X_t | \mathcal{F}_{t-1}) = \begin{cases} 0.2718X_{t-1,1} + 0.1389X_{t-2,1} + 0.0905X_{t-3,1} + 0.0554X_{t-4,1} + 0.0761X_{t-5,1} + 0.6157 & 0 < t \leq 1159, \\ (0.0436) & (0.0365) & (0.0286) & (0.0338) & (0.0318) & (0.0650) \\ 0.3542X_{t-1,2} + 0.1026X_{t-1,2} + 0.0663X_{t-1,2} + 0.0579X_{t-1,2} + 0.0848X_{t-1,2} + 0.0434X_{t-1,2} + 1.2173 & 1159 < t \leq 5886, \\ (0.0174) & (0.0167) & (0.0161) & (0.0151) & (0.0154) & (0.0152) & (0.0138) & (0.0848) \\ 0.4655X_{t-1,3} + 0.2113X_{t-1,3} + 0.0095X_{t-1,3} + 0.0172X_{t-1,3} + 0.0356X_{t-1,3} + 1.1921 & 5886 < t \leq 6495, \\ (0.0597) & (0.0610) & (0.0421) & (0.0384) & (0.0647) & (0.0467) & (0.2632) \\ 0.1988X_{t-1,4} + 0.1343X_{t-1,4} + 3.7911 & 6495 < t \leq 6774, \\ (0.0686) & (0.0552) & (0.4654) \end{cases}$$

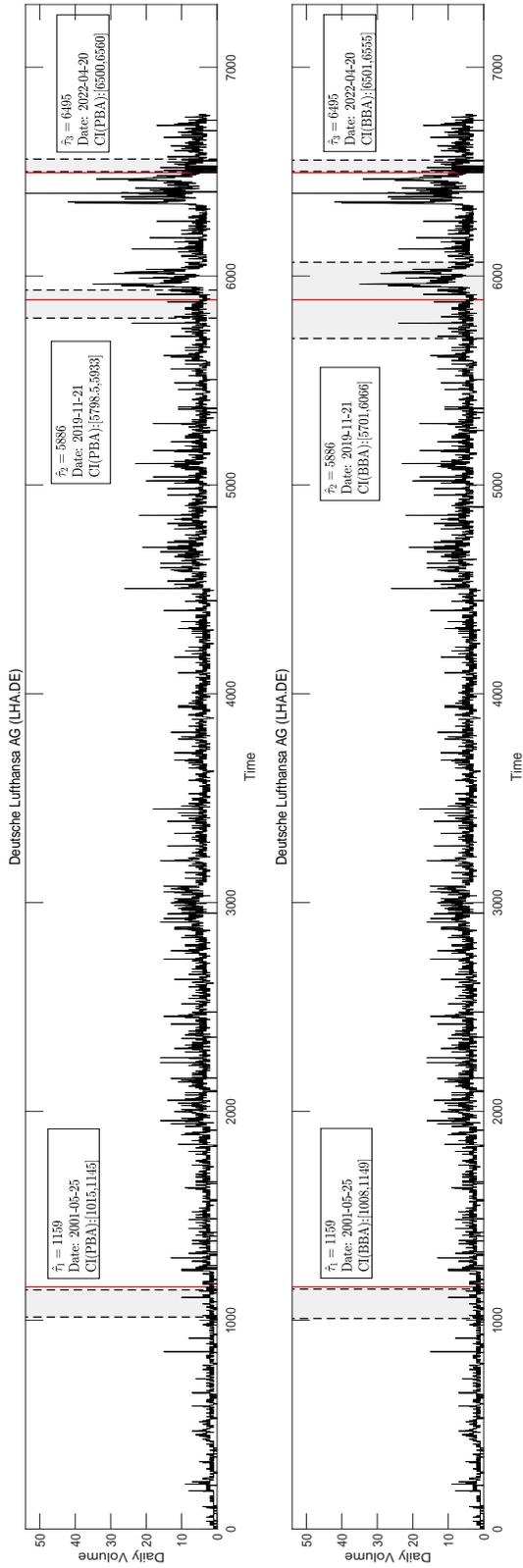


Figure 5: Sample path of the LHA.DE daily trades volume and estimate results given by the LRSM. The red lines are the LRSM change-points estimates and the shaded regions are the corresponding PBA and BBA CIs.

