Computational Statistics and Data Analysis 177 (2023) 107579

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

Modeling and inference for multivariate time series of counts based on the INGARCH scheme

Sangyeol Lee [a], Dongwon Kim [a], Byungsoo Kim [b,*,1]

[a] Department of Statistics, Seoul National University, Seoul 08826, Republic of Korea
[b] Department of Statistics, Yeungnam University, Gyeongsan 38541, Republic of Korea

ARTICLE INFO

Article history:
Received 28 September 2021
Received in revised form 17 July 2022
Accepted 18 July 2022
Available online 22 July 2022

Keywords:
Multivariate time series of counts
MINGARCH model
QMLE
MDPDE
Change point test

ABSTRACT

Modeling multivariate time series of counts using the integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) scheme is proposed. The key idea is to model each component of the time series with a univariate INGARCH model, where the conditional distribution is modeled with a one-parameter exponential family distribution, and to use a (nonlinear) parametric function of all components to recursively produce the conditional means. It is shown that the proposed multivariate INGARCH (MINGARCH) model is strictly stationary and ergodic. For inference, the quasi-maximum likelihood estimator (QMLE) and the minimum density power divergence estimator (MDPDE) for robust estimation are adopted, and their consistency and asymptotic normality are verified. As an application, the change point test based on the QMLE and MDPDE is illustrated. The Monte Carlo simulation study and real data analysis using the number of weekly syphilis cases in the United States are conducted to confirm the validity of the proposed method.

1. Introduction

In this study, we model a multivariate time series of counts using the integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) scheme and investigate the asymptotic behavior of the quasi-maximum likelihood estimator (QMLE) and the minimum density power divergence estimator (MDPDE) for robust estimation, and apply it to the change point test for model parameters. Since the pioneering work by McKenzie (1985), Al-Osh and Alzaid (1987), Ferland et al. (2006), and Fokianos et al. (2009), integer-valued time series has grown to be an area in time series analysis, and the integer-valued autoregressive (INAR) and INGARCH models have become iconic models that are widely used for diverse research topics in economics, finance, social, physical, and medical sciences and engineering. For a general overview, see Weiß (2018). Many authors have developed theories and applications for these models, including Davis and Wu (2009), Jazi et al. (2012), Zhu (2012a,b), Christou and Fokianos (2014), Davis and Liu (2016), Lee et al. (2016), and Chen et al. (2019).

In the aforementioned studies, the time series of the counts was univariate. However, because time series of counts are often multivariate, multivariate versions of INAR and INGARCH models have also been developed in the literature. Pedeli and Karlis (2011) considered bivariate INAR models with Poisson and negative binomial (NB) innovations. See Darolles et al. (2019) and the papers cited therein. By contrast, Liu (2012) proposed bivariate Poisson-INGARCH models constructed using a trivariate reduction method. See Lee et al. (2018) and Cui and Zhu (2018). Fokianos et al. (2020) recently proposed the

* Corresponding author.
E-mail address: bkim@yu.ac.kr (B. Kim).

[1] Department of Statistics, Yeungnam University, 280 Daehak-ro, Gyeongsan, Gyeongbuk 38541, Republic of Korea.

https://doi.org/10.1016/j.csda.2022.107579

Poisson multivariate INGARCH (MINGARCH) and log-MINGARCH models and established asymptotic theorems for the QMLE of model parameters. Those authors invested more effort to specify the joint distribution of multivariate time series and the marginal distributions of its components, which inevitably escalates the complexity of modeling in comparison with the univariate INGARCH case. To ease the burden of modeling, we propose a simpler scheme that applies the univariate INGARCH model of Davis and Liu (2016) to the component processes of multivariate time series, wherein their marginal distributions are assumed to belong to a one-parameter exponential family, while their conditional means are recursively calculated with a recurrent function of all components.

This simplification significantly increases the tractability of the models because it is not concerned with the specification of the joint distributions. Moreover, as the marginal distributions can differ from each other and the conditional mean equation is not necessarily linear, unlike in previous studies, model flexibility can be enhanced. This modeling scheme is inspired by the viewpoint that the conditional mean equation constitutes the main body of modeling while the specification of underlying joint distributions is not a primary concern, particularly when parameter estimation and subsequent prediction are the main tasks in inference. See Ahmad and Francq (2016), Fokianos et al. (2020), and Lee and Lee (2021). Although the investigation of the joint distribution of the multivariate time series of counts is an important subject, it is not our main concern in the current study. The focus is on developing an easy-to-use MINGARCH model and laying down a theoretical background, such as stationarity and ergodicity.

For parameter estimation, we adopt the QMLE assuming that the components of the time series are independent and derive the strong consistency and asymptotic normality of the QMLE. Moreover, we study a robust estimation method using MDPDE, originally proposed by Basu et al. (1998), to cope with model bias and outliers. We employ MDPDE because it efficiently makes robust inferences under various circumstances, and is capable of adjusting the trade-off between efficiency and robustness by controlling the tuning parameter. See Kang and Lee (2014a), Kim and Lee (2017, 2020b), and Kim et al. (2021), who deal with the MDPDE for INGARCH models and demonstrate the validity of the MDPDE.

As an application, we also consider the problem of detecting a change point based on the cumulative sum (CUSUM) test because of its importance in practice. The change point problem has a long history of returning to Page (1955). Since then, the CUSUM test has been acclaimed as a popular method for detecting parameter changes in the underlying models. For a review, we refer the reader to Csörgő and Horváth (1997), Lee et al. (2003), and Chen and Gupta (2012). Several authors have studied the change point test for INGARCH models as well, including Fokianos and Fried (2010, 2012), Franke et al. (2012), Hudecová (2013), Fokianos et al. (2014), Kang and Lee (2014b), Lee et al. (2016), Hudecová et al. (2017), Lee et al. (2018), and Lee and Lee (2019). For relevant references regarding the MDPDE-based CUSUM test for univariate INGARCH models, refer to Kang and Song (2020) and Kim and Lee (2020a). These previous studies provide the basic tools for establishing the asymptotic theorems in this study. However, they impose some variant regularity conditions, including redundant conditions. All these conditions have been reorganized with clarifications.

The remainder of this study is organized as follows: Sections 2 and 3 introduce the MINGARCH model, QMLE, and MDPDE and investigate their asymptotic properties. Section 4 considers a change point test based on these estimators. Section 5 conducts Monte Carlo simulations and Section 6 provides a real data analysis using the number of weekly syphilis cases in the United States. Finally, Section 7 concludes the paper. The proof of the main theorem is provided in the Appendix A.

2. MINGARCH model and QMLE

Let $Y_t = (Y_{t1}, \ldots, Y_{tm})^T$, $t \geq 1$, be the time series of counts taking values in $\mathbb{N}_0^m$, where $\mathbb{N}_0 = \{0, 1, \ldots\}$. Considering the univariate INGARCH model of Davis and Liu (2016), we aim to build a model for $\{Y_t\}$ such that each component of Y_t has a one-parameter exponential family conditional distribution. More precisely, let

$$p_i(y|\eta) = \exp\{\eta y - A_i(\eta)\}h_i(y), \quad y \in \mathbb{N}_0,$$

which stands for the probability mass function of one-parameter exponential family, wherein η is the natural parameter, $A_i(\eta)$ and $h_i(y)$ are known functions, and both A_i and $B_i = A_i'$, where A_i' stands for the derivative of A_i, are strictly increasing. We then consider the following model.

$$Y_{ti}|\mathcal{F}_{t-1} \sim p_i(y|\eta_{ti}),\ i = 1, \ldots, m, \quad X_t := E(Y_t|\mathcal{F}_{t-1}) = f_\theta(X_{t-1}, Y_{t-1}), \tag{1}$$

where $\mathcal{F}_{t-1}$ denotes the σ-field generated by $Y_{t-1}, Y_{t-2}, \ldots$, $B_i(\eta_{ti}) = X_{ti}$ (ith component of X_t), and f_θ is a non-negative function defined on $[0, \infty)^m \times \mathbb{N}_0^m$ depending on the parameter $\theta \in \Theta \subset \mathbb{R}^d$ for some $d = 1, 2, \ldots$. Although each component of Y_t is modeled using a univariate INGARCH model, the dependence structure is imposed by the conditional mean process X_t. Note that $B_i'(\eta_{ti})$ is the conditional variance of Y_{ti}. Here, the symbols $X_t(\theta)$ and $\eta_t(\theta) = (\eta_{t1}(\theta), \ldots, \eta_{tm}(\theta))^T := B^{-1}(X_t(\theta)) := (B_1^{-1}(X_{t1}(\theta)), \ldots, B_m^{-1}(X_{tm}(\theta)))^T$ is also used to represent X_t and $\eta_t = (\eta_{t1}, \ldots, \eta_{tm})^T$. Then, we can verify that the following theorem holds, the proof of which is outlined in the Appendix A.

Theorem 1. *If the following contraction condition holds for all $x, x' \in [0, \infty)^m$ and $y, y' \in \mathbb{N}_0^m$:*

$$\max_{1 \leq i \leq m} \sup_{\theta \in \Theta} |f_{i,\theta}(x, y) - f_{i,\theta}(x', y')| \leq \lambda_1 ||x - x'||_1 + \lambda_2 ||y - y'||_1 \tag{2}$$

with constants $\lambda_1, \lambda_2 \geq 0$ satisfying $\lambda_1 + \lambda_2 < 1$, where $f_{i,\theta}$ denotes the ith component of f_θ and $||x||_1 = \sum_{i=1}^m |x_i|$ for $x = (x_1, \ldots, x_m)^T$, then there exists a strict stationary and ergodic solution of $\{Y_t\}$ satisfying (1) and the expression $X_t(\theta) = f_\infty^\theta(Y_{t-1}, Y_{t-2}, \ldots)$ is allowed for some non-negative measurable function f_∞^θ defined on $(\mathbb{N}_0^m)^\infty$.

Remark 1. The following implies (2):

$$\sup_{\theta\in\Theta} ||f_\theta(x, y) - f_\theta(x', y')||_1 \leq \lambda_1 ||x - x'||_1 + \lambda_2 ||y - y'||_1. \tag{3}$$

However, the opposite may not be true unless $m = 1$.

The key feature of this modeling scheme is that the formulation is composed only of the marginal distributions of the components of the time series of counts, which remarkably simplifies the modeling task. An advantage of this model is that the components of the time series can have distinctive distributions, which enhances the model's flexibility.

As a popular example of f_θ, we refer to the linear function of the form:

$$f_\theta(x, y) = W + Ax + By,$$

where $\theta = vec(W, A, B)$ denotes the vectorization of an m-dimensional vector W with positive entries and $m \times m$ matrices $A = (a_{ij})$ and $B = (b_{ij})$ with non-negative entries satisfying either

$$\sup_{\theta\in\Theta} \left\{ \sum_{j=1}^m (a_{ij} + b_{ij}) \right\} < 1, \;\; i = 1, \ldots, m, \tag{4}$$

for a compact subset Θ of $\mathbb{R}^{m+2m^2}$, which implies (2), or

$$\sup_{\theta\in\Theta} \left\{ \max_{1\leq j\leq m} \left(\sum_{i=1}^m a_{ij} \right) + \max_{1\leq j\leq m} \left(\sum_{i=1}^m b_{ij} \right) \right\} < 1, \tag{5}$$

which implies (3). Then, Model (1) becomes a linear stationary MINGARCH model:

$$Y_{ti}|\mathcal{F}_{t-1} \sim p_i(y|\eta_{ti}), \quad X_t = W + AX_{t-1} + BY_{t-1}. \tag{6}$$

Considering (4) and (5), we further assume that the true parameter θ_0 is the interior of a compact neighborhood

$$\Theta_1 = \left\{ \theta : \omega_* \leq w_i \leq \omega^*, \; 0 \leq \sum_{j=1}^m (a_{ij} + b_{ij}) \leq 1 - \epsilon, \; i = 1, \ldots, m \right\}$$

or

$$\Theta_2 = \left\{ \theta : \omega_* \leq w_i \leq \omega^*, \; 0 \leq \max_{1\leq j\leq m} \left(\sum_{i=1}^m a_{ij} \right) + \max_{1\leq j\leq m} \left(\sum_{i=1}^m b_{ij} \right) \leq 1 - \epsilon \right\},$$

for some $0 < \omega_* < \omega^*$ and $\epsilon > 0$, where w_i denotes the ith component of W. Moreover, for p_i, we can adopt a Poisson family with $Y_{ti}|\mathcal{F}_{t-1} \sim \text{Poisson}(X_{ti})$ or an NB family with $Y_{ti}|\mathcal{F}_{t-1} \sim \text{NB}(r_i, p_{ti}), \;\; X_{ti} = r_i(1 - p_{ti})/p_{ti}$. Model (6) is universally applicable in practice when either A or B is set diagonal, as this reduces the number of parameters, as in Heinen and Rengifo (2003) and Lee et al. (2018). In practice, too many parameters hamper the accuracy of the parameter estimation; thus, the diagonal setting is recommendable for enhancing the practicality of the MINGARCH models.

Remark 2. If $p(\cdot|\eta)$ is an exponential family probability mass function with canonical function A and if $Z_1 \sim p(\cdot|\eta_1)$ and $Z_2 \sim p(\cdot|\eta_2)$ are two independent random variables with $EZ_i = B(\eta_i)$, where $B = A'$ denotes the derivative of A, we can see that

$$E|Z_1 - Z_2| = |EZ_1 - EZ_2|. \tag{7}$$

This property plays a key role in verifying Theorem 1 regarding the geometric moment condition (GMC) of Wu and Shao (2004). The class of probability mass functions in Theorem 1 does not have to be restricted to the exponential family and can be extended to any class of probability mass functions, with each member of the family satisfying (7). However, this extension is not important for practical usage, particularly when the QMLE method is adopted (see the discussion in Remark 4 below).

