Robust Change Point Test for General Integer-Valued Time Series Models Based on Density Power Divergence

Byungsoo Kim [1],* and Sangyeol Lee [2]

[1]   Department of Statistics, Yeungnam University, Gyeongsan 38541, Korea
[2]   Department of Statistics, Seoul National University, Seoul 08826, Korea; sylee@stats.snu.ac.kr
*   Correspondence: bkim@yu.ac.kr

Abstract: In this study, we consider the problem of testing for a parameter change in general integer-valued time series models whose conditional distribution belongs to the one-parameter exponential family when the data are contaminated by outliers. In particular, we use a robust change point test based on density power divergence (DPD) as the objective function of the minimum density power divergence estimator (MDPDE). The results show that under regularity conditions, the limiting null distribution of the DPD-based test is a function of a Brownian bridge. Monte Carlo simulations are conducted to evaluate the performance of the proposed test and show that the test inherits the robust properties of the MDPDE and DPD. Lastly, we demonstrate the proposed test using a real data analysis of the return times of extreme events related to Goldman Sachs Group stock.

Keywords: integer-valued time series; one-parameter exponential family; minimum density power divergence estimator; density power divergence; robust change point test

1. Introduction

Integer-valued time series models have received widespread attention from researchers and practitioners in diverse research areas. Since the works of McKenzie [1] as well as Al-Osh and Alzaid [2], integer-valued autoregressive (INAR) models have gained popularity in the analysis of correlated time series of counts. Later, as an alternative, Ferland et al. [3] proposed using Poisson integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) models (see Engle [4] and Bollerslev [5]). Since then, INGARCH models have been studied by many authors, such as Fokianos et al. [6], who developed Poisson autoregressive (Poisson AR) models with nonlinear specifications for their intensity processes. The Poisson assumption on INGARCH models has been extended to include negative binomial INGARCH (NB-INGARCH) models (Davis and Wu [7] and Christou and Fokianos [8]), zero-inflated generalized Poisson INGARCH models (Zhu [9,10] and Lee et al. [11]), and one-parameter exponential distribution AR models (Davis and Liu [12]). The latter are also known as general integer-valued time series models and have been studied by, among others, Diop and Kengne [13] and Lee and Lee [14], who considered change point tests for these models.

The change point problem is a core issue in time series analysis because changes can occur in underlying model parameters owing to critical events or policy changes, and ignoring such changes can result in false conclusions. Numerous studies exist on change point analysis in time series models; refer to Kang and Lee [15] and Lee and Lee [14], and the articles cited therein, for the background and history of change points in integer-valued time series models. Lee and Lee [14] conducted a comparison study of the performance of various cumulative sum (CUSUM) tests using score vectors and residuals through the Monte Carlo simulations. In their work, the conditional maximum likelihood

estimator (CMLE) is used for the parameter estimation and also the construction of the CUSUM tests. However, the CMLE is often damaged by outliers, and so is the performance of the CMLE-based CUSUM test. In general, outliers easily mislead the CUSUM test since they can be mistakenly taken for abrupt changes; in the opposite, they can misidentify change points in their presence on time series. Among the robust estimation methods, we adopt the minimum density power divergence estimator (MDPDE) approach—proposed by Basu et al. [16]—as a remedy and propose to use the density power divergence (DPD)-based test as a robust change point test.

The MDPDE method is well known for consistently making robust inferences in various situations, and the trade-off between efficiency and robustness is managed via the tuning parameter. Basu et al. [16] introduced the MDPDE using the independent and identically distributed observations, and later, Ghosh and Basu [17] extended their method to the independent but not identically distributed samples. For earlier works in the context of time series, see Lee and Song [18], Kim and Lee [19], Kang and Lee [20], and Kim and Lee [21], who deal with the MDPDE for GARCH models, multivariate times series, and (zero-inflated) Poisson AR models. Kim and Lee [22] demonstrated that the MDPDE for general integer-valued time series models has strong robust properties, with little loss in asymptotic efficiency relative to the CMLE. This motivates us to use the MDPDE to construct a robust change point test for general integer-valued time series models. More precisely, we anticipate that the robust property of the MDPDE would be inherited to the proposed change point test, so that the influence of outliers should be reduced when performing a parameter change test in the presence of outliers. Although the problem of testing for a parameter change in integer-valued time series models has been investigated by many researchers, the testing procedure for observations with outliers has not been widely studied. This motivates us to develop a MDPDE-based robust change point test for general integer-valued time series models.

Kang and Song [23] proposed an estimate-based robust CUSUM test that uses the MDPDE to detect parameter changes in Poisson AR models. However, this type of test is known to suffer from severe size distortions, especially when the true parameter lies at the boundary of the parameter space. Thus, we use the test deduced based on an empirical version of the DPD, which is the objective function of the MDPDE. Song and Kang [24] and Kang and Song [25] applied DPD-based change point tests in GARCH models and Poisson AR models, respectively. However, the DPD approach basically shares the same spirit as the score-based CUSUM test of Lee and Lee [14] (see Remark 3 in Section 2.2), in that both are based on derivatives of objective functions. Thus, the idea is easily adapted to one-parameter exponential family AR models. As for a parameter change test for independent samples based on divergence measures, see Batsidis et al. [26,27], who consider the ϕ-divergence as a measure. We also refer to Martín and Pardo [28], who point out the importance of a Wald-type test based on DPD in dealing with the change point problem.