The corresponding estimated model based on 'UAL' data set is as follows:

$$E(X_t | \mathcal{F}_{t-1}) \begin{cases} 0.4411X_{t-1,1} + 0.0410X_{t-2,1} + 0.0895X_{t-3,1} + 0.0803X_{t-4,1} + 0.1100X_{t-5,1} + 0.0179X_{t-6,1} + 0.0830X_{t-7,1} + 0.8574 & 0 < t \leq 3480, \\ (0.0344) & (0.0219) & (0.0252) & (0.0210) & (0.0235) & (0.0213) & (0.0202) & (0.1057) \\ 0.4552X_{t-3481,2} + 0.0339X_{t-3482,2} + 0.1179X_{t-3483,2} + 0.1596X_{t-3484,2} + 0.0392X_{t-3485,2} + 0.1522X_{t-3486,2} + 0.7725 & 3480 < t \leq 4338. \\ (0.0521) & (0.0405) & (0.0430) & (0.0486) & (0.0435) & (0.0461) & (0.2826) \end{cases}$$

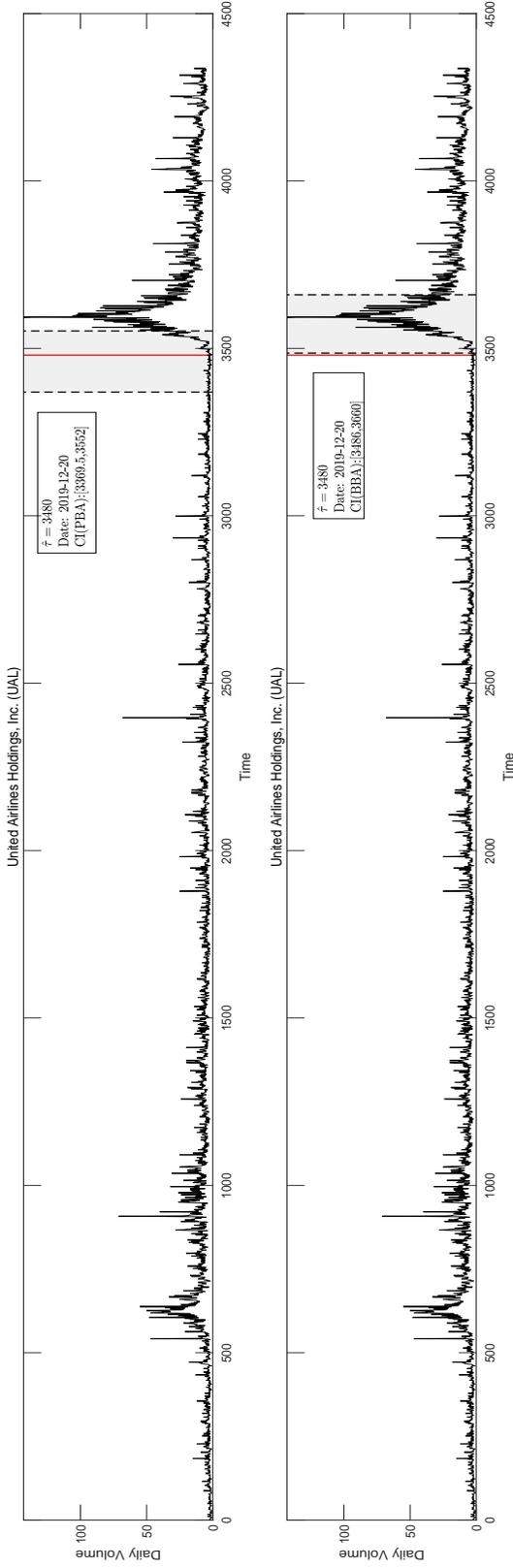


Figure 6: Sample path of the UAL daily trades volume and estimate results given by the LRSM.

The red line is the LRSM change-points estimate and the shaded regions are the corresponding PBA and BBA CIs.

References

- Al-Osh, M.A., Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8(3), 261-275.
- Aue, A., Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, 34(1), 1-16.
- Baudry, J. P., Maugis, C., Michel, B. (2012). Slope heuristics: overview and implementation. *Statistics and Computing*, 22, 455-470.
- Boysen, L., Kempe, A., Liebscher, V., Munk, A., and Wittich, O. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *The Annals of Statistics*, 37(1), 157-183.
- Casini, A., Perron, P. (2018). Structural breaks in time series. *arXiv preprint arXiv:1805.03807*.
- Chan, N.H., Yau, C.Y., Zhang, R.M. (2014). Group LASSO for structural break time series. *Journal of the American Statistical Association*, 109(506), 590-599.
- Chattopadhyay, S., Maiti, R., Das, S., and Biswas, A. (2021). Change-point analysis through integer-valued autoregressive process with application to some COVID-19 data. *Statistica Neerlandica*, 76(1), 4-34.
- Chen, H., Ren, H., Yao, F., and Zou, C. (2021). Data-driven selection of the number of change-points via error rate control. *Journal of the American Statistical Association*, 118(542), 1415-1428.
- Cui, Y., Wu, R., Zheng, Q. (2021). Estimation of change-point for a class of count time series models. *Scandinavian Journal of Statistics*, 48(4), 1277-1313.
- Davis, R.A., Lee, T.C.M., Rodriguez-Yam, G.A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473), 223-239.
- Davis, R.A., Yau, C.Y. (2013). Consistency of minimum description length model selection for piecewise stationary time series models. *Electronic Journal of Statistics*, 7, 381-411.
- Diop, M.L., Kengne, W. (2021b). Piecewise autoregression for general integer-valued time series. *Journal of Statistical Planning and Inference*, 211, 271-286.
- Grunwald, G.K., Hyndman, R.J., Tedesco L., Tweedie, R.L. Non-Gaussian Conditional Linear AR(1) Models. (2000) *Australian and New Zealand Journal of Statistics*, 42(4), 479-495.
- Jackson, B., Sargle, J.D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., Tsai, T. T. (2005). An algorithm for optimal partitioning of data on an interval. *IEEE Signal Process. Letters*, 12(2), 105-108.
- Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2), 633-643.
- Jentsch, C., Weiß, C. H. (2019). Bootstrapping INAR models. *Bernoulli*, 25(3), 2359-2408.