Remark 3. Given the initial vectors X_0, Y_0, specific distributions p_i, and f_θ, $\{Y_t\}$ with Y_{ti} following a conditional law p_i can be generated with iid uniform random vectors $U_t = (U_{t1}, \ldots, U_{tm})^T$ with $U_{ti} \sim U(0,1)$. U_t can be generated from multivariate normal random vectors. Specifically, we generate iid $G_t = (G_{t1}, \ldots, G_{tm})^T \sim \mathcal{N}(0, \mathcal{S})$ with $\mathcal{S}_{ii} = 1$ and $\mathcal{S}_{ij} = \rho_{ij} = \rho_{ji} \in (0,1)$. Then, by setting $U_t = (\Phi(G_{t1}), \ldots, \Phi(G_{tm}))^T$, where Φ denotes the standard normal distribution, we obtain $Y_t = (F^{-1}_{1,X_{t1}}(U_{t1}), \ldots, F^{-1}_{m,X_{tm}}(U_{tm}))^T$ with $F_{i,x}(z) = \sum_{y \le z} p_i(y|\eta)$ and $\eta = B_i^{-1}(x)$. We can also consider using copulas to generate $\{U_t\}$, refer to Strelen and Nassaj (2007) and Law (2014).

To estimate the true parameter θ_0, we consider the pseudo-likelihood function:

$$\mathcal{L}(\theta) = \Pi_{t=1}^n \Pi_{i=1}^m \exp\{\tilde{\ell}_{ti}(\theta) + \log h_i(Y_{ti})\} \tag{8}$$

with

$$\tilde{\ell}_{ti}(\theta) = \tilde{\eta}_{ti}(\theta) Y_{ti} - A_i(\tilde{\eta}_{ti}(\theta)) \text{ and } \tilde{\eta}_{ti}(\theta) = B_i^{-1}(\tilde{X}_{ti}(\theta)),$$

where $\tilde{X}_t$ approximates X_t constructed with a preassigned initial vector $\tilde{X}_1$, namely, $\tilde{X}_t = f_\theta(\tilde{X}_{t-1}, Y_{t-1})$, $t \ge 2$. Specifically, we have $\tilde{\ell}_{ti}(\theta) = -\tilde{X}_{ti}(\theta) + Y_{ti} \log \tilde{X}_{ti}(\theta)$ for the Poisson family and $\tilde{\ell}_{ti}(\theta) = Y_{ti} \log(\tilde{X}_{ti}(\theta)/(\tilde{X}_{ti}(\theta) + r_i)) - r_i \log(\tilde{X}_{ti}(\theta) + r_i)$ for the NB family.

The QMLE is then defined as the maximizer of the log-likelihood function:

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}}\, \widetilde{\mathcal{L}}_n(\theta) := \underset{\theta \in \Theta}{\operatorname{argmax}}\, \frac{1}{n} \sum_{t=1}^n \sum_{i=1}^m \tilde{\ell}_{ti}(\theta).$$

To investigate the asymptotic property of the QMLE, we impose the following regularity conditions, where V and $\rho \in (0,1)$ denote a generic integrable random variable and constant, respectively, and $\|\cdot\|$ denotes the L^2 norm for vectors and matrices.

(A0) Condition (2) holds and $f_\theta = (f_{1,\theta}, \ldots, f_{m,\theta})^T$ satisfies

$$\inf_{1 \le i \le m, \theta \in \Theta} \inf_{x \in [0,\infty)^m, y \in \mathbb{N}_0^m} f_{i,\theta}(x, y) \ge c_*$$

for some $c_* > 0$.

(A1) θ_0 is an interior point in the compact parameter space $\Theta \subset \mathbb{R}^d$.

(A2) For all i, it holds that $EY_{1i}^4 < \infty$, $E\left(\sup_{\theta\in\Theta} X_{1i}(\theta)\right)^4 < \infty$, and $E\{\sup_{\theta\in\Theta}(Y_{1i}|\eta_{1i}(\theta)| + |A_i(\eta_{1i}(\theta))|)\} < \infty$.

(A3) There exists $\underline{c} > 0$ such that for all i, t,

$$\inf_{\theta\in\Theta} \inf_{0 \le \delta \le 1} B_i'((1-\delta)\eta_{ti}(\theta) + \delta\tilde{\eta}_{ti}(\theta)) \ge \underline{c},$$

where B_i' denotes the derivative of B_i.

(A4) If there exists $t > 1$ such that $X_t(\theta) = X_t(\theta_0)$ a.s., then $\theta = \theta_0$.

(A5) For all i, t, $\partial^2 X_{ti}(\theta)/\partial\theta\partial\theta^T$ is continuous on Θ,

$$E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial X_{1i}(\theta)}{\partial\theta}\right\|\right)^4 < \infty \text{ and } E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial^2 X_{1i}(\theta)}{\partial\theta\partial\theta^T}\right\|\right)^2 < \infty.$$

(A6) For all i, t,

$$\sup_{\theta\in\Theta}\left\|\frac{\partial\tilde{X}_{ti}(\theta)}{\partial\theta} - \frac{\partial X_{ti}(\theta)}{\partial\theta}\right\| \le V\rho^t \text{ a.s.}$$

(A7) For each i, $\nu^T \partial X_{ti}(\theta_0)/\partial\theta = 0$ a.s. if and only if $\nu = 0$.

(A8) For all i, t,

$$\sup_{\theta\in\Theta}\sup_{0\le\delta\le 1}\left|\frac{B_i''\left((1-\delta)\eta_{ti}(\theta) + \delta\tilde{\eta}_{ti}(\theta)\right)}{B_i'\left((1-\delta)\eta_{ti}(\theta) + \delta\tilde{\eta}_{ti}(\theta)\right)^{5/2}}\right| \le K \text{ for some } K > 0,$$

where B_i'' denotes the second derivative of B_i.

These conditions can be found in Lee and Lee (2019) and Kim and Lee (2020a,b) who deal with univariate INGARCH models. Some of these studies assumed an approximation condition similar to **(A6)** for the second derivatives of the conditional mean. This condition is not easy to check, as might be anticipated at the first glimpse, but it turns out to be redundant in this study. As shown in Theorem 1, **(A0)** ensures the existence of a stationary and ergodic solution to (1). **(A2)** and **(A5)** are the moment conditions related to Y_{ti}, the conditional mean process X_{ti}, and the first and second derivatives of X_{ti}, which are essential for proving the asymptotic properties of QMLE and MDPDE. Regarding the first part of **(A2)**, which depends on conditional distributions, we refer to Section 3.2 of Kim and Lee (2020b). The last part of **(A2)** and **(A4)** are imposed to ensure consistency of the QMLE: see the proof of Theorem 2 in Appendix A. **(A4)** was also used for consistency of the MDPDE. **(A3)** and **(A8)** are required to derive the upper bounds of certain terms obtained from the mean value theorem in the proofs. For an example, see the proof of Lemma 1(iii) in the Appendix A. In practice, past observations and the conditional mean process comprise unobservable time series, and thus, **(A6)** is introduced to approximate $\partial X_{ti}(\theta)/\partial\theta$ with its counterpart $\partial \tilde{X}_{ti}(\theta)/\partial\theta$ comprising only observable time series, essential for proving Lemma 1(iii) and (A.4) in Appendix A. The non-singularity of J and J_α holds for Model (1) because of **(A7)**, where J and J_α are defined in Theorems 3 and 4, respectively. The strong consistency and asymptotic normality addressed below are verified using **(A0)**–**(A8)**, see the Appendix A for the proofs.

Theorem 2. *Suppose that **(A0)**–**(A4)** holds. Then, as $n \to \infty$,*

$$\hat{\theta}_n \longrightarrow \theta_0 \ \text{a.s.}$$

Theorem 3. *Under **(A0)**–**(A8)**, $\hat{\theta}_n$ is asymptotically normal, as n tends to ∞, that is, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, J^{-1} I J^{-1})$, where*

$$I = E\left(\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial \ell_{ti}(\theta_0)}{\partial\theta}\frac{\partial \ell_{tj}(\theta_0)}{\partial\theta^T}\right), \quad J = -E\left(\sum_{i=1}^{m} \frac{\partial^2 \ell_{ti}(\theta_0)}{\partial\theta\partial\theta^T}\right),$$

and $\ell_{ti}(\theta) = \eta_{ti}(\theta)Y_{ti} - A_i(\eta_{ti}(\theta))$.

Remark 4. We can verify that the above theorem holds for any stationary and ergodic process $\{Y_t\}$, not necessarily following Model (1), as far as we can express $X_t = g_\theta(Y_{t-1}, Y_{t-2}, \ldots)$ and $\tilde{X}_t = g_\theta(Y_{t-1}, \ldots, Y_1, y_0, \ldots)$ for some measurable function $g_\theta : (\mathbb{N}_0^m)^\infty \to [0,\infty)^m$, where each entry of g_θ is larger than a positive constant for all θ and the aforementioned regularity conditions are satisfied, as in Ahmad and Francq (2016), Douc et al. (2017), and Lee and Lee (2021). This suggests that, in practice, any one-parameter exponential family, including Poisson and NB distributions, can replace p_i in (8) when obtaining the QMLE. Additionally, even when the true p_i are distinct, they can be taken to be the same in the inferences.

3. MDPDE

Basu et al. (1998) introduced density power divergence d_α between the two density functions g and h defined as:

$$d_\alpha(g,h) := \begin{cases} \int\{g^{1+\alpha}(y) - (1+\frac{1}{\alpha})h(y)g^\alpha(y) + \frac{1}{\alpha}h^{1+\alpha}(y)\}dy, & \alpha > 0, \\ \int h(y)(\log h(y) - \log g(y))dy, & \alpha = 0, \end{cases}$$

and defined the MDPDE based on this. For $Y_1, \ldots, Y_n$ generated from Model (1), the MDPDE for θ_0 is defined as:

$$\hat{\theta}_{\alpha,n} = \operatorname*{argmin}_{\theta\in\Theta} \tilde{\mathcal{L}}_{\alpha,n}(\theta) = \operatorname*{argmin}_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{m} \tilde{\ell}_{\alpha,t,i}(\theta),$$

where

$$\tilde{\ell}_{\alpha,t,i}(\theta) = \begin{cases} \sum_{y=0}^{\infty} p_i^{1+\alpha}(y|\tilde{\eta}_{ti}(\theta)) - \left(1+\frac{1}{\alpha}\right)p_i^\alpha(Y_{ti}|\tilde{\eta}_{ti}(\theta)), & \alpha > 0, \\ -\log p_i(Y_{ti}|\tilde{\eta}_{ti}(\theta)), & \alpha = 0, \end{cases} \tag{9}$$

and $\tilde{\eta}_{ti}(\theta) = B_i^{-1}(\tilde{X}_{ti}(\theta))$ is the same as that in the previous section. Particularly, the MDPDE with $\alpha = 0$ and 1 becomes an MLE and the L^2-distance estimator, respectively. This means that the tuning parameter α controls the trade-off between robustness and asymptotic efficiency. We note that $\tilde{\ell}_{0,t,i}(\theta) = -(\tilde{\ell}_{ti}(\theta) + \log h_i(Y_{ti}))$. In practice, the infinite sum term in equation (9) can be approximated by $\sum_{y=0}^{k} p_i^{1+\alpha}(y|\tilde{\eta}_{ti}(\theta))$ for some k that satisfies $\sum_{y=0}^{k} p_i(y|\tilde{\eta}_{ti}(\theta)) > 1-\epsilon$, where ϵ is a pre-determined small value, such as 10^{-6}. In this case, we particularly have $|\sum_{y=0}^{k+1} p_i^{1+\alpha}(y|\tilde{\eta}_{ti}(\theta)) - \sum_{y=0}^{k} p_i^{1+\alpha}(y|\tilde{\eta}_{ti}(\theta))| < \epsilon$.

The following shows the consistency and asymptotic normality of the MDPDE, the proof of which is provided in the Appendix A.

Theorem 4. *Under **(A0)**-**(A4)**, as $n \to \infty$, $\hat{\theta}_{\alpha,n}$ converges to θ_0 a.s., and under **(A0)**–**(A8)**, $\sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0) \xrightarrow{d} N(0, J_\alpha^{-1} K_\alpha J_\alpha^{-1})$, where*

$$J_\alpha = -E\left(\sum_{i=1}^{m} \frac{\partial^2 \ell_{\alpha,t,i}(\theta_0)}{\partial\theta\partial\theta^T}\right) \text{ and } K_\alpha = E\left(\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial \ell_{\alpha,t,i}(\theta_0)}{\partial\theta}\frac{\partial \ell_{\alpha,t,j}(\theta_0)}{\partial\theta^T}\right),$$

and $\ell_{\alpha,t,i}(\theta)$ is the same as $\tilde{\ell}_{\alpha,t,i}(\theta)$ in (9) with $\eta_{ti}(\theta)$ replaced by $\tilde{\eta}_{ti}(\theta)$.