Monte Carlo simulations are conducted to evaluate the performance of the proposed test. Here, we compare the DPD-based test and the score-based CUSUM test to demonstrate the superiority of the proposed test in the presence of outliers. Then, we provide a real data analysis of the return times of extreme events related to Goldman Sachs Group (GS) stock to illustrate the proposed test. The paper proceeds as follows. Section 2 constructs the DPD-based change point test for general integer-valued time series models, and states its weak convergence theorem. Section 3 presents a simulation study and a real data analysis. Section 4 concludes the paper. All proofs are provided in the Appendix A.

2. Construction of the MDPDE and Change Point Test

2.1. MDPDE for General Integer-Valued Time Series Models

Let $Y_1, Y_2, \ldots$ be the observations generated from general integer-valued time series models with the conditional distribution of the one-parameter exponential family:

$$Y_t | \mathcal{F}_{t-1} \sim p(y | \eta_t), \quad X_t := E(Y_t | \mathcal{F}_{t-1}) = f_\theta(X_{t-1}, Y_{t-1}), \tag{1}$$

where $\mathcal{F}_{t-1}$ is a σ-field generated by $Y_{t-1}, Y_{t-2}, \ldots$ and $f_\theta(x, y)$ is a non-negative bivariate function defined on $[0, \infty) \times \mathbb{N}_0$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, depending on the parameter $\theta \in \Theta \subset \mathbb{R}^d$, and satisfies $\inf_{\theta \in \Theta} f_\theta(x, y) \geq x^*$ for some $x^* > 0$ for all x, y. Here, $p(\cdot|\cdot)$ is a probability mass function, given by

$$p(y|\eta) = \exp\{\eta y - A(\eta)\}h(y), \quad y \geq 0,$$

where η is the natural parameter and $A(\eta)$ and $h(y)$ are known functions. This distribution family includes several famous discrete distributions, such as the Poisson, negative binomial, and binomial distributions. If $B(\eta) = A'(\eta)$, $B(\eta_t)$ and $B'(\eta_t)$ become the conditional mean and variance of Y_t, and $X_t = B(\eta_t)$. The derivative of $A(\eta)$ exists for the exponential family; see Lehmann and Casella [29]. Since $B'(\eta_t) = Var(Y_t|\mathcal{F}_{t-1}) > 0$, $B(\eta)$ is strictly increasing, and since $B(\eta_t) = E(Y_t|\mathcal{F}_{t-1}) > 0$, $A(\eta)$ is also strictly increasing. To emphasize the role of θ, we also use $X_t(\theta)$ and $\eta_t(\theta) = B^{-1}(X_t(\theta))$ to stand for X_t and η_t, respectively.

Davis and Liu [12] showed that the assumption below ensures the strict stationarity and ergodicity of $\{(X_t, Y_t)\}$:

(A0) For all $x, x' \geq 0$ and $y, y' \in \mathbb{N}_0$,

$$\sup_{\theta \in \Theta} |f_\theta(x, y) - f_\theta(x', y')| \leq \omega_1 |x - x'| + \omega_2 |y - y'|,$$

where $\omega_1, \omega_2 \geq 0$ satisfy $\omega_1 + \omega_2 < 1$.

They also demonstrated that there exists a measurable function $f_\infty^\theta : \mathbb{N}_0^\infty \to [0, \infty)$, such that $X_t(\theta) = f_\infty^\theta(Y_{t-1}, Y_{t-2}, \ldots)$ almost surely (a.s.).

Meanwhile, the DPD d_α between two density functions g and h is defined as

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

For a parametric family $\{G_\theta, \theta \in \Theta\}$ with densities given by $\{g_\theta\}$ and a distribution H with density h, the minimum DPD functional $T_\alpha(H)$ is defined by $d_\alpha(h, g_{T_\alpha(H)}) = \min_{\theta \in \Theta} d_\alpha(h, g_\theta)$. In particular, if $H = G_{\theta_0} \in \{G_\theta\}$, $T_\alpha(G_{\theta_0}) = \theta_0$. Then, given a random sample $Y_1, \ldots, Y_n$ with unknown density h, the MDPDE is defined by

$$\hat{\theta}_{\alpha, n} = \operatorname*{argmin}_{\theta \in \Theta} L_{\alpha, n}(\theta),$$

where $L_{\alpha, n}(\theta) = \frac{1}{n} \sum_{t=1}^n l_{\alpha, t}(\theta)$ and

$$l_{\alpha, t}(\theta) = \begin{cases} \int g_\theta^{1+\alpha}(y)dy - \left(1 + \frac{1}{\alpha}\right)g_\theta^\alpha(Y_t), & \alpha > 0, \\ -\log g_\theta(Y_t), & \alpha = 0. \end{cases}$$

When $\alpha = 0$ and 1, the MDPDE becomes the MLE and the L^2-distance estimator, respectively. Basu et al. [16] revealed that $\hat{\theta}_{\alpha, n}$ is consistent for $T_\alpha(H)$ and asymptotically normal. Furthermore, the estimator is robust against outliers, but still exhibits high efficiency when the true distribution belongs to a parametric family $\{G_\theta\}$ and α is close to zero. The tuning parameter α controls the trade-off between robustness and asymptotic efficiency. A large α escalates the robustness while a small α yields greater efficiency. The conditional version of the MDPDE is defined similarly (cf. Section 2 of Kim and Lee [22]).