- Kang, J., Lee, S. (2009). Parameter change test for random coefficient integer-valued autoregressive processes with application to polio data analysis. *Journal of Time Series Analysis*, 30(2), 239-258.
- Khoo, W. C., Ong, S. H., Biswas, A. (2017). Modeling time series of counts with a new class of INAR (1) model. *Statistical Papers*, 58, 393-416.
- Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, 19(4), 439-455.
- Lee, S., Ha, J., Na, O., and Na, S. (2003). The cusum test for parameter change in time series models. *Scandinavian Journal of Statistics*, 30(4), 781-796.
- Lee, S., Jo, M. (2022). Bivariate random coefficient integer-valued autoregressive models: Parameter estimation and change point test. *Journal of Time Series Analysis*, doi:10.1111/jtsa.12662.
- Lesigne, E., Voln, D. (2001). Large deviations for martingales. *Stochastic Processes and Their Applications*, 96(1), 143-159.
- McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin*, 21, 645-650.
- Monteiro, M., Scotto, M. G., Pereira, I. (2012). Integer-valued self-exciting threshold autoregressive processes. *Communications in Statistics-Theory and Methods*, 41(15), 2717-2737.
- Ng, W.L., Pan, S., Yau, C.Y. (2022). Bootstrap inference for multiple change-points in time series. *Econometric Theory*, 38(4), 752-792.
- Niu, Y.S., Hao, N., Zhang, H. (2016). Multiple change-point detection: a selective overview. *Statistical Science*, 31(4), 611-623.
- Page, E.S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2), 100-105.
- Page, E.S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4), 523-527.
- Ristić, M.M., Bakouch, H.S., Nastić, A.S. (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *Journal of Statistical Planning and Inference*, 139(7), 2218-2226.
- Safikhani, A., Shojaie, A. (2022). Joint structural break detection and parameter estimation in high-dimensional nonstationary VAR models. *Journal of the American Statistical Association*, 117(537), 251-264.
- Sheng, D.S., Wang, D.H. (2023). Change-points analysis for generalized integer-valued autoregressive model via minimum description length principle. *arXiv preprint arXiv:2307.00466*
- Steutel, F.W., Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 7(5), 893-899.
- Truong, C., Oudre, L., Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, 167, 107299.

- Wei, C.H. (2007). Controlling correlated processes of Poisson counts. *Quality and reliability engineering international*, 23(6), 741-754.
- Wei, C.H. (2009). Controlling jumps in correlated processes of Poisson counts. *Applied Stochastic Models in Business and Industry*, 25(5), 551-564.
- Wei, C.H. (2009). EWMA monitoring of correlated processes of Poisson counts. *Quality Technology and Quantitative Management*, 6(2), 137-153.
- Wei, C.H., Testik, M.C. (2009). CUSUM monitoring of first-order integer-valued autoregressive processes of Poisson counts. *Journal of quality technology*, 41(4), 389-400.
- Wei, C.H. (2010). The INARCH(1) Model for Overdispersed Time Series of Counts. *Communications in Statistics - Simulation and Computation*, 39(6), 1269-1291.
- Yau, C.Y., Zhao, Z. (2016). Inference for multiple change points in time series via likelihood ratio scan statistics. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 78(4), 895-916.
- Yao, Y.C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *The Annals of Statistics*, 15(3), 1321-1328.
- Yao, Y.C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statistics and Probability Letters*, 6(3), 181-189.
- Yu, G.H., Kim, S.G. (2020). Parameter change test for periodic integer-valued autoregressive process. *Communications in Statistics-Theory and Methods*, 49(12), 2898-2912.
- Yu, K., Wang, H., Wei, C.H. (2022). An empirical-likelihood-based structural-change test for INAR processes. *Journal of Statistical Computation and Simulation*, 93(3), 442-458.