MDPDE is a functional tool for robust estimation. Because α controls the trade-off between robustness and asymptotic efficiency, selecting the optimal α is an important issue in practice. Generally, as α increases, the robustness increases, but the efficiency decreases. When the number of outliers is small, their role diminishes when the dimension of the time series increases. Indeed, an increase in dimension affects inferences as the abnormality of outliers can be obscured because it makes the data scattered and sparse. Thus, a smaller α would be favorable for the construction of the MDPDE, because a larger α would lose efficiency. In practice, there are several methods for selecting the optimal α. Hong and Kim (2001) proposed a method of selecting an optimal α that minimizes the trace of the estimated asymptotic variance of $\hat{\theta}_{\alpha,n}$ ($\widehat{\text{As.var}}(\hat{\theta}_{\alpha,n})$), which is computed as:

$$\widehat{\text{As.var}}(\hat{\theta}_{\alpha,n}) = \left(\sum_{t=1}^{n}\sum_{i=1}^{m}\frac{\partial^2\tilde{\ell}_{\alpha,t,i}(\hat{\theta}_{\alpha,n})}{\partial\theta\partial\theta^T}\right)^{-1}\left(\sum_{t=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial\tilde{\ell}_{\alpha,t,i}(\hat{\theta}_{\alpha,n})}{\partial\theta}\frac{\partial\tilde{\ell}_{\alpha,t,j}(\hat{\theta}_{\alpha,n})}{\partial\theta^T}\right)\left(\sum_{t=1}^{n}\sum_{i=1}^{m}\frac{\partial^2\tilde{\ell}_{\alpha,t,i}(\hat{\theta}_{\alpha,n})}{\partial\theta\partial\theta^T}\right)^{-1}.$$

This method selects an optimal α such that $\hat{\theta}_{\alpha,n}$ is more robust and stable than the QMLE. Contrastingly, Warwick (2005) and Warwick and Jones (2005) suggested a criterion that minimizes the trace of the estimated asymptotic mean squared error ($\widehat{\text{AMSE}}$) defined by

$$\widehat{\text{AMSE}} = (\hat{\theta}_{\alpha,n} - \hat{\theta}_{1,n})(\hat{\theta}_{\alpha,n} - \hat{\theta}_{1,n})^T + \widehat{\text{As.var}}(\hat{\theta}_{\alpha,n}),$$

where $\hat{\theta}_{1,n}$ denotes the MDPDE with $\alpha = 1$. This method tends to select parameters that are more robust than those in Hong and Kim (2001)'s method because of the penalty term. However, in the presence of a change point in the model parameters, this criterion may no longer be valid; thus, one can opt to choose a small α, for example, $\alpha = 0.1$, as suggested by previous studies; see Section 5 of Lee and Na (2005). In the real data analysis, we evaluate both methods in the selection of α and compare the result with that obtained from $\alpha = 0.1$; see Section 6.

4. Change point test

4.1. QMLE-based test

In this section, we consider CUSUM tests based on standardized residuals scaled with conditional means and score vectors to assess the hypotheses:

$$H_0 : \theta \text{ does not change over } Y_1, \cdots, Y_n \ \ vs. \ \ H_1 : \text{not } H_0.$$

Kang and Lee (2014b) and Lee and Lee (2019) considered the CUSUM test based on the residuals or standardized residuals. Here, we propose a modified version of the CUSUM test based on the residuals $\epsilon_t = (\epsilon_{t1}, \ldots, \epsilon_{tm})^T$ with $\epsilon_{ti} = (Y_{ti} - X_{ti}(\theta_0))/\sqrt{X_{ti}(\theta_0)}$.

Because $\{\epsilon_t, \mathcal{F}_t\}$ forms a square-integrable stationary ergodic martingale difference sequence under **(A0)** and **(A2)**, from the functional central limit theorem, we have

$$T_n^{res} := \max_{1\le k\le n}\frac{1}{n}\left(\sum_{t=1}^{k}\epsilon_t - \frac{k}{n}\sum_{t=1}^{n}\epsilon_t\right)^T\Gamma^{-1}\left(\sum_{t=1}^{k}\epsilon_t - \frac{k}{n}\sum_{t=1}^{n}\epsilon_t\right) \xrightarrow{d} \sup_{0\le s\le 1}||\mathbf{B}_m^{\circ}(s)||^2,$$

where $\Gamma = E\epsilon_1\epsilon_1^T$, which is assumed to be non-singular, and $\mathbf{B}_m^{\circ}(s)$ is an m-dimensional Brownian bridge.

As ϵ_t is unobservable, we employ $\hat{\epsilon}_t = (\hat{\epsilon}_{t1}, \ldots, \hat{\epsilon}_{tm})^T$ with $\hat{\epsilon}_{ti} = (Y_{ti} - \hat{X}_{ti})/\sqrt{\hat{X}_{ti}}$, where $\hat{X}_t = f_{\hat{\theta}_n}(\hat{X}_{t-1}, Y_{t-1})$ for $t \ge 2$ and $\hat{X}_1$ is an arbitrarily chosen initial value. We then consider the following test:

$$\widehat{T}_n^{res} := \max_{1\le k\le n}\frac{1}{n}\left(\sum_{t=1}^{k}\hat{\epsilon}_t - \frac{k}{n}\sum_{t=1}^{n}\hat{\epsilon}_t\right)^T\hat{\Gamma}_n^{-1}\left(\sum_{t=1}^{k}\hat{\epsilon}_t - \frac{k}{n}\sum_{t=1}^{n}\hat{\epsilon}_t\right),$$

where $\hat{\Gamma}_n = \frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t\hat{\epsilon}_t^T$. Because $\widehat{T}_n^{res} = T_n^{res} + o_P(1)$, the proof of which is provided in the Appendix A, we obtain the following theorem.

Theorem 5. *Suppose that conditions **(A0)**–**(A8)** hold. Then, under H_0, as $n \to \infty$,*

$$\widehat{T}_n^{res} \xrightarrow{d} \sup_{0\le s\le 1}||\mathbf{B}_m^{\circ}(s)||^2.$$

We can also consider the CUSUM test based on the score vectors as follows:

$$\widehat{T}_n^{score} := \max_{1\le k\le n} \frac{k^2}{n} \frac{\partial \widetilde{\mathcal{L}}_k(\hat{\theta}_n)}{\partial\theta^T} \hat{I}_n^{-1} \frac{\partial \widetilde{\mathcal{L}}_k(\hat{\theta}_n)}{\partial\theta},$$

where $\hat{I}_n = n^{-1}\sum_{t=1}^n \sum_{i=1}^m \sum_{j=1}^m (\partial\tilde{\ell}_{ti}(\hat{\theta}_n)/\partial\theta)(\partial\tilde{\ell}_{tj}(\hat{\theta}_n)/\partial\theta^T)$, which is a consistent estimator of I following the arguments in Lemma A5 of Kim and Lee (2020a). Then, provided that I is non-singular, we can verify that the following holds, the proof of which is provided in the Appendix A.

Theorem 6. *Suppose that conditions **(A0)–(A8)** hold. Then, under H_0, as $n\to\infty$,*

$$\widehat{T}_n^{score} \xrightarrow{d} \sup_{0\le s\le 1} ||\mathbf{B}_d^\circ(s)||^2.$$

In the univariate INGARCH model, the residual-based CUSUM test outperformed the score vector-based CUSUM test to a certain extent (Lee and Lee, 2019). Thus, $\widehat{T}_n^{res}$ is preferred to $\widehat{T}_n^{score}$ in practice. However, as seen in Kim and Lee (2020a), the score vector-based CUSUM test based on MDPDE is favored when a robust test is requested.

4.2. MDPDE-based test

To construct a robust test using the objective function of the MDPDE, we employ the following test:

$$\widehat{T}_n^{\alpha} := \max_{1\le k\le n} \frac{k^2}{n} \frac{\partial \widetilde{\mathcal{L}}_{\alpha,k}(\hat{\theta}_{\alpha,n})}{\partial\theta^T} \widehat{K}_{\alpha,n}^{-1} \frac{\partial \widetilde{\mathcal{L}}_{\alpha,k}(\hat{\theta}_{\alpha,n})}{\partial\theta}$$

with

$$\widehat{K}_{\alpha,n} = \frac{1}{n}\sum_{t=1}^n \sum_{i=1}^m \sum_{j=1}^m \frac{\partial\tilde{\ell}_{\alpha,t,i}(\hat{\theta}_{\alpha,n})}{\partial\theta} \frac{\partial\tilde{\ell}_{\alpha,t,j}(\hat{\theta}_{\alpha,n})}{\partial\theta^T},$$

which can be shown to be a consistent estimator of K_α like Lemma A5 in Kim and Lee (2020a). Note that $\widehat{T}_n^{\alpha}$ with $\alpha = 0$ is the same as $\widehat{T}_n^{score}$. In the univariate INGARCH model, Kim and Lee (2020a) showed that $\widehat{T}_n^{\alpha}$ outperforms $\widehat{T}_n^{score}$ when outliers contaminate the data. Then, provided that K_α is non-singular, we can obtain the following theorem, see the Appendix A for its proof.

Theorem 7. *Suppose that conditions **(A0)–(A8)** hold. Then, under H_0, as $n\to\infty$,*

$$\widehat{T}_n^{\alpha} \xrightarrow{d} \sup_{0\le s\le 1} \|\mathbf{B}_d^\circ(s)\|^2.$$

Remark 5. For all tests $\widehat{T}_n^{res}$, $\widehat{T}_n^{score}$, and $\widehat{T}_n^{\alpha}$, when the null hypothesis is rejected, the location of the change point is estimated as the time point that maximizes the test statistics. For example, for $\widehat{T}_n^{res}$, it is given as:

$$\hat{\tau} := \operatorname*{argmax}_{1\le k\le n} \frac{1}{n}\left(\sum_{t=1}^k \hat{\epsilon}_t - \frac{k}{n}\sum_{t=1}^n \hat{\epsilon}_t\right)^T \hat{\Gamma}_n^{-1} \left(\sum_{t=1}^k \hat{\epsilon}_t - \frac{k}{n}\sum_{t=1}^n \hat{\epsilon}_t\right).$$

5. Simulations

In this section, we evaluate the performance of the QMLE and MDPDE and the change-point tests $\widehat{T}_n^{res}$, $\widehat{T}_n^{score}$, and $\widehat{T}_n^{\alpha}$ for the following 3-dimensional linear MINGARCH(1,1) model:

$$\begin{aligned} &Y_{ti} \mid \mathcal{F}_{t-1} \sim \text{Poisson}\,(X_{ti})\,, \quad i = 1, 2,\\ &Y_{t3} \mid \mathcal{F}_{t-1} \sim \text{NB}\,(r, p_t)\,, \quad p_t = r/(X_{t3}+r),\\ &X_t = W + AX_{t-1} + BY_{t-1}, \end{aligned}$$

where $W = (\omega_1, \omega_2, \omega_3)^T$ is a 3-dimensional vector, $A = (a_{ij})$ is a 3×3 diagonal matrix, and $B = (b_{ij})$ is a 3×3 matrix. We also consider the same settings with the outliers. In this case, the time series is generated from $Y_t + q_t Z_t$, where Y_t is a 3-dimensional MINGARCH process with the parameters above, q_t are iid Bernoulli random variables with success probability q ($q = 0$ means no outliers), and $Z_t = (Z_{t1}, Z_{t2}, Z_{t3})^T$ are iid 3-dimensional random vectors, where Z_{ti}, $i = 1, 2, 3$, are independent Poisson random variables with intensities $\lambda_i > 0$. Moreover, $\{Y_t\}$, $\{q_t\}$, and $\{Z_t\}$ are independent. We set

$\lambda = (\lambda_1, \lambda_2, \lambda_3)^T$. In the parameter estimation, as X_t is unobservable, we follow the least-squares scheme to obtain an initial estimate of θ; namely, it is obtained as the least-squares estimator that minimizes $CSS(\theta) := \sum_{t=2}^{n} ||Y_t - W - AX_{t-1} - BY_{t-1}||^2$ using proxy $\tilde{X}_t = \sum_{i=1}^{m} Y_{t-i}/m,\ t \geq m+1$ and $\tilde{X}_t = \sum_{i=1}^{t-1} Y_i/(t-1),\ t \leq m$ with $m = 5$ and $\tilde{X}_1 = \sum_{i=1}^{n} Y_i/n$, as a substitute for X_t. All tables reporting the results of our experiments are provided in the supplementary material.