For $Y_1, \ldots, Y_n$ generated from (1), the MDPDE for general integer-valued time series models is defined as

$$\hat{\theta}_{\alpha, n} = \operatorname*{argmin}_{\theta \in \Theta} \widetilde{L}_{\alpha, n}(\theta) = \operatorname*{argmin}_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^n \tilde{l}_{\alpha, t}(\theta), \tag{2}$$

where

$$
\tilde{l}_{\alpha,t}(\theta) = \begin{cases} \sum_{y=0}^{\infty} p^{1+\alpha}(y|\tilde{\eta}_t(\theta)) - \left(1 + \frac{1}{\alpha}\right) p^{\alpha}(Y_t|\tilde{\eta}_t(\theta)), & \alpha > 0, \\ -\log p(Y_t|\tilde{\eta}_t(\theta)), & \alpha = 0, \end{cases}
\tag{3}
$$

and $\tilde{\eta}_t(\theta) = B^{-1}(\widetilde{X}_t(\theta))$ is updated recursively using the following equations:

$$
\widetilde{X}_t(\theta) = f_\theta(\widetilde{X}_{t-1}(\theta), Y_{t-1}), \ t = 2, 3, \ldots, \ \widetilde{X}_1(\theta) = \widetilde{X}_1,
$$

with an arbitrarily chosen initial value $\widetilde{X}_1$. The MDPDE with $\alpha = 0$ becomes the CMLE from (3).

Kim and Lee [22] showed that under the regularity conditions **(A0)–(A9)** stated below, the MDPDE is strongly consistent and asymptotically normal. Conditions **(A10)** and **(A11)** are imposed to derive the limiting null distribution of the DPD-based change point test in Section 2.2. Below, V and $\rho \in (0,1)$ represent a generic integrable random variable and a constant, respectively; the symbol $\| \cdot \|$ denotes the L^2-norm for matrices and vectors; and $E(\cdot)$ is taken under θ_0, where θ_0 denotes the true value of θ.

(A1) θ_0 is an interior point in the compact parameter space $\Theta \subset \mathbb{R}^d$.
(A2) $E\left(\sup_{\theta \in \Theta} X_1(\theta)\right)^4 < \infty$.
(A3) $\inf_{\theta \in \Theta} \inf_{0 \leq \delta \leq 1} B'((1-\delta)\eta_t(\theta) + \delta\tilde{\eta}_t(\theta)) \geq \underline{c}$ for some $\underline{c} > 0$.
(A4) $EY_1^4 < \infty$.
(A5) If there exists $t \geq 1$, such that $X_t(\theta) = X_t(\theta_0)$ a.s., then $\theta = \theta_0$.
(A6) $\sup_{\theta \in \Theta} \sup_{0 \leq \delta \leq 1} \left| \frac{B''((1-\delta)\eta_t(\theta) + \delta\tilde{\eta}_t(\theta))}{B'((1-\delta)\eta_t(\theta) + \delta\tilde{\eta}_t(\theta))^{5/2}} \right| \leq K$ for some $K > 0$.
(A7) The mapping $\theta \mapsto f_\infty^\theta$ is twice continuously differentiable with respect to θ, and satisfies

$$
E\left(\sup_{\theta \in \Theta} \left\| \frac{\partial f_\infty^\theta(Y_0, Y_{-1}, \ldots)}{\partial \theta} \right\| \right)^4 < \infty \quad \text{and} \quad E\left(\sup_{\theta \in \Theta} \left\| \frac{\partial^2 f_\infty^\theta(Y_0, Y_{-1}, \ldots)}{\partial \theta \partial \theta^T} \right\| \right)^2 < \infty.
$$

(A8) $\sup_{\theta \in \Theta} \left\| \frac{\partial \widetilde{X}_t(\theta)}{\partial \theta} - \frac{\partial X_t(\theta)}{\partial \theta} \right\| \leq V\rho^t$ a.s.
(A9) $\nu^T \frac{\partial X_t(\theta_0)}{\partial \theta} = 0$ a.s. implies $\nu = 0$.
(A10) $\sup_{\theta \in \Theta} \left\| \frac{\partial^2 \widetilde{X}_t(\theta)}{\partial \theta \partial \theta^T} - \frac{\partial^2 X_t(\theta)}{\partial \theta \partial \theta^T} \right\| \leq V\rho^t$ a.s.
(A11) $\sup_{\theta \in \Theta} \sup_{0 \leq \delta \leq 1} \left| \frac{B^{(3)}((1-\delta)\eta_t(\theta) + \delta\tilde{\eta}_t(\theta))}{B'((1-\delta)\eta_t(\theta) + \delta\tilde{\eta}_t(\theta))^4} \right| \leq M$ for some $M > 0$.

Proposition 1. *Under* **(A0)–(A5)**, $\hat{\theta}_{\alpha,n} \longrightarrow \theta_0$ *a.s. as* $n \to \infty$, *and further, under* **(A0)–(A9)**,

$$
\sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0) \xrightarrow{d} N(0, J_\alpha^{-1} K_\alpha J_\alpha^{-1}) \text{ as } n \to \infty,
$$

where

$$
J_\alpha = -E\left(\frac{\partial^2 l_{\alpha,t}(\theta_0)}{\partial \theta \partial \theta^T}\right), \ K_\alpha = E\left(\frac{\partial l_{\alpha,t}(\theta_0)}{\partial \theta} \frac{\partial l_{\alpha,t}(\theta_0)}{\partial \theta^T}\right)
$$

and $l_{\alpha,t}(\theta)$ *is defined by substituting* $\eta_t(\theta)$ *for* $\tilde{\eta}_t(\theta)$ *in* (3).