Part 1. We compare the sample mean, standard deviation (SD), root mean squared error (RMSE), and average estimated asymptotic standard deviation (AEASD) of the QMLE and MDPDE under different parameter settings for $n = 500, 1000$, $q = 0, 0.05, 0.1$, and various λ values. The number of repetitions in each simulation is 1000.

$$\textit{Parameter 1}: W = \begin{pmatrix} 1 \\ 1 \\ 1.5 \end{pmatrix},\ A = \begin{pmatrix} 0.2 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.0 \\ 0.0 & 0.0 & 0.2 \end{pmatrix},\ \text{and } B = \begin{pmatrix} 0.2 & 0.1 & 0.0 \\ 0.0 & 0.3 & 0.2 \\ 0.1 & 0.1 & 0.2 \end{pmatrix},$$

$$\textit{Parameter 2}: W = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix},\ A = \begin{pmatrix} 0.2 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix},\ \text{and } B = \begin{pmatrix} 0.3 & 0.2 & 0.0 \\ 0.0 & 0.1 & 0.2 \\ 0.1 & 0.1 & 0.2 \end{pmatrix},$$

$$\textit{Parameter 3}: W = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix},\ A = \begin{pmatrix} 0.2 & 0.0 & 0.1 \\ 0.0 & 0.2 & 0.0 \\ 0.2 & 0.0 & 0.1 \end{pmatrix},\ \text{and } B = \begin{pmatrix} 0.2 & 0.0 & 0.2 \\ 0.0 & 0.3 & 0.0 \\ 0.1 & 0.0 & 0.2 \end{pmatrix}.$$

Herein, we commonly use

$$\mathcal{S} = \begin{pmatrix} 1 & 0.1 & -0.2 \\ 0.1 & 1 & 0 \\ -0.2 & 0 & 1 \end{pmatrix} \text{ and } r = 2,$$

where $\mathcal{S}$ is that of Remark 3. We put the symbol * on the minimal MSEs for each parameter setting in Tables S.1–S.12. Table S.13 describes the sample mean, SD, RMSE, and AEASD of QMLE when A is a non-diagonal matrix for *Parameter 3*. The results of 4-dimensional Poisson MINGARCH models for *Parameter 4* as below are displayed in Table S.14.

$$\textit{Parameter 4}: W = \begin{pmatrix} 2 \\ 2 \\ 1 \\ 2 \end{pmatrix},\ A = \begin{pmatrix} 0.2 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.3 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.2 \end{pmatrix},\ B = \begin{pmatrix} 0.3 & 0.2 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.2 & 0.0 \\ 0.1 & 0.1 & 0.2 & 0.0 \\ 0.1 & 0.0 & 0.1 & 0.3 \end{pmatrix},$$

$$\text{and } \mathcal{S} = \begin{pmatrix} 1 & 0.1 & -0.2 & 0 \\ 0.1 & 1 & 0 & 0.2 \\ -0.2 & 0 & 1 & -0.1 \\ 0 & 0.2 & -0.1 & 1 \end{pmatrix}.$$

We report the results for *Parameter 1* and *Parameter 2* in Tables S.1–S.6 and Tables S.7–S.12, respectively. Table S.1 shows that the QMLE is more efficient than the MDPDE in all measures when outliers do not contaminate the data. Specifically, SD and AEASD of QMLE and MDPDE with $\alpha = 0.1, 0.2$, and 0.3 have similar values and the difference between the SD and AEASD values of MDPDE with $\alpha = 0.5$ and 1 becomes larger compared to the other cases of α. By contrast, Table S.2 shows that when the data is contaminated by outliers with $q = 0.05$, MDPDE outperforms QMLE. Moreover, Table S.3 exhibits that QMLE's performance becomes worsened as the influence of outliers gets stronger in the case of $q = 0.1$ and bigger λ's, which implies that the stronger the influence of the outliers is, the larger the measurement values of QMLE are, and there is a higher chance to find an MDPDE outperforming QMLE. Tables S.4–S.6 for $n = 1000$ show a pattern similar to Tables S.1–S.3 for $n = 500$, but all the measurement values become noticeably smaller. Tables S.7–S.12 list the results for *Parameter 2*, which show a pattern similar to Tables S.1–S.6. Meanwhile, Table S.13 exhibits that the RMSE and SD for the case of a full matrix A have small values when no outliers exist, while the AEASD becomes larger than that for a diagonal matrix case but tends to decrease as the sample size increases. Also, Table S.14 shows that the results for *Parameter 4*, handling a 4-dimensional case, are like those for the 3-dimensional case in the presence of no outliers.

Part 2. Next, we compare the performance of the proposed change point tests at the nominal level of 0.05 for $n = 500, 1000$ with 1000 repetitions. For the critical values, we use 3.0467 for $\widehat{T}_n^{res}$ and 7.8888 for $\widehat{T}_n^{score}$ and $\widehat{T}_n^{\alpha}$, which are empirically obtained 0.95th quantiles of $\sup_{0\leq s\leq 1} ||\mathbf{B}_3^{\circ}(s)||^2$ and $\sup_{0\leq s\leq 1} ||\mathbf{B}_{15}^{\circ}(s)||^2$, respectively. In examining the power, we assume that the parameter change occurs at $\tau = n/2 + 1$ as the power at this point becomes significantly larger than that at other locations. We also report the RMSE of the estimated location of the change point (i.e., RMSE of $\hat{\tau}$) in the parentheses.

Part 2-1. We compare the performance of $\widehat{T}_n^{res}$, $\widehat{T}_n^{score}$, and $\widehat{T}_n^{\alpha}$ with $\alpha = 0.1, 0.2, 0.3, 0.5, 1$ when the parameter changes from (W_0, A_0, B_0) to (W_1, A_1, B_1) as follows:

Case 1: $W_1 = (1+\delta)W_0$, $A_1 = A_0$, and $B_1 = B_0$; namely, only W changes,
Case 2: $W_1 = W_0$, $A_1 = (1+\delta)A_0$, and $B_1 = B_0$; namely, only A changes,

Case 3: $W_1 = W_0$, $A_1 = A_0$, and $B_1 = (1+\delta)B_0$; namely, only B changes, wherein we use the parameter settings:

$$\textit{Case 1}: W_0 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} (:= W_{0,1}) \text{ and } \begin{pmatrix} 2 \\ 4 \\ 4 \end{pmatrix} (:= W_{0,2}),\ A_0 = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.3 \end{pmatrix},$$

$$B_0 = \begin{pmatrix} 0.2 & 0.2 & 0 \\ 0 & 0.1 & 0.2 \\ 0.3 & 0.1 & 0.1 \end{pmatrix}, \text{ and } r = 2,$$

$$\textit{Case 2}: W_0 = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix},\ A_0 = \begin{pmatrix} 0.3 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.2 \end{pmatrix} (:= A_{0,1}) \text{ and } \begin{pmatrix} 0.6 & 0 & 0 \\ 0 & 0.6 & 0 \\ 0 & 0 & 0.4 \end{pmatrix} (:= A_{0,2}),$$

$$B_0 = \begin{pmatrix} 0.1 & 0.1 & 0.1 \\ 0 & 0.1 & 0 \\ 0 & 0.1 & 0.2 \end{pmatrix}, \text{ and } r = 3,$$

$$\textit{Case 3}: W_0 = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix},\ A_0 = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.2 \end{pmatrix},$$

$$B_0 = \begin{pmatrix} 0.15 & 0.1 & 0 \\ 0 & 0.15 & 0 \\ 0.1 & 0 & 0.15 \end{pmatrix} (:= B_{0,1}) \text{ and } \begin{pmatrix} 0.3 & 0.2 & 0 \\ 0 & 0.3 & 0 \\ 0.2 & 0 & 0.3 \end{pmatrix} (:= B_{0,2}),\ \text{and } r = 5.$$

Here,

$$\mathcal{S} = \begin{pmatrix} 1 & 0.2 & -0.1 \\ 0.2 & 1 & 0.1 \\ -0.1 & 0.1 & 1 \end{pmatrix}$$

is used for all cases. Further, we consider $\delta = 0, 0.25, 0.5, 0.75, 1$ and $0, -1/5, -1/3, -3/7, -1/2$.

Part 2-2. We consider the same settings as in *Part 2-1* with outliers and $q = 0.1$.

Tables S.15–S.20 show the results for *Part 2-1* (with no outliers), indicating that all the tests have no size distortions, and the size in each case gets closer to the nominal level as the sample size increases. Overall, $\widehat{T}_n^{res}$ tends to produce better power than the other two tests, and neither $\widehat{T}_n^{score}$ nor $\widehat{T}_n^{\alpha}$ with $\alpha = 0.1$ completely outperforms the other. However, as α increases, the power of $\widehat{T}_n^{\alpha}$ tends to decrease gradually, which is consistent with our intuition.

Tables S.21–S.26 illustrate the results for *Part 2-2* (with outliers) and show a significant difference exists between $\widehat{T}_n^{score}$ and $\widehat{T}_n^{\alpha}$. It is shown that $\widehat{T}_n^{score}$ and $\widehat{T}_n^{\alpha}$ with $\alpha = 0.1$ have no severe size distortions but $\widehat{T}_n^{\alpha}$ with $\alpha > 0.1$ tends to be oversized in most cases. Contrastingly, $\widehat{T}_n^{res}$ is shown to make a stable test except for some cases, as shown in Table S.24. The tables also show that $\widehat{T}_n^{\alpha}$ with $\alpha = 0.1$ largely outperforms the other tests in terms of power. Notably, $\widehat{T}_n^{score}$ produces the lowest power in all cases, whereas $\widehat{T}_n^{\alpha}$ with $\alpha = 0.1$ yields greater power than $\widehat{T}_n^{res}$ in *Case 1* and *Case 2*; however, the result is reversed in *Case 3*. The obtained results also show that $\widehat{T}_n^{\alpha}$ with $\alpha = 0.1$ and 0.2 tends to outperform the other tests regarding the accuracy of estimating the change point location when outliers exist, but $\widehat{T}_n^{res}$ exhibits the highest accuracy in the absence of outliers.

These findings confirm the adequacy of the MDPDE-based change point test in the presence of outliers. Especially, the MDPDE-based test with $\alpha = 0.1$ is demonstrated to be the most suitable in our simulation settings. In the literature, $\alpha = 0.1 \sim 0.3$ is often recommended for practical usage, refer to Lee and Na (2005), Section 5, to cope with a possible chance of outliers or change points. In a multi-dimensional time series, the impact of a small number of outliers tends to decrease. As mentioned earlier, employing an MDPDE-based test with a small α seems to be a reasonable choice for practical usage. Moreover, selecting a universally good α is by no means possible under the circumstances of having change points with high possibilities. Henceforth, we harness the MDPDE-based test with $\alpha = 0.1$ as well as an optimally selected α in our empirical study addressed below.

6. Real data analysis

In this section, we illustrate a real data example using the number of weekly syphilis cases in the United States from January 2007 to December 2010. Time series data can be obtained from the syph dataset in the ZIM R-package of Yang et al. (2018). Among these datasets, we use the 3-dimensional count time series of the weekly number of syphilis cases in Ohio (OH), Florida (FL), and Alabama (AL) with 209 observations. The sample mean and variance are 2.507 and 8.511 for OH, 7.478 and 176.058 for FL, and 1.330 and 4.761 for AL, respectively. Evidently, a prominent outlier exists in FL at $t = 50$

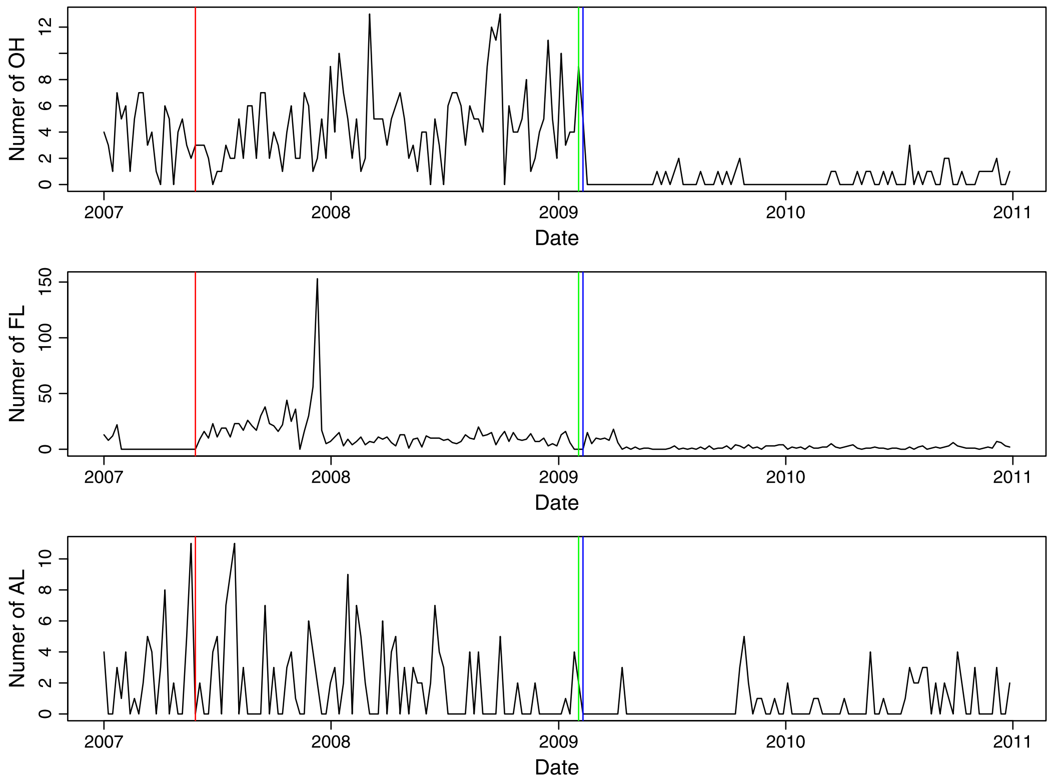

Fig. 1. Weekly count series of OH (top), FL (middle), and AL (bottom).

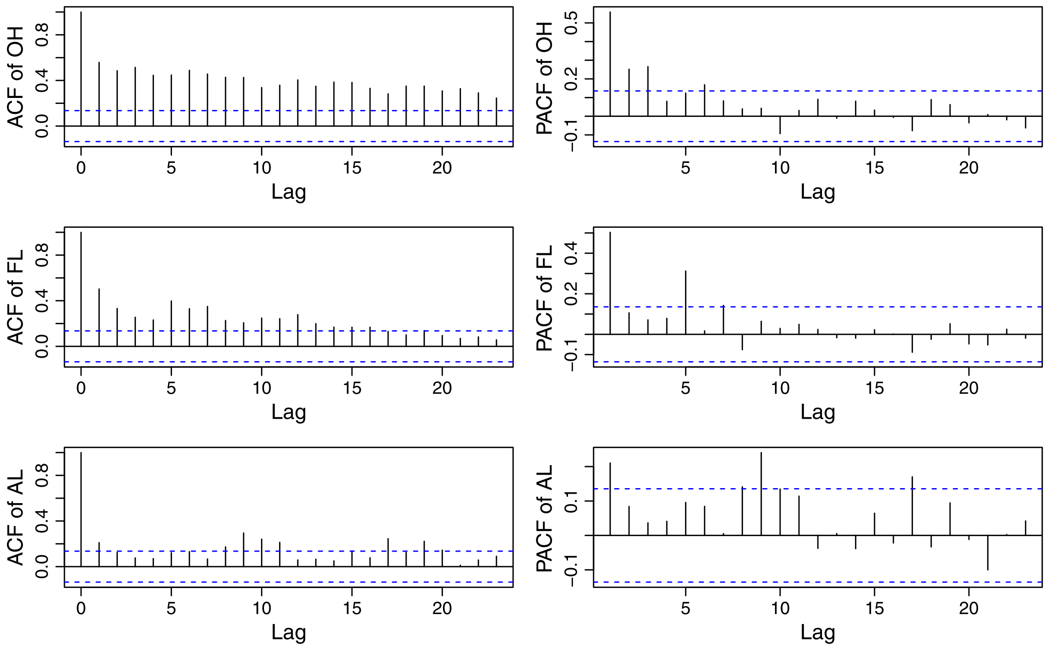

Fig. 2. ACF and PACF of OH (top), FL (middle), and AL (bottom).

(see the time series plot in Fig. 1). The autocorrelation function (ACF) and partial autocorrelation function (PACF) of the three series, and the cross-correlation function (CCF) are shown in Figs. 2 and 3, respectively.

We fit a linear 3-dimensional MINGARCH(1,1) model with diagonal matrix A to this dataset, $Y_t = (Y_{t1}, Y_{t2}, Y_{t3})^T$, $t = 1, \ldots, 209$, as described in Section 5 for the preference of a diagonal matrix model. As a candidate for conditional

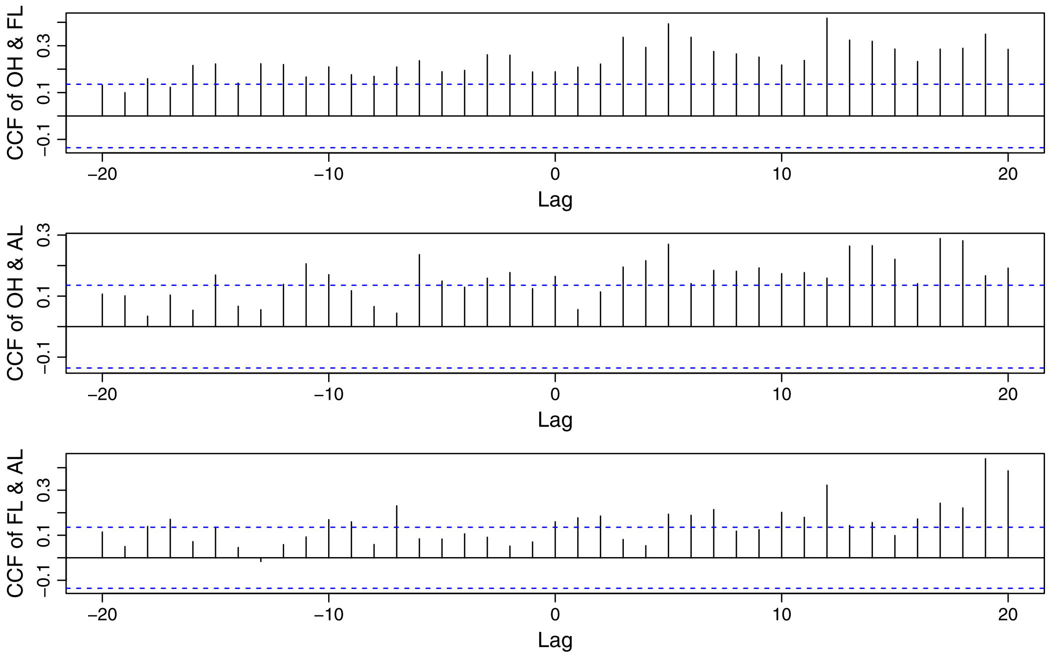

Fig. 3. CCF of OH & FL (top), OH & AL (middle), and FL & AL (bottom).

distributions, we consider an NB(r)-INGARCH(1,1) model, $r = 1, \cdots, 30$, for each Y_{ti}, $i = 1, 2, 3$, and choose r, denoted by $\hat{r}_i$, $i = 1, 2, 3$, to maximize the values of the corresponding log-likelihood functions. For each i, we fit the NB($\hat{r}_i$)- and Poisson-INGARCH(1,1) models to Y_{ti} and obtain the estimated conditional mean values, $\hat{X}_{ti}$. Then, by comparing the two Pearson correlations between Y_{ti} and $\hat{X}_{ti}$, $t = 1, \ldots, 209$, obtained from the two models, we select the one that yields a larger correlation. From this procedure, we select Poisson, NB(2), and Poisson distributions for the conditional distributions of Y_{ti}, $i = 1, 2, 3$.

Based on the selected model, we then determine an optimal α to construct MDPDE that minimizes $\mathrm{tr}(\widehat{\mathrm{As.var}}(\hat{\theta}_{\alpha,209}))$ for $\alpha = 0, 0.1, \ldots, 1$. Consequently, we obtain $\alpha = 0.5$, which is the same as that obtained from $\mathrm{tr}(\widehat{\mathrm{AMSE}})$. In our analysis, however, we not only employ $\alpha = 0.5$ but also $\alpha = 0.1$, as the aforementioned model selection procedure can be misleading because of the possible presence of change points; further, $\alpha = 0.1$ was revealed to have an excellent performance in our simulation study, often recommended by previous studies as well, see Lee and Na (2005).

The parameter estimates, standard errors, $\mathrm{tr}(\widehat{\mathrm{As.var}}(\hat{\theta}_{\alpha,209}))$, and $\mathrm{tr}(\widehat{\mathrm{AMSE}})$ based on the whole observations are displayed in Table 1, wherein the standard errors of $\hat{\theta}_{\alpha,n}$ are evaluated by the diagonal values of the square root of $\widehat{\mathrm{As.var}}(\hat{\theta}_{\alpha,n})$. Especially, it can be seen that $\widehat{b}_{ij} + \widehat{b}_{ji} > 0$ for $i \neq j$ for the QMLE because of the positive cross-correlations between the three series. By contrast, as α increases, b_{ij}, $i \neq j$, shrinks gradually, which indicates that the estimated parameters vary according to the choice of α.

Next, we implement the CUSUM tests, $\widehat{T}_n^{res}$, $\widehat{T}_n^{score}$, $\widehat{T}_n^{0.1}$, and $\widehat{T}_n^{0.5}$ at the nominal level of 0.05, to test for the existence of a change point and detect a change point at the 110th, 111th, 111th, and 22nd time lags, respectively. Here, p-values are obtained from a multivariate Brownian bridge distribution, approximated by the partial sum processes of iid normal random variables of size 10^4 with 10^8 repetitions. They are 0.000, 0.000, 8.080×10^{-6}, and 3.748×10^{-6} for $\widehat{T}_n^{res}$, $\widehat{T}_n^{score}$, $\widehat{T}_n^{0.1}$, and $\widehat{T}_n^{0.5}$, respectively. The locations of the change points are plotted in Fig. 1 as the red ($\widehat{T}_n^{0.5}$), green ($\widehat{T}_n^{res}$), and blue ($\widehat{T}_n^{score}$ and $\widehat{T}_n^{0.1}$) vertical lines. The results demonstrate that the change points detected by $\widehat{T}_n^{res}$, $\widehat{T}_n^{score}$, and $\widehat{T}_n^{0.1}$ are located at almost the same time lag, whereas $\widehat{T}_n^{0.5}$ points to an outrageous time lag far ahead of these points. This result indicates that the optimally chosen α can be misleading in the presence of a change point, and $\alpha = 0.1$ is preferable to 0.5 in this specific example.

Finally, we compare the forecasting performance of our models with a diagonal matrix A with those of the Poisson MINGARCH and log-MINGARCH models with a non-diagonal matrix A from Fokianos et al. (2020), named F1 and F2, respectively. The model selection procedure mentioned above is implemented on $(Y_1, \cdots, Y_{100})$ and the final model is determined as the MINGARCH(1,1) model with Poisson, NB(2), and Poisson marginal distributions. Furthermore, the optimal α for the MDPDE based on $\mathrm{tr}(\widehat{\mathrm{As.var}}(\hat{\theta}_{\alpha,100}))$ is determined as 0.6. For comparison, a one-step-ahead out-of-sample forecast is conducted with a moving window of size 100, that is, we predict Y_{t+101} according to $\hat{Y}_{t+101} = \hat{X}_{t+101}$, where $\hat{X}_{t+101}$ is computed using training data $(Y_{t+1}, \cdots, Y_{t+100})$, $t = 0, 1, \ldots, 108$. As a result, it is obtained that the RMSEs of the forecast based on our models with the QMLE and MDPDE with $\alpha = 0.1$ and 0.6 are 2.323, 2.341, 2.232, respectively, and the RMSEs of F1 and F2 with the QMLE are 2.342, 2.392, respectively. If $\mathrm{tr}(\widehat{\mathrm{AMSE}})$ is used to find the optimal α, $\alpha = 0.7$ is selected, and

Table 1
Parameter estimates (standard error), $\mathrm{tr}(\widehat{\mathrm{As.var}}(\hat{\theta}_{\alpha,209}))$, and $\mathrm{tr}(\widehat{\mathrm{AMSE}})$ for the real data.

α	ω_1	ω_2	ω_3	a_{11}	a_{22}	a_{33}	b_{11}	b_{12}	b_{13}
0(QMLE)	0.881	2.033	0.954	0.154	0.111	0.059	0.499	0.007	0.048
	(0.396)	(2.185)	(1.399)	(0.140)	(0.198)	(0.711)	(0.065)	(0.008)	(0.073)
0.1	0.363	1.094	0.655	0.336	0.072	0.001	0.410	0.017	0.000
	(0.457)	(0.301)	(0.470)	(0.322)	(0.097)	(0.448)	(0.110)	(0.012)	(0.048)
0.5	0.060	0.411	0.317	0.596	0.383	0.000	0.354	0.000	0.000
	(0.035)	(0.162)	(0.234)	(0.093)	(0.139)	(0.421)	(0.083)	(0.009)	(0.035)
1.0	0.075	0.346	0.284	0.554	0.397	0.000	0.380	0.000	0.000
	(0.042)	(1.136)	(0.463)	(0.138)	(1.646)	(1.293)	(0.113)	(0.016)	(0.041)
α	b_{21}	b_{22}	b_{23}	b_{31}	b_{32}	b_{33}		$\mathrm{tr}(\widehat{\mathrm{As.var}}(\hat{\theta}_{\alpha,209}))$	$\mathrm{tr}(\widehat{\mathrm{AMSE}})$
0(QMLE)	0.264	0.468	0.266	0.082	0.000	0.184		10.08	14.46
	(0.335)	(0.196)	(1.567)	(0.069)	(0.022)	(0.107)			
0.1	0.171	0.646	0.160	0.058	0.001	0.084		1.148	2.183
	(0.272)	(0.129)	(0.449)	(0.048)	(0.016)	(0.067)			
0.5	0.000	0.508	0.000	0.002	0.005	0.022		0.346	0.358
	(0.169)	(0.084)	(0.104)	(0.034)	(0.021)	(0.047)			
1.0	0.000	0.445	0.000	0.007	0.002	0.020		9.276	9.276
	(1.517)	(0.921)	(0.443)	(0.025)	(0.019)	(0.084)			

the corresponding RMSE is given as 2.155. This result shows that the MDPDE with an optimal α outperforms the others, and our QMLE also compares well with QMLE based on F1 and F2, indicating that MDPDE is preferable to QMLE based on F1 and F2. However, none of these methods completely outperform the others in all circumstances, and the practitioner's preference should be dependent on a given situation. Overall, our findings strongly confirm the adequacy of our method.

7. Concluding remarks

In this study, we proposed an MINGARCH model with its components following an INGARCH model whose conditional distribution follows a one-parameter exponential family linked with a dynamical equation that recursively evolves the conditional means. This modeling procedure is simple and flexible. The model parameters were estimated by maximizing a quasi-likelihood function, which yielded a consistent QMLE. As the model is merely an approximation of a real phenomenon, we also considered a robust estimation procedure using the MDPDE. As an application, we illustrated the parameter change test, which employs the CUSUM test based on standardized residuals scaled with the conditional mean and score vectors used for obtaining the QMLE and MDPDE. Monte Carlo simulations and real data analysis demonstrated the validity of the proposed method.

Acknowledgements

We thank the Editor, an AE, and two referees for their careful reading and valuable comments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2021R1A2C1004009) (S. Lee) and (NRF-2019R1C1C1004662) (B. Kim).

Appendix A. Proofs

In this Appendix, we prove the theorems in the previous sections. Although the proofs for the consistency and asymptotic normality and the Browian bridge result on the CUSUM tests have some overlaps with those in Lee and Lee (2019) and Kim and Lee (2020a), they are subtly different in detail because of the different sets of conditions.