Remark 1. *In our empirical study, discussed in Section 3.2, we select an optimal α using the method of Warwick [30] and Warwick and Jones [31]. We choose α that minimizes the trace of the estimated asymptotic mean squared error ($\widehat{AMSE}$):*

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

where $\hat{\theta}_{1,n}$ is the MDPDE with $\alpha = 1$ and $\widehat{As.var}(\hat{\theta}_{\alpha,n})$ is the estimate of the asymptotic variance of $\hat{\theta}_{\alpha,n}$, computed as

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

Remark 2. *Instead of* **(A6)**, *Kim and Lee* [22] *assumed*

$$\sup_{\theta \in \Theta} \sup_{0 \le \delta \le 1} \left| \frac{B''((1-\delta)\eta_t(\theta) + \delta \tilde{\eta}_t(\theta))}{B'((1-\delta)\eta_t(\theta) + \delta \tilde{\eta}_t(\theta))^3} \right| \le K \text{ for some } K > 0$$

to prove Proposition 1. *Note that this condition is satisfied directly if* **(A3)** *and* **(A6)** *hold. In our study, we alter the above condition to* **(A6)** *to prove Lemma* A1 *in the Appendix* A, *which is needed to obtain the limiting null distribution of the DPD-based change point test in Section* 2.2.

The following INGARCH(1,1) models are typical examples of general integer-valued time series models:

$$Y_t | \mathcal{F}_{t-1} \sim p(y|\eta_t), \quad X_t = d + a X_{t-1} + b Y_{t-1},$$

where $X_t = B(\eta_t) = E(Y_t|\mathcal{F}_{t-1})$, $\theta = (d, a, b)^T \in \Theta \subset (0, \infty) \times [0, \infty)^2$ with $a + b < 1$, and Θ is compact. Condition **(A0)** trivially holds, and the process $\{(X_t, Y_t), \ t \ge 1\}$ has a strictly stationary and ergodic solution. Condition **(A1)** can be replaced with the following:

(A1)′ The true parameter θ_0 lies in a compact neighborhood $\Theta \in \mathbb{R}_+^3$ of θ_0, where

$$\Theta \in \{\theta = (d, a, b)^T \in \mathbb{R}_+^3 : 0 < d_L \le d \le d_U, \ \epsilon \le a + b \le 1 - \epsilon\} \text{ for some } d_L, d_U, \epsilon > 0.$$

Moreover, we can express

$$X_t(\theta) = \frac{d}{1-a} + b \sum_{k=0}^{\infty} a^k Y_{t-k-1} \quad \text{and} \quad \widetilde{X}_t(\theta) = \frac{d}{1-a} + b \sum_{k=0}^{t-2} a^k Y_{t-k-1},$$

where the initial value $\widetilde{X}_1$ is taken as $d/(1-a)$ for simplicity. Based on the above and **(A4)**, the conditions **(A2)**, **(A5)**, and **(A7)**–**(A10)** are all satisfied for INGARCH(1,1) models, as proven by Theorem 3 of Kang and Lee [15]. Kim and Lee [22] showed recently that the following Poisson and negative binomial INGARCH(1,1) models satisfy **(A3)** and **(A4)**. Furthermore, following the arguments presented in Section 3.2 of their study, **(A6)** holds for these models as well. Below, we show that **(A11)** holds for Poisson and negative binomial INGARCH(1,1) models.

- *Poisson INGARCH(1,1) model:*

$$Y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(X_t), \quad X_t = d + a X_{t-1} + b Y_{t-1}.$$

In this model, $\eta_t(\theta) = \log(X_t(\theta))$ and $A(\eta_t(\theta)) = e^{\eta_t(\theta)}$. Since $B'(\eta) = B^{(3)}(\eta)$, **(A11)** holds owing to **(A3)**.

- *NB-INGARCH(1,1) model:*

$$Y_t | \mathcal{F}_{t-1} \sim \text{NB}(r, p_t), \quad X_t = \frac{r(1-p_t)}{p_t} = d + a X_{t-1} + b Y_{t-1},$$

where $\text{NB}(r, p)$ denotes a negative binomial distribution with parameters $r \in \mathbb{N}$ and $p \in (0, 1)$. To be more specific, it counts the number of failures before the r-th success occurs in a sequence of Bernoulli trials with success probability p. Here, r is assumed to be known. In this model, $\eta_t(\theta) = \log(X_t(\theta)/(X_t(\theta) + r))$ and $A(\eta_t(\theta)) = r \log(r/(1 - e^{\eta_t(\theta)}))$. From the fact that $B'(\eta) = re^{\eta}/(1 - e^{\eta})^2$ and $B^{(3)}(\eta) = re^{\eta}(e^{2\eta} +$

$4e^{\eta} + 1)/(1 - e^{\eta})^4$, we have $B^{(3)}(\eta)/B'(\eta)^4 = (1 - e^{\eta})^4(e^{2\eta} + 4e^{\eta} + 1)/r^3 e^{3\eta}$, which is positive and strictly decreasing on $\eta < 0$. Moreover, since $d_L/(d_L + r) \leq e^{\eta_t(\theta)} < 1$, it holds that

$$\frac{B^{(3)}(\eta_t(\theta))}{B'(\eta_t(\theta))^4} \leq \frac{6(1 - d_L/(d_L + r))^4}{r^3(d_L/(d_L + r))^3} = \frac{6r}{d_L^3(d_L + r)}$$

and $B^{(3)}(\tilde{\eta}_t(\theta))/B'(\tilde{\eta}_t(\theta))^4$ also has the same upper bound. Hence, **(A11)** is satisfied.