Proof of Theorem 1. As the lines in the proofs of Propositions 1 and 2 in Appendix C of Davis and Liu (2016) can be applied to Model (1), we only outline the proof without going into great detail. For $u=(u_1,\ldots,u_m)^T$ and $x=(x_1,\ldots,x_m)^T\in\mathbb{R}^m$, we define $g_u(x)=f_\theta(x,\mathbb{F}_x^{-1}(u))$, where $\mathbb{F}_x^{-1}(u)=(F_{1,x_1}^{-1}(u_1),\ldots,F_{m,x_m}^{-1}(u_m))^T$ with $F_{i,x_i}(z)=\sum_{y\le z}p_i(y|\eta_i)$ and $\eta_i=B_i^{-1}(x_i)$. Namely, F_{i,x_i} is the distribution function of $p_i(\cdot|\eta_i)$ with $\sum_{z=0}^{\infty}zp_i(z|\eta_i)=x_i$. Letting $U_t=(U_{t1},\ldots,U_{tm})^T$ be iid random vectors, where $U_{ti}\sim U(0,1)$, $1\le i\le m$, we define $X_t=g_{U_t}(X_{t-1})=f_\theta(X_{t-1},Y_{t-1})$ with $Y_t=\mathbb{F}_{X_{t-1}}^{-1}(U_t)$. Then, we can find a stationary distribution π such that $\{X_t;t\ge1\}$, recursively defined with an initial random vector $X_0\sim\pi$, satisfies the GMC owing to (2) and thus is strictly stationary, which also guarantees the stationarity of $\{Y_t;t\ge1\}$. The $\{(Y_t,X_t);t\ge1\}$ can extend to $\{(Y_t,X_t);t=0,\pm1,\pm2,\ldots\}$ using Kolmogorov's existence theorem, and the existence of f_∞^θ and the ergodicity also can be proven similarly to Proposition 2 of Davis and Liu (2016). Setting $\mathcal{F}_t=\sigma(Y_t,Y_{t-1},\ldots)$, we assert the theorem. □

Proof of Theorem 2. First, we can check that

$$\sup_{\theta\in\Theta}|\tilde{\eta}_{ti}(\theta)-\eta_{ti}(\theta)|+\sup_{\theta\in\Theta}|\tilde{X}_{ti}(\theta)-X_{ti}(\theta)|\le V\rho^t\quad\text{a.s.}\tag{A.1}$$

holds owing to **(A0)**, **(A2)**, and **(A3)**, which implies

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{m}\tilde{\ell}_{ti}(\theta)-\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{m}\ell_{ti}(\theta)\right|\longrightarrow 0 \ \text{ a.s.,} \tag{A.2}$$

due to **(A2)**. Second, since $E\left(\sup_{\theta\in\Theta}|\ell_{ti}(\theta)|\right)<\infty$ owing to **(A2)**, by the uniform strong law of large numbers and ergodicity, we have

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{m}\ell_{ti}(\theta)-E\sum_{i=1}^{m}\ell_{ti}(\theta)\right|\longrightarrow 0 \ \text{ a.s.,}$$

so that by (A.2),

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{m}\tilde{\ell}_{ti}(\theta)-E\sum_{i=1}^{m}\ell_{ti}(\theta)\right|\longrightarrow 0 \ \text{ a.s.} \tag{A.3}$$

Note that $g(\eta):=\eta X_{ti}(\theta_0)-A_i(\eta)$ is concave with a maximum at $\eta=B_i^{-1}(X_{ti}(\theta_0))$, which is $\eta_{ti}(\theta_0)$, and thus, $E\ell_{ti}(\theta)=E\{\eta_{ti}(\theta)X_{ti}(\theta_0)-A_i(\eta_{ti}(\theta))\}$ has a unique maximum at θ_0, owing to **(A4)**. Namely,

$$E\sum_{i=1}^{m}\ell_{ti}(\theta)<E\sum_{i=1}^{m}\ell_{ti}(\theta_0) \ \text{ for all } \ \theta\neq\theta_0,$$

which ensures the strong consistency of the QMLE $\hat{\theta}_n$, owing to **(A1)** and (A.3). This completes the proof. □

Below we address the first and second derivatives of $\ell_{ti}(\theta)$:

$$\begin{aligned}
\frac{\partial\ell_{ti}(\theta)}{\partial\theta}&=h_0(X_{ti}(\theta))\frac{\partial X_{ti}(\theta)}{\partial\theta},\\
\frac{\partial^2\ell_{ti}(\theta)}{\partial\theta\partial\theta^T}&=h_0(X_{ti}(\theta))\frac{\partial^2 X_{ti}(\theta)}{\partial\theta\partial\theta^T}+m_0(X_{ti}(\theta))\frac{\partial X_{ti}(\theta)}{\partial\theta}\frac{\partial X_{ti}(\theta)}{\partial\theta^T},
\end{aligned}$$

with

$$\begin{aligned}
h_0(x)&=\frac{1}{B_i'(B_i^{-1}(x))}(Y_{ti}-x),\\
m_0(x)&=\frac{\partial h_0(x)}{\partial x}=-\frac{B_i''(B_i^{-1}(x))}{B_i'(B_i^{-1}(x))^3}(Y_{ti}-x)-\frac{1}{B_i'(B_i^{-1}(x))}.
\end{aligned}$$

In particular, we can express

$$\begin{aligned}
h_0(X_{ti}(\theta))&=\frac{1}{B_i'(\eta_{ti}(\theta))}(Y_{ti}-X_{ti}(\theta)),\\
m_0(X_{ti}(\theta))&=-\frac{B_i''(\eta_{ti}(\theta))}{B_i'(\eta_{ti}(\theta))^3}(Y_{ti}-X_{ti}(\theta))-\frac{1}{B_i'(\eta_{ti}(\theta))}.
\end{aligned}$$

Lemma 1. *Suppose that **(A0)**-**(A8)** hold. Then, we have for all i, j,*

(i) $E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial^2\ell_{ti}(\theta)}{\partial\theta\partial\theta^T}\right\|\right)<\infty,$
(ii) $E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial\ell_{ti}(\theta)}{\partial\theta}\frac{\partial\ell_{tj}(\theta)}{\partial\theta^T}\right\|\right)<\infty,$
(iii) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{i=1}^{m}\sup_{\theta\in\Theta}\left\|\frac{\partial\ell_{ti}(\theta)}{\partial\theta}-\frac{\partial\tilde{\ell}_{ti}(\theta)}{\partial\theta}\right\|\longrightarrow 0$ a.s.

Proof. Using **(A3)** and **(A8)**, we can easily check the following:

$$|h_0(X_{ti}(\theta))|\le\frac{1}{\underline{c}}(Y_{ti}+X_{ti}(\theta)) \text{ and } |m_0(X_{ti}(\theta))|\le\frac{K}{\underline{c}^{1/2}}(Y_{ti}+X_{ti}(\theta))+\frac{1}{\underline{c}},$$

so that

$$\left\| \frac{\partial \ell_{ti}(\theta)}{\partial \theta} \frac{\partial \ell_{tj}(\theta)}{\partial \theta^T} \right\| \leq |h_0(X_{ti}(\theta))||h_0(X_{tj}(\theta))| \left\| \frac{\partial X_{ti}(\theta)}{\partial \theta} \frac{\partial X_{tj}(\theta)}{\partial \theta^T} \right\|$$

and $$\left\| \frac{\partial^2 \ell_{ti}(\theta)}{\partial \theta \partial \theta^T} \right\| \leq |h_0(X_{ti}(\theta))| \left\| \frac{\partial^2 X_{ti}(\theta)}{\partial \theta \partial \theta^T} \right\| + |m_0(X_{ti}(\theta))| \left\| \frac{\partial X_{ti}(\theta)}{\partial \theta} \frac{\partial X_{ti}(\theta)}{\partial \theta^T} \right\|.$$

Then, (i) and (ii) are asserted by using the Cauchy-Schwarz inequality, **(A2)**, and **(A5)**.

Next, we verify (iii). Note that

$$|h_0(\tilde{X}_{ti}(\theta))| \leq \frac{1}{\underline{c}}(Y_{ti} + X_{ti}(\theta) + |X_{ti}(\theta) - \tilde{X}_{ti}(\theta)|),$$

and by the mean value theorem, we can express

$$\begin{aligned} |h_0(X_{ti}(\theta)) - h_0(\tilde{X}_{ti}(\theta))| &= |m_0(X_{ti}(\theta)^*)||X_{ti}(\theta) - \tilde{X}_{ti}(\theta)| \\ &\leq \left\{ \frac{K}{\underline{c}^{1/2}}(Y_{ti} + X_{ti}(\theta) + |X_{ti}(\theta) - \tilde{X}_{ti}(\theta)|) + \frac{1}{\underline{c}} \right\} |X_{ti}(\theta) - \tilde{X}_{ti}(\theta)|, \end{aligned}$$

where $X_{ti}(\theta)^*$ is an intermediate point between $X_{ti}(\theta)$ and $\tilde{X}_{ti}(\theta)$. Also,

$$\left\| \frac{\partial \ell_{ti}(\theta)}{\partial \theta} - \frac{\partial \tilde{\ell}_{ti}(\theta)}{\partial \theta} \right\| \leq |h_0(\tilde{X}_{ti}(\theta))| \left\| \frac{\partial X_{ti}(\theta)}{\partial \theta} - \frac{\partial \tilde{X}_{ti}(\theta)}{\partial \theta} \right\| + |h_0(X_{ti}(\theta)) - h_0(\tilde{X}_{ti}(\theta))| \left\| \frac{\partial X_{ti}(\theta)}{\partial \theta} \right\|.$$

Then, owing to Lemma 2.1 of Straumann and Mikosch (2006), **(A2)**, **(A5)**, **(A6)**, and (A.1), the supremum of the RHS of the above inequality decays exponentially to 0 a.s. This completes the proof. □

Lemma 2. *Suppose that **(A0)** - **(A8)** hold and let* $\hat{\theta}_n^L = \text{argmax}_{\theta \in \Theta} \mathcal{L}_n(\theta)$ *with*

$$\mathcal{L}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{m} \ell_{ti}(\theta).$$

Then, we have that as $n \to \infty$,

$$\hat{\theta}_n^L \longrightarrow \theta_0 \quad \text{a.s.}$$

and

$$\sqrt{n}(\hat{\theta}_n^L - \theta_0) \xrightarrow{d} N(0, J^{-1} I J^{-1}).$$

Proof. The strong convergence is already proved in the proof of Theorem 2. Thus, asymptotic normality remains to be verified. As $n^{-1/2} \sum_{t=1}^{n} \sum_{i=1}^{m} \partial \ell_{ti}(\hat{\theta}_n^L)/\partial \theta = 0$, using the mean value theorem, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{i=1}^{m} \frac{\partial \ell_{ti}(\theta_0)}{\partial \theta} = \Big(-\frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{m} \frac{\partial^2 \ell_{ti}(\theta_n^*)}{\partial \theta \partial \theta^T} \Big) \sqrt{n}(\hat{\theta}_n^L - \theta_0),$$

where θ_n^* denotes a point lying between θ_0 and $\hat{\theta}_n^L$, which is actually a generic symbol for the component-wise intermediate points $\theta_n^*(i, j)$ for the (i, j)th entry of the $m \times m$ matrices. Using Lemma 1(i), the dominated convergence theorem, and ergodicity, we can show that $-n^{-1} \sum_{t=1}^{n} \sum_{i=1}^{m} \partial^2 \ell_{ti}(\theta_n^*)/\partial \theta \partial \theta^T$ converges to J in probability. Then, with Lemma 1(ii), applying the invariance principle to the martingale differences $\{\sum_{i=1}^{m} \partial \ell_{ti}(\theta_0)/\partial \theta\}$, we assert the lemma. □

Proof of Theorem 3. Note that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{i=1}^{m} \frac{\partial \ell_{ti}(\hat{\theta}_n^L)}{\partial \theta} - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{i=1}^{m} \frac{\partial \ell_{ti}(\hat{\theta}_n)}{\partial \theta} = \Big(\frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{m} \frac{\partial^2 \ell_{ti}(\theta_n^{**})}{\partial \theta \partial \theta^T} \Big) \sqrt{n}(\hat{\theta}_n^L - \hat{\theta}_n),$$

where θ_n^{**} is an intermediate point between $\hat{\theta}_n$ and $\hat{\theta}_n^L$. Then, since

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{i=1}^{m} \frac{\partial \ell_{ti}(\hat{\theta}_n^L)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{i=1}^{m} \frac{\partial \tilde{\ell}_{ti}(\hat{\theta}_n)}{\partial \theta} = 0,$$

we can write

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{i=1}^{m}\frac{\partial\tilde{\ell}_{ti}(\hat{\theta}_n)}{\partial\theta}-\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{i=1}^{m}\frac{\partial\ell_{ti}(\hat{\theta}_n)}{\partial\theta}=\Big(\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{m}\frac{\partial^2\ell_{ti}(\theta_n^{**})}{\partial\theta\partial\theta^T}\Big)\sqrt{n}(\hat{\theta}_n^L-\hat{\theta}_n),$$

and thus, by Lemma 1(iii) and the fact that $n^{-1}\sum_{t=1}^{n}\sum_{i=1}^{m}\partial^2\ell_{ti}(\theta_n^{**})/\partial\theta\partial\theta^T\xrightarrow{P}-J$, we have $\sqrt{n}(\hat{\theta}_n^L-\hat{\theta}_n)=o_P(1)$. Then the theorem is validated by Lemma 2. □

Proof of Theorem 4. Since the proof is similar to that of Kim and Lee (2020b) except for substituting $\ell_{\alpha,t}(\theta)$ and $\tilde{\ell}_{\alpha,t}(\theta)$ in their study with $\sum_{i=1}^{m}\ell_{\alpha,t,i}(\theta)$ and $\sum_{i=1}^{m}\tilde{\ell}_{\alpha,t,i}(\theta)$, respectively, we only sketch the proof without detailing algebras. Similarly to Lemmas 2 and 3 of Kim and Lee (2020b), we can check that