In addition to the above models, general integer-valued time series models also include nonlinear models, such as the integer-valued threshold GARCH (INTGARCH) model:

$$Y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(X_t), \quad X_t = d + aX_{t-1} + b_1 \max(Y_{t-1} - l, 0) + b_2 \min(Y_{t-1}, l),$$

where $\theta = (d, a, b_1, b_2)^T \in \Theta \subset (0, \infty) \times [0, \infty)^3$ with $a + \max(b_1, b_2) < 1$, Θ is compact, and l is a non-negative integer value. For more details, see Remark 3 in Kim and Lee [22].

2.2. DPD-Based Change Point Test

As a robust test for parameter changes in general integer-valued time series models, we propose a DPD-based test for the following hypotheses:

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

To construct the test, we employ the objective function of the MDPDE. That is, our test is constructed using the empirical version of the DPD. Let $\widetilde{L}_{\alpha,n}$ be that in (2). To implement our test, we employ the following test statistic:

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

where

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

is a consistent estimator of K_{α}. For the consistency of $\widehat{K}_{\alpha}$, see Lemma A5 in Appendix A.

Using the mean value theorem (MVT), we have the following, for each $s \in [0, 1]$,

$$\frac{[ns]}{\sqrt{n}} \frac{\partial \widetilde{L}_{\alpha,[ns]}(\hat{\theta}_{\alpha,n})}{\partial \theta} = \frac{[ns]}{\sqrt{n}} \frac{\partial \widetilde{L}_{\alpha,[ns]}(\theta_0)}{\partial \theta} + \frac{[ns]}{n} \frac{\partial^2 \widetilde{L}_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial \theta \partial \theta^T} \sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0), \tag{4}$$

where $\theta^*_{\alpha,n,s}$ is an intermediate point between $\hat{\theta}_{\alpha,n}$ and θ_0. From $\partial \widetilde{L}_{\alpha,n}(\hat{\theta}_{\alpha,n})/\partial \theta = 0$, we obtain that, for $s = 1$,

$$0 = \sqrt{n} \frac{\partial \widetilde{L}_{\alpha,n}(\theta_0)}{\partial \theta} + \frac{\partial^2 \widetilde{L}_{\alpha,n}(\theta^*_{\alpha,n,1})}{\partial \theta \partial \theta^T} \sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0).$$

Furthermore, since J_{α} is nonsingular (cf. proof of Lemma 7 in Kim and Lee [22]), this can be expressed as

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0) &= J_{\alpha}^{-1} \sqrt{n} \frac{\partial \widetilde{L}_{\alpha,n}(\theta_0)}{\partial \theta} + J_{\alpha}^{-1} \frac{\partial^2 \widetilde{L}_{\alpha,n}(\theta^*_{\alpha,n,1})}{\partial \theta \partial \theta^T} \sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0) + \sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0) \\
&= J_{\alpha}^{-1} \sqrt{n} \frac{\partial \widetilde{L}_{\alpha,n}(\theta_0)}{\partial \theta} + J_{\alpha}^{-1} \left(\frac{\partial^2 \widetilde{L}_{\alpha,n}(\theta^*_{\alpha,n,1})}{\partial \theta \partial \theta^T} + J_{\alpha} \right) \sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0).
\end{aligned}$$

Substituting the above into (4) yields

$$
\begin{aligned}
\frac{[ns]}{\sqrt{n}} \frac{\partial \widetilde{L}_{\alpha,[ns]}(\hat{\theta}_{\alpha,n})}{\partial \theta} &= \frac{[ns]}{\sqrt{n}} \frac{\partial \widetilde{L}_{\alpha,[ns]}(\theta_0)}{\partial \theta} + \frac{[ns]}{n} \frac{\partial^2 \widetilde{L}_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial \theta \partial \theta^T} J_\alpha^{-1} \sqrt{n} \frac{\partial \widetilde{L}_{\alpha,n}(\theta_0)}{\partial \theta} \\
&\quad + \frac{[ns]}{n} \frac{\partial^2 \widetilde{L}_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial \theta \partial \theta^T} J_\alpha^{-1} \left(\frac{\partial^2 \widetilde{L}_{\alpha,n}(\theta^*_{\alpha,n,1})}{\partial \theta \partial \theta^T} + J_\alpha \right) \sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0).
\end{aligned}
\tag{5}
$$

In Appendix A, we show that the first two terms on the right-hand side of (5) converge weakly to $K_\alpha^{1/2} B_d^o(s)$, where B_d^o is a d-dimensional standard Brownian bridge and the last term is asymptotically negligible. Therefore, we obtain the following theorem.

Theorem 1. *Suppose that conditions* **(A0)–(A11)** *hold. Then, under H_0, we have*

$$
K_\alpha^{-1/2} \frac{[ns]}{\sqrt{n}} \frac{\partial \widetilde{L}_{\alpha,[ns]}(\hat{\theta}_{\alpha,n})}{\partial \theta} \xrightarrow{w} B_d^o(s).
$$

Therefore,

$$
\widehat{T}_n^\alpha \xrightarrow{d} \sup_{0 \le s \le 1} \| B_d^o(s) \|^2.
$$

We reject H_0 if $\widehat{T}_n^\alpha$ is large; see Table 1 of Lee et al. [32] for the critical values. When a change point is detected, its location is estimated as

$$
\underset{1 \le k \le n}{\operatorname{argmax}} \frac{k^2}{n} \frac{\partial \widetilde{L}_{\alpha,k}(\hat{\theta}_{\alpha,n})}{\partial \theta^T} \widehat{K}_\alpha^{-1} \frac{\partial \widetilde{L}_{\alpha,k}(\hat{\theta}_{\alpha,n})}{\partial \theta}.
$$

Remark 3. *The proposed test $\widehat{T}_n^\alpha$ with $\alpha = 0$ is the same as the score-vector-based CUSUM test proposed by Lee and Lee [14], given by*

$$
\widehat{T}_n^{score} = \max_{1 \le k \le n} \frac{1}{n} \left(\sum_{t=1}^k \frac{\partial \tilde{l}_{0,t}(\hat{\theta}_{0,n})}{\partial \theta^T} \right) \widehat{I}_n^{-1} \left(\sum_{t=1}^k \frac{\partial \tilde{l}_{0,t}(\hat{\theta}_{0,n})}{\partial \theta} \right),
$$

where $\tilde{l}_{0,t}(\theta)$ is defined in (3), $\hat{\theta}_{0,n}$ is the CMLE, and $\widehat{I}_n = n^{-1} \sum_{t=1}^n \partial^2 \tilde{l}_{0,t}(\hat{\theta}_{0,n})/\partial \theta \partial \theta^T$. In the next section, we compare the performance of $\widehat{T}_n^\alpha$ with that of $\widehat{T}_n^{score}$ in the presence of outliers.

3. Empirical Studies

3.1. Simulation

In this section, we evaluate the performance of the proposed test $\widehat{T}_n^\alpha$ (with $\alpha > 0$) through simulations, focusing on the comparison with $\widehat{T}_n^{score}$. First, we consider the Poisson INGARCH models:

$$
Y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(X_t), \quad X_t = d + a X_{t-1} + b Y_{t-1},
\tag{6}
$$

where X_1 is set to 0 for the data generation and $\widetilde{X}_1$ is set as the sample mean of the data. The sample sizes considered are $n = 500$ and 1000, with 1000 repetitions for each simulation. For the comparison, we examine the empirical size and power at the nominal level of 0.05, which has a corresponding critical value of 3.004. To calculate the empirical size and power for each test, we consider cases with $\theta = (d, a, b) = (1, 0.2, 0.2)$, $(1, 0.2, 0.4)$, $(1, 0.2, 0.7)$ and those in which $\theta = (d, a, b) = (1, 0.2, 0.2)$ changes to $\theta' = (d', a', b') = (1.5, 0.2, 0.2)$, $(1, 0.4, 0.2)$, $(1, 0.2, 0.4)$ at the middle time $t = [n/2]$, respectively.

Table 1 presents the results when the data are not contaminated by outliers, showing that both tests ($\widehat{T}_n^{score}$ and $\widehat{T}_n^{\alpha}$) exhibit reasonable size, even when $a + b$ is close to 1. When $n = 500$, $\widehat{T}_n^{score}$ outperforms $\widehat{T}_n^{\alpha}$ in terms of power; however, as the sample size increases to $n = 1000$, $\widehat{T}_n^{\alpha}$ exhibits similar power to that of $\widehat{T}_n^{score}$, particularly when α is small. The power of $\widehat{T}_n^{\alpha}$ tends to decrease as α increases, confirming that an MDPDE with large α results in a loss of efficiency.

Table 1. Empirical sizes and powers for Poisson integer-valued generalized autoregressive conditional heteroscedastic (INGARCH)(1,1) models when no outliers exist.

				$\widehat{T}_n^{\alpha}$ with α				
	$\theta = (d, a, b)$	n	$\widehat{T}_n^{score}$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
	(1, 0.2, 0.2)	500	0.084	0.053	0.059	0.059	0.058	0.059
		1000	0.065	0.047	0.053	0.053	0.051	0.059
Sizes	(1, 0.2, 0.4)	500	0.049	0.040	0.043	0.045	0.047	0.047
		1000	0.033	0.039	0.045	0.047	0.050	0.053
	(1, 0.2, 0.7)	500	0.031	0.028	0.030	0.029	0.029	0.034
		1000	0.050	0.051	0.047	0.044	0.046	0.051
	$\theta' = (d', a', b')$	n	$\theta = (d, a, b) = (1, 0.2, 0.2)$ changes to $\theta' = (d', a', b')$					
	(1.5, 0.2, 0.2)	500	0.836	0.776	0.764	0.741	0.687	0.525
		1000	0.912	0.914	0.911	0.910	0.901	0.871
Powers	(1, 0.4, 0.2)	500	0.782	0.704	0.695	0.661	0.591	0.454
		1000	0.951	0.942	0.939	0.937	0.917	0.886
	(1, 0.2, 0.4)	500	0.819	0.804	0.800	0.795	0.736	0.634
		1000	0.996	0.996	0.996	0.993	0.991	0.978