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{m}\tilde{\ell}_{\alpha,t,i}(\theta)-E\sum_{i=1}^{m}\ell_{\alpha,t,i}(\theta)\right|\longrightarrow 0\ \text{ a.s.}$$
$$\text{and }\ E\sum_{i=1}^{m}\ell_{\alpha,t,i}(\theta)>E\sum_{i=1}^{m}\ell_{\alpha,t,i}(\theta_0)\ \text{ for all }\ \theta\neq\theta_0,$$

which imply the consistency of $\hat{\theta}_{\alpha,n}$. Next, note that $\{\sum_{i=1}^{m}\partial\ell_{\alpha,t,i}(\theta_0)/\partial\theta\}$ forms a sequence of stationary martingale differences. In a similar manner to Lemmas 6 and 7 of Kim and Lee (2020b), we can readily check that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{i=1}^{m}\sup_{\theta\in\Theta}\left\|\frac{\partial\ell_{\alpha,t,i}(\theta)}{\partial\theta}-\frac{\partial\tilde{\ell}_{\alpha,t,i}(\theta)}{\partial\theta}\right\|\longrightarrow 0\ \text{ a.s.,}\tag{A.4}$$
$$\hat{\theta}_{\alpha,n}^L\longrightarrow\theta_0\ \text{ a.s.,}$$
$$\sqrt{n}(\hat{\theta}_{\alpha,n}^L-\theta_0)\xrightarrow{d}N(0,J_\alpha^{-1}K_\alpha J_\alpha^{-1}),$$

and J_α is non-singular, where $\hat{\theta}_{\alpha,n}^L=\text{argmin}_{\theta\in\Theta}n^{-1}\sum_{t=1}^{n}\sum_{i=1}^{m}\ell_{\alpha,t,i}(\theta)$. Using all these results and following the lines in the proof of Theorem 2 of Kim and Lee (2020b), we can show that the MDPDE is asymptotically normal. This completes the theorem. □

Proof of Theorem 5. Put $e_{ti}=Y_{ti}-X_{ti}(\theta_0)$ and $\hat{e}_{ti}=Y_{ti}-X_{ti}(\hat{\theta}_n)$. Note that due to **(A0)**, **(A2)**, and the fact that $\sup_{\theta\in\Theta}|\tilde{X}_{ti}(\theta)-X_{ti}(\theta)|\le V\rho^t$ a.s.,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left|(Y_{ti}-\tilde{X}_{ti}(\theta))\tilde{X}_{ti}(\theta)^{-1/2}-(Y_{ti}-X_{ti}(\theta))X_{ti}(\theta)^{-1/2}\right|=o_P(1),$$

so that we only have to verify that for each $1\le i\le m$,

$$\frac{1}{\sqrt{n}}\max_{1\le k\le n}\Bigg|\sum_{t=1}^{k}\hat{e}_{ti}X_{ti}(\hat{\theta}_n)^{-1/2}-\Big(\frac{k}{n}\Big)\sum_{t=1}^{n}\hat{e}_{ti}X_{ti}(\hat{\theta}_n)^{-1/2}$$
$$-\Big\{\sum_{t=1}^{k}e_{ti}X_{ti}(\theta_0)^{-1/2}-\Big(\frac{k}{n}\Big)\sum_{t=1}^{n}e_{ti}X_{ti}(\theta_0)^{-1/2}\Big\}\Bigg|=o_P(1).$$

We express

$$\begin{aligned}(\hat{e}_{ti}-e_{ti})X_{ti}(\hat{\theta}_n)^{-1/2}&=(X_{ti}(\theta_0)-X_{ti}(\hat{\theta}_n))(X_{ti}(\hat{\theta}_n)^{-1/2}-X_{ti}(\theta_0)^{-1/2})\\&\quad+(X_{ti}(\theta_0)-X_{ti}(\hat{\theta}_n))X_{ti}(\theta_0)^{-1/2}\\&:=\Lambda_{t1}+\Lambda_{t2}.\end{aligned}$$

As $|\Lambda_{t1}|\le||\hat{\theta}_n-\theta_0||^2\sup_{\theta\in\Theta}\left|\left|\partial X_{ti}(\theta)/\partial\theta\right|\right|\cdot||q_t(\theta)||$ with $q_t(\theta)=-X_{ti}(\theta)^{-3/2}(\partial X_{ti}(\theta)/\partial\theta)/2$, using **(A5)** and Theorem 3, we have

$$\frac{1}{\sqrt{n}}\max_{1\le k\le n}\left|\sum_{t=1}^{k}\Lambda_{t1}-\Big(\frac{k}{n}\Big)\sum_{t=1}^{n}\Lambda_{t1}\right|=o_P(1).\tag{A.5}$$

Also, as $\Lambda_{t2}=(\theta_0-\hat{\theta}_n)^T\frac{\partial X_{ti}(\theta_0)}{\partial\theta}X_{ti}(\theta_0)^{-1/2}+\xi_t$ with $|\xi_t|\le||\hat{\theta}_n-\theta_0||^2\sup_{\theta\in\Theta}\left|\left|\partial^2X_{ti}(\theta)/\partial\theta\partial\theta^T\right|\right|\cdot X_{ti}(\theta_0)^{-1/2}/2$, we get

$$\frac{1}{\sqrt{n}}\max_{1\le k\le n}\left|\sum_{t=1}^{k}\Lambda_{t2}-\left(\frac{k}{n}\right)\sum_{t=1}^{n}\Lambda_{t2}\right|=o_P(1). \tag{A.6}$$

Combining (A.5) and (A.6), we have

$$\frac{1}{\sqrt{n}}\max_{1\le k\le n}\left|\sum_{t=1}^{k}(\hat{e}_{ti}-e_{ti})X_{ti}(\hat{\theta}_n)^{-1/2}-\left(\frac{k}{n}\right)\sum_{t=1}^{n}(\hat{e}_{ti}-e_{ti})X_{ti}(\hat{\theta}_n)^{-1/2}\right|=o_P(1). \tag{A.7}$$

Meanwhile, using the mean value theorem, we can express

$$e_{ti}(X_{ti}(\hat{\theta}_n)^{-1/2}-X_{ti}(\theta_0)^{-1/2})=(\hat{\theta}_n-\theta_0)^T q_t(\theta_0)e_{ti}+(\hat{\theta}_n-\theta_0)^T\Big(q_t(\theta_n^*)-q_t(\theta_0)\Big)e_{ti},$$

where θ_n^* is an intermediate point between $\hat{\theta}_n$ and θ_0. Then, using **(A5)**, Theorem 3, Donsker's invariance principle, and the dominated convergence theorem, we can obtain

$$\frac{1}{\sqrt{n}}\max_{1\le k\le n}\left|\sum_{t=1}^{k}e_{ti}(X_{ti}(\hat{\theta}_n)^{-1/2}-X_{ti}(\theta_0)^{-1/2})-\left(\frac{k}{n}\right)\sum_{t=1}^{n}e_{ti}(X_{ti}(\hat{\theta}_n)^{-1/2}-X_{ti}(\theta_0)^{-1/2})\right|=o_P(1),$$

which together with (A.7) establishes the theorem. □

Proof of Theorem 6. From Lemma 1(iii), we have

$$\sup_{\theta\in\Theta}\sup_{0\le s\le 1}\frac{[ns]}{\sqrt{n}}\left\|\frac{\partial\widetilde{\mathcal{L}}_{[ns]}(\theta)}{\partial\theta}-\frac{\partial\mathcal{L}_{[ns]}(\theta)}{\partial\theta}\right\|=o_P(1).$$

By the mean value theorem, we can write for each $s\in[0,1]$,

$$\frac{[ns]}{\sqrt{n}}\frac{\partial\mathcal{L}_{[ns]}(\hat{\theta}_n)}{\partial\theta}=\frac{[ns]}{\sqrt{n}}\frac{\partial\mathcal{L}_{[ns]}(\theta_0)}{\partial\theta}+\frac{[ns]}{n}\frac{\partial^2\mathcal{L}_{[ns]}(\theta_{n,s}^*)}{\partial\theta\partial\theta^T}\sqrt{n}(\hat{\theta}_n-\theta_0),$$

where $\theta_{n,s}^*$ is an intermediate point between $\hat{\theta}_n$ and θ_0, chosen to satisfy $\theta_{n,s}^*=\theta_{n,s'}^*$ whenever $[ns]=[ns']$. Since $\partial\widetilde{L}_n(\hat{\theta}_n)/\partial\theta=0$, we obtain that for $s=1$,

$$0=\frac{[ns]}{n}\sqrt{n}\frac{\partial\mathcal{L}_n(\theta_0)}{\partial\theta}+\frac{[ns]}{n}\frac{\partial^2\mathcal{L}_n(\theta_{n,1}^*)}{\partial\theta\partial\theta^T}\sqrt{n}(\hat{\theta}_n-\theta_0)+o_P(1),$$

which implies

$$\begin{aligned}\frac{[ns]}{\sqrt{n}}\frac{\partial\widetilde{\mathcal{L}}_{[ns]}(\hat{\theta}_n)}{\partial\theta}&=\frac{[ns]}{\sqrt{n}}\frac{\partial\mathcal{L}_{[ns]}(\theta_0)}{\partial\theta}-\frac{[ns]}{n}\sqrt{n}\frac{\partial\mathcal{L}_n(\theta_0)}{\partial\theta}\\&\quad+\frac{[ns]}{n}\left(\frac{\partial^2\mathcal{L}_{[ns]}(\theta_{n,s}^*)}{\partial\theta\partial\theta^T}-\frac{\partial^2\mathcal{L}_n(\theta_{n,1}^*)}{\partial\theta\partial\theta^T}\right)\sqrt{n}(\hat{\theta}_n-\theta_0)+o_P(1),\end{aligned} \tag{A.8}$$

uniformly in s. Since $\{\sum_{i=1}^m\partial\ell_{ti}(\theta_0)/\partial\theta\}$ is a sequence of stationary martingale differences, using the functional central limit theorem, we have

$$I^{-1/2}\frac{[ns]}{\sqrt{n}}\frac{\partial\mathcal{L}_{[ns]}(\theta_0)}{\partial\theta}\xrightarrow{d}\mathbf{B}_d(s),$$

where $\mathbf{B}_d(s)$ is a d-dimensional Brownian motion, and thus,

$$\frac{[ns]}{\sqrt{n}}\frac{\partial\mathcal{L}_{[ns]}(\theta_0)}{\partial\theta}-\frac{[ns]}{n}\sqrt{n}\frac{\partial\mathcal{L}_n(\theta_0)}{\partial\theta}\xrightarrow{d}I^{1/2}\mathbf{B}_d^{\circ}(s). \tag{A.9}$$

Furthermore, following the arguments in the proof of Lemma A4 in Kim and Lee (2020a), we can check

$$\max_{1\le k\le n}\frac{k}{n}\left\|\frac{\partial^2\mathcal{L}_k(\bar{\theta}_{n,k})}{\partial\theta\partial\theta^T}+J\right\|\longrightarrow 0\ \text{a.s.},$$

for any $\{\bar{\theta}_{n,k};\,1\le k\le n,n\ge 1\}$ satisfying $\|\bar{\theta}_{n,k}-\theta_0\|\le\|\hat{\theta}_n-\theta_0\|$, which implies

$$\sup_{0\le s\le 1}\left\| \frac{[ns]}{n}\left(\frac{\partial^2 \mathcal{L}_{[ns]}(\theta^*_{n,s})}{\partial\theta\partial\theta^T} - \frac{\partial^2 \mathcal{L}_n(\theta^*_{n,1})}{\partial\theta\partial\theta^T}\right)\sqrt{n}(\hat\theta_n-\theta_0)\right\|$$
$$\le 2\max_{1\le k\le n}\frac{k}{n}\left\|\frac{\partial^2 \mathcal{L}_k(\theta^*_{n,k})}{\partial\theta\partial\theta^T}+J\right\|\left\|\sqrt{n}(\hat\theta_n-\theta_0)\right\| = o_P(1), \tag{A.10}$$

where $\theta^*_{n,k}$ denotes $\theta^*_{n,s}$ with s satisfying $[ns]=k$. From (A.8), (A.9), and (A.10), the theorem is validated. □

Proof of Theorem 7. Let $\mathcal{L}_{\alpha,n}(\theta)=n^{-1}\sum_{t=1}^n\sum_{i=1}^m \ell_{\alpha,t,i}(\theta)$. By virtue of (A.4), we have

$$\sup_{0\le s\le 1}\sup_{\theta\in\Theta}\frac{[ns]}{\sqrt{n}}\left\|\frac{\partial\widetilde{\mathcal{L}}_{\alpha,[ns]}(\theta)}{\partial\theta}-\frac{\partial\mathcal{L}_{\alpha,[ns]}(\theta)}{\partial\theta}\right\| = o_P(1).$$