To evaluate the robustness of the proposed test, we assume that contaminated data $Y_{c,t}$ are observed instead of Y_t in (6) (cf. Fried et al. [33]):

$$Y_{c,t} = Y_t + P_t Y_{o,t}, \tag{7}$$

where P_t are independent and identically distributed (iid) Bernoulli random variables with success probability p and $Y_{o,t}$ are iid Poisson random variables with mean γ. We assume that Y_t, P_t, and $Y_{o,t}$ are all independent. In this simulation, we consider the cases $p = 0.01$, 0.03 and $\gamma = 5$, 10. The results are reported in Tables 2–5, showing that $\widehat{T}_n^{score}$ suffers from size distortions that become more severe as either p or γ increase. In contrast, $\widehat{T}_n^{\alpha}$ compensates for this defect remarkably well, yielding comparable power to that of $\widehat{T}_n^{score}$ when $n = 1000$. This indicates that as more data are contaminated by outliers, $\widehat{T}_n^{\alpha}$ increasingly outperforms $\widehat{T}_n^{score}$.

Table 2. Empirical sizes and powers for Poisson INGARCH(1,1) models when $p = 0.01$ and $\gamma = 5$.

| | $\theta = (d, a, b)$ | n | $\widehat{T}_n^{score}$ | $\widehat{T}_n^\alpha$ with α | | | | |
				$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
	(1, 0.2, 0.2)	500	0.108	0.048	0.046	0.046	0.052	0.057
		1000	0.110	0.048	0.044	0.041	0.050	0.053
Sizes	(1, 0.2, 0.4)	500	0.070	0.041	0.041	0.046	0.046	0.049
		1000	0.078	0.041	0.042	0.041	0.045	0.043
	(1, 0.2, 0.7)	500	0.057	0.035	0.039	0.038	0.045	0.045
		1000	0.061	0.041	0.042	0.045	0.044	0.049
	$\theta' = (d', a', b')$	n	$\theta = (d, a, b) = (1, 0.2, 0.2)$ changes to $\theta' = (d', a', b')$					
	(1.5, 0.2, 0.2)	500	0.792	0.736	0.735	0.723	0.676	0.569
		1000	0.901	0.898	0.903	0.903	0.896	0.856
Powers	(1, 0.4, 0.2)	500	0.766	0.684	0.686	0.667	0.626	0.525
		1000	0.944	0.934	0.935	0.931	0.915	0.864
	(1, 0.2, 0.4)	500	0.871	0.806	0.804	0.787	0.752	0.647
		1000	0.997	0.993	0.993	0.992	0.990	0.960

Table 3. Empirical sizes and powers for Poisson INGARCH(1,1) models when $p = 0.01$ and $\gamma = 10$.

| | $\theta = (d, a, b)$ | n | $\widehat{T}_n^{score}$ | $\widehat{T}_n^\alpha$ with α | | | | |
				$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
	(1, 0.2, 0.2)	500	0.246	0.069	0.075	0.070	0.069	0.079
		1000	0.317	0.071	0.062	0.070	0.070	0.062
Sizes	(1, 0.2, 0.4)	500	0.234	0.053	0.060	0.061	0.052	0.051
		1000	0.262	0.059	0.070	0.072	0.071	0.060
	(1, 0.2, 0.7)	500	0.127	0.040	0.040	0.037	0.041	0.038
		1000	0.115	0.045	0.044	0.049	0.048	0.050
	$\theta' = (d', a', b')$	n	$\theta = (d, a, b) = (1, 0.2, 0.2)$ changes to $\theta' = (d', a', b')$					
	(1.5, 0.2, 0.2)	500	0.840	0.785	0.791	0.769	0.742	0.649
		1000	0.874	0.863	0.874	0.869	0.868	0.862
Powers	(1, 0.4, 0.2)	500	0.835	0.743	0.759	0.740	0.694	0.590
		1000	0.911	0.910	0.913	0.908	0.902	0.879
	(1, 0.2, 0.4)	500	0.920	0.829	0.835	0.831	0.787	0.697
		1000	0.997	0.992	0.995	0.997	0.994	0.965

Table 4. Empirical sizes and powers for Poisson INGARCH(1,1) models when $p = 0.03$ and $\gamma = 5$.

| | $\theta = (d, a, b)$ | n | $\widehat{T}_n^{score}$ | $\widehat{T}_n^\alpha$ with α | | | | |
				$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
	(1, 0.2, 0.2)	500	0.213	0.060	0.059	0.058	0.062	0.074
		1000	0.229	0.052	0.055	0.063	0.062	0.061
Sizes	(1, 0.2, 0.4)	500	0.176	0.052	0.057	0.064	0.066	0.060
		1000	0.173	0.047	0.055	0.054	0.055	0.059
	(1, 0.2, 0.7)	500	0.073	0.030	0.039	0.037	0.037	0.045
		1000	0.086	0.039	0.035	0.040	0.042	0.039
	$\theta' = (d', a', b')$	n	$\theta = (d, a, b) = (1, 0.2, 0.2)$ changes to $\theta' = (d', a', b')$					
	(1.5, 0.2, 0.2)	500	0.804	0.693	0.715	0.709	0.687	0.616
		1000	0.867	0.859	0.867	0.867	0.859	0.847
Powers	(1, 0.4, 0.2)	500	0.786	0.662	0.693	0.681	0.634	0.561
		1000	0.908	0.896	0.903	0.899	0.893	0.868
	(1, 0.2, 0.4)	500	0.915	0.787	0.797	0.792	0.773	0.672
		1000	0.998	0.994	0.995	0.993	0.986	0.965