Also, the mean value theorem yields that for each $s\in[0,1]$,

$$\frac{[ns]}{\sqrt{n}}\frac{\partial\mathcal{L}_{\alpha,[ns]}(\hat\theta_{\alpha,n})}{\partial\theta}=\frac{[ns]}{\sqrt{n}}\frac{\partial\mathcal{L}_{\alpha,[ns]}(\theta_0)}{\partial\theta}+\frac{[ns]}{n}\frac{\partial^2\mathcal{L}_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial\theta\partial\theta^T}\sqrt{n}(\hat\theta_{\alpha,n}-\theta_0),$$

where $\theta^*_{\alpha,n,s}$ is an intermediate point between $\hat\theta_{\alpha,n}$ and θ_0. Similarly to (A.8), using $\partial\widetilde{L}_{\alpha,n}(\hat\theta_{\alpha,n})/\partial\theta=0$, we can express

$$\begin{aligned}\frac{[ns]}{\sqrt{n}}\frac{\partial\widetilde{\mathcal{L}}_{\alpha,[ns]}(\hat\theta_{\alpha,n})}{\partial\theta} &= \frac{[ns]}{\sqrt{n}}\frac{\partial\mathcal{L}_{\alpha,[ns]}(\theta_0)}{\partial\theta}-\frac{[ns]}{n}\sqrt{n}\frac{\partial\mathcal{L}_{\alpha,n}(\theta_0)}{\partial\theta}\\ &\quad+\frac{[ns]}{n}\left(\frac{\partial^2\mathcal{L}_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial\theta\partial\theta^T}-\frac{\partial^2\mathcal{L}_{\alpha,n}(\theta^*_{\alpha,n,1})}{\partial\theta\partial\theta^T}\right)\sqrt{n}(\hat\theta_{\alpha,n}-\theta_0)+o_P(1)\\ &:= M_{n1}(s)+M_{n2}(s)+o_P(1).\end{aligned}\tag{A.11}$$

Because $\{\sum_{i=1}^m\partial\ell_{\alpha,t,i}(\theta_0)/\partial\theta\}$ forms a stationary martingale difference sequence, we have

$$K_\alpha^{-1/2}\frac{[ns]}{\sqrt{n}}\frac{\partial\mathcal{L}_{\alpha,[ns]}(\theta_0)}{\partial\theta}\xrightarrow{d}\mathbf{B}_d(s),$$

and henceforth,

$$M_{n1}\xrightarrow{d}K_\alpha^{1/2}\mathbf{B}_d^\circ. \tag{A.12}$$

Moreover, similarly to (A.10), we can have

$$\sup_{0\le s\le 1}||M_{n2}(s)||\le 2\max_{1\le k\le n}\frac{k}{n}\left\|\frac{\partial^2\mathcal{L}_{\alpha,k}(\theta^*_{\alpha,n,k})}{\partial\theta\partial\theta^T}+J_\alpha\right\|\left\|\sqrt{n}(\hat\theta_{\alpha,n}-\theta_0)\right\|=o_P(1), \tag{A.13}$$

where $\theta^*_{\alpha,n,k}$ is a point lying between θ_0 and $\hat\theta_{\alpha,n}$. Combining (A.11), (A.12) and (A.13), we assert the theorem. □

Appendix B. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2022.107579.

References

Ahmad, A., Francq, C., 2016. Poisson QMLE of count time series models. J. Time Ser. Anal. 37, 291–314. https://doi.org/10.1111/jtsa.12167. https://onlinelibrary.wiley.com/doi/abs/10.1111/jtsa.12167.

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

Basu, A., Harris, I.R., Hjort, N.L., Jones, M.C., 1998. Robust and efficient estimation by minimising a density power divergence. Biometrika 85, 549–559. https://doi.org/10.1093/biomet/85.3.549.

Chen, C.W.S., Khamthong, K., Lee, S., 2019. Markov switching integer-valued generalized auto-regressive conditional heteroscedastic models for Dengue counts. J. R. Stat. Soc., Ser. C, Appl. Stat. 68, 963–983. https://doi.org/10.1111/rssc.12344. https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12344.

Chen, J., Gupta, A.K., 2012. Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance. Birkhäuser, Boston.

Christou, V., Fokianos, K., 2014. Quasi-likelihood inference for negative binomial time series models. J. Time Ser. Anal. 35, 55–78. https://doi.org/10.1111/jtsa.12050. https://onlinelibrary.wiley.com/doi/abs/10.1111/jtsa.12050.

Csörgő, M., Horváth, L., 1997. Limit Theorems in Change-Point Analysis. Wiley, New York. https://www.wiley.com/en-in/Limit+Theorems+in+Change+Point+Analysis-p-9780471955221.

Cui, Y., Zhu, F., 2018. A new bivariate integer-valued GARCH model allowing for negative cross-correlation. Test 27, 428–452. https://doi.org/10.1007/s11749-017-0552-4.

Darolles, S., Le Fol, G., Lu, Y., Sun, R., 2019. Bivariate integer-autoregressive process with an application to mutual fund flows. J. Multivar. Anal. 173, 181–203. https://doi.org/10.1016/j.jmva.2019.02.015. https://www.sciencedirect.com/science/article/pii/S0047259X18304329.

Davis, R.A., Liu, H., 2016. Theory and inference for a class of nonlinear models with application to time series of counts. Stat. Sin. 26, 1673–1707. https://doi.org/10.5705/ss.2014.145t.

Davis, R.A., Wu, R., 2009. A negative binomial model for time series of counts. Biometrika 96, 735–749. https://doi.org/10.1093/biomet/asp029.

Douc, R., Fokianos, K., Moulines, E., 2017. Asymptotic properties of quasi-maximum likelihood estimators in observation-driven time series models. Electron. J. Stat. 11, 2707–2740. https://doi.org/10.1214/17-EJS1299.

Ferland, R., Latour, A., Oraichi, D., 2006. Integer-valued GARCH process. J. Time Ser. Anal. 27, 923–942. https://doi.org/10.1111/j.1467-9892.2006.00496.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.2006.00496.x.

Fokianos, K., Fried, R., 2010. Interventions in INGARCH processes. J. Time Ser. Anal. 31, 210–225. https://doi.org/10.1111/j.1467-9892.2010.00657.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.2010.00657.x.

Fokianos, K., Fried, R., 2012. Interventions in log-linear Poisson autoregression. Stat. Model. 12, 299–322. https://doi.org/10.1177/1471082X1201200401.

Fokianos, K., Gombay, E., Hussein, A., 2014. Retrospective change detection for binary time series models. J. Stat. Plan. Inference 145, 102–112. https://doi.org/10.1016/j.jspi.2013.08.017. https://www.sciencedirect.com/science/article/pii/S0378375813002127.

Fokianos, K., Rahbek, A., Tjøstheim, D., 2009. Poisson autoregression. J. Am. Stat. Assoc. 104, 1430–1439. https://doi.org/10.1198/jasa.2009.tm08270.

Fokianos, K., Støve, B., Tjøstheim, D., Doukhan, P., 2020. Multivariate count autoregression. Bernoulli 26, 471–499. https://doi.org/10.3150/19-BEJ1132.

Franke, J., Kirch, C., Kamgaing, J.T., 2012. Changepoints in times series of counts. J. Time Ser. Anal. 33, 757–770. https://doi.org/10.1111/j.1467-9892.2011.00778.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.2011.00778.x.

Heinen, A., Rengifo, E., 2003. Multivariate modelling of time series count data: an autoregressive conditional Poisson model. CORE Discussion Paper 2003/25. Université Catholique de Louvain, Center for Operations Research and Econometrics (CORE). https://ssrn.com/abstract=640321.

Hong, C., Kim, Y., 2001. Automatic selection of the tuning parameter in the minimum density power divergence estimation. J. Korean Stat. Soc. 30, 453–465. https://www.kss.or.kr/bbs/board.php?bo_table=jkss_online_view&PIDX=13.

Hudecová, Š., 2013. Structural changes in autoregressive models for binary time series. J. Stat. Plan. Inference 143, 1744–1752. https://doi.org/10.1016/j.jspi.2013.05.009. https://www.sciencedirect.com/science/article/pii/S0378375813001122.

Hudecová, Š., Hušková, M., Meintanis, S.G., 2017. Tests for structural changes in time series of counts. Scand. J. Stat. 44, 843–865. https://doi.org/10.1111/sjos.12278. https://onlinelibrary.wiley.com/doi/abs/10.1111/sjos.12278.

Jazi, M.A., Jones, G., Lai, C.D., 2012. First-order integer valued AR processes with zero inflated Poisson innovations. J. Time Ser. Anal. 33, 954–963. https://doi.org/10.1111/j.1467-9892.2012.00809.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.2012.00809.x.

Kang, J., Lee, S., 2014a. Minimum density power divergence estimator for Poisson autoregressive models. Comput. Stat. Data Anal. 80, 44–56. https://doi.org/10.1016/j.csda.2014.06.009. https://www.sciencedirect.com/science/article/pii/S0167947314001807.

Kang, J., Lee, S., 2014b. Parameter change test for Poisson autoregressive models. Scand. J. Stat. 41, 1136–1152. https://doi.org/10.1111/sjos.12088. https://onlinelibrary.wiley.com/doi/abs/10.1111/sjos.12088.

Kang, J., Song, J., 2020. A robust approach for testing parameter change in Poisson autoregressive models. J. Korean Stat. Soc. 49, 1285–1302. https://doi.org/10.1007/s42952-020-00056-7.

Kim, B., Lee, S., 2017. Robust estimation for zero-inflated Poisson autoregressive models based on density power divergence. J. Stat. Comput. Simul. 87, 2981–2996. https://doi.org/10.1080/00949655.2017.1351563.

Kim, B., Lee, S., 2020a. Robust change point test for general integer-valued time series models based on density power divergence. Entropy 22, 493:1–493:22. https://doi.org/10.3390/e22040493. https://www.mdpi.com/1099-4300/22/4/493.

Kim, B., Lee, S., 2020b. Robust estimation for general integer-valued time series models. Ann. Inst. Stat. Math. 72, 1371–1396. https://doi.org/10.1007/s10463-019-00728-0.

Kim, B., Lee, S., Kim, D., 2021. Robust estimation for bivariate Poisson INGARCH models. Entropy 23, 367:1–367:25. https://doi.org/10.3390/e23030367. https://www.mdpi.com/1099-4300/23/3/367.

Law, A.M., 2014. Simulation Modeling and Analysis, 5th ed. McGraw-Hill, New York. https://www.mheducation.com/highered/product/simulation-modeling-analysis-law/M9780073401324.authorbio.html.

Lee, S., Ha, J., Na, O., Na, S., 2003. The CUSUM test for parameter change in time series models. Scand. J. Stat. 30, 781–796. https://doi.org/10.1111/1467-9469.00364. https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9469.00364.

Lee, S., Lee, S., 2021. Exponential family QMLE-based CUSUM test for integer-valued time series. Commun. Stat., Simul. Comput. https://doi.org/10.1080/03610918.2021.1897620 (online published).

Lee, S., Lee, Y., Chen, C.W.S., 2016. Parameter change test for zero-inflated generalized Poisson autoregressive models. Statistics 50, 540–557. https://doi.org/10.1080/02331888.2015.1083020.

Lee, S., Na, O., 2005. Test for parameter change based on the estimator minimizing density-based divergence measures. Ann. Inst. Stat. Math. 57, 553–573. https://doi.org/10.1007/BF02509239.

Lee, Y., Lee, S., 2019. CUSUM test for general nonlinear integer-valued GARCH models: comparison study. Ann. Inst. Stat. Math. 71, 1033–1057. https://doi.org/10.1007/s10463-018-0676-7.

Lee, Y., Lee, S., Tjøstheim, D., 2018. Asymptotic normality and parameter change test for bivariate Poisson INGARCH models. Test 27, 52–69. https://doi.org/10.1007/s11749-016-0510-6.

Liu, H., 2012. Some models for time series of counts. Ph.D. thesis. Columbia University, USA.

McKenzie, E., 1985. Some simple models for discrete variate time series. J. Am. Water Resour. Assoc. 21, 645–650. https://doi.org/10.1111/j.1752-1688.1985.tb05379.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1752-1688.1985.tb05379.x.

Page, E.S., 1955. A test for a change in a parameter occurring at an unknown point. Biometrika 42, 523–527. https://doi.org/10.1093/biomet/42.3-4.523.

Pedeli, X., Karlis, D., 2011. A bivariate INAR(1) process with application. Stat. Model. 11, 325–349. https://doi.org/10.1177/1471082X1001100403.

Straumann, D., Mikosch, T., 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. Ann. Stat. 34, 2449–2495. https://doi.org/10.1214/009053606000000803.

Strelen, J.C., Nassaj, F., 2007. Analysis and generation of random vectors with copulas. In: Henderson, S.G., Biller, B., Hsieh, M.H., Shortle, J., Tew, J.D., Barton, R.R. (Eds.), Proceedings of the 2007 Winter Simulation Conference, pp. 488–496.

Warwick, J., 2005. A data-based method for selecting tuning parameters in minimum distance estimators. Comput. Stat. Data Anal. 48, 571–585. https://doi.org/10.1016/j.csda.2004.03.006. https://www.sciencedirect.com/science/article/pii/S0167947304000672.

Warwick, J., Jones, M.C., 2005. Choosing a robustness tuning parameter. J. Stat. Comput. Simul. 75, 581–588. https://doi.org/10.1080/00949650412331299120.

Weiß, C.H., 2018. An Introduction to Discrete-Valued Time Series. Wiley, New York.

Wu, W., Shao, X., 2004. Limit theorems for iterated random functions. J. Appl. Probab. 41, 425–436. https://doi.org/10.1239/jap/1082999076.

Yang, M., Zamba, G., Cavanaugh, J., 2018. ZIM: zero-inflated models (ZIM) for count time series with excess zeros. r package version 1.1.0. https://CRAN.R-project.org/package=ZIM.

Zhu, F., 2012a. Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. J. Math. Anal. Appl. 389, 58–71. https://doi.org/10.1016/j.jmaa.2011.11.042. https://www.sciencedirect.com/science/article/pii/S0022247X11010626.

Zhu, F., 2012b. Zero-inflated Poisson and negative binomial integer-valued GARCH models. J. Stat. Plan. Inference 142, 826–839. https://doi.org/10.1016/j.jspi.2011.10.002. https://www.sciencedirect.com/science/article/pii/S037837581100379X.