Table 5. Empirical sizes and powers for Poisson INGARCH(1,1) models when $p = 0.03$ and $\gamma = 10$.

| | $\theta = (d, a, b)$ | n | $\widehat{T}_n^{score}$ | $\widehat{T}_n^\alpha$ with α | | | | |
				$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
	(1, 0.2, 0.2)	500	0.475	0.083	0.082	0.083	0.091	0.102
		1000	0.592	0.092	0.097	0.104	0.109	0.097
Sizes	(1, 0.2, 0.4)	500	0.556	0.071	0.078	0.080	0.068	0.065
		1000	0.621	0.092	0.113	0.115	0.108	0.071
	(1, 0.2, 0.7)	500	0.296	0.050	0.056	0.056	0.053	0.040
		1000	0.289	0.060	0.062	0.057	0.060	0.055
	$\theta' = (d', a', b')$	n	$\theta = (d, a, b) = (1, 0.2, 0.2)$ changes to $\theta' = (d', a', b')$					
	(1.5, 0.2, 0.2)	500	0.834	0.760	0.800	0.801	0.782	0.719
		1000	0.889	0.821	0.852	0.867	0.860	0.866
Powers	(1, 0.4, 0.2)	500	0.850	0.738	0.783	0.786	0.759	0.688
		1000	0.897	0.848	0.887	0.889	0.895	0.880
	(1, 0.2, 0.4)	500	0.951	0.817	0.847	0.842	0.815	0.728
		1000	0.997	0.991	0.992	0.992	0.983	0.969

Next, we consider the following NB-INGARCH(1,1) models:

$$Y_t | \mathcal{F}_{t-1} \sim \mathrm{NB}(r, p_t), \quad X_t = \frac{r(1 - p_t)}{p_t} = d + aX_{t-1} + bY_{t-1}, \tag{8}$$

where X_1 and $\widetilde{X}_1$ are 0 and the sample mean of the data, respectively. We set $r = 10$, and use the same parameter settings as in the Poisson INGARCH model case. In order to evaluate the robustness of the test, we observe contaminated data $Y_{c,t}$, as in (7), where Y_t are generated from (8), P_t are iid Bernoulli random variables with success probability p, and $Y_{o,t}$ are iid NB$(10, \kappa)$ random variables. We consider the cases $p = 0.01$, 0.03 and $\kappa = 0.6$, 0.5. The results are reported in Tables 6–10, showing similar results to those in Tables 1–5. Our findings show that the DPD-based test performs reasonably well in terms of both size and power, regardless of the existence of outliers. In addition, we confirm that the proposed test outperforms the score-based CUSUM test when the data are contaminated by outliers.

Table 6. Empirical sizes and powers for negative binomial INGARCH (NB-INGARCH)(1,1) models when no outliers exist.

| | $\theta = (d, a, b)$ | n | $\widehat{T}_n^{score}$ | $\widehat{T}_n^\alpha$ with α | | | | |
				$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
	(1, 0.2, 0.2)	500	0.076	0.050	0.052	0.054	0.061	0.071
		1000	0.061	0.055	0.052	0.052	0.055	0.059
Sizes	(1, 0.2, 0.4)	500	0.040	0.041	0.038	0.040	0.045	0.048
		1000	0.049	0.053	0.056	0.057	0.062	0.060
	(1, 0.2, 0.7)	500	0.047	0.046	0.043	0.038	0.042	0.043
		1000	0.041	0.044	0.048	0.048	0.047	0.043
	$\theta' = (d', a', b')$	n	$\theta = (d, a, b) = (1, 0.2, 0.2)$ changes to $\theta' = (d', a', b')$					
	(1.5, 0.2, 0.2)	500	0.821	0.759	0.735	0.706	0.640	0.505
		1000	0.953	0.942	0.936	0.932	0.919	0.881
Powers	(1, 0.4, 0.2)	500	0.759	0.689	0.646	0.611	0.558	0.454
		1000	0.967	0.964	0.959	0.955	0.940	0.881
	(1, 0.2, 0.4)	500	0.733	0.719	0.718	0.702	0.650	0.544
		1000	0.984	0.984	0.981	0.975	0.961	0.908

Table 7. Empirical sizes and powers for NB-INGARCH(1,1) models when $p = 0.01$ and $\kappa = 0.6$.

| | $\theta = (d, a, b)$ | n | $\widehat{T}_n^{score}$ | $\widehat{T}_n^\alpha$ with α | | | | |
				$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
	(1, 0.2, 0.2)	500	0.158	0.062	0.066	0.066	0.071	0.071
		1000	0.173	0.069	0.066	0.067	0.068	0.061
Sizes	(1, 0.2, 0.4)	500	0.105	0.045	0.045	0.049	0.047	0.039
		1000	0.112	0.058	0.058	0.062	0.057	0.047
	(1, 0.2, 0.7)	500	0.045	0.031	0.035	0.038	0.041	0.038
		1000	0.065	0.042	0.045	0.044	0.041	0.045
	$\theta' = (d', a', b')$	n	$\theta = (d, a, b) = (1, 0.2, 0.2)$ changes to $\theta' = (d', a', b')$					
	(1.5, 0.2, 0.2)	500	0.803	0.705	0.714	0.695	0.647	0.516
		1000	0.945	0.931	0.931	0.930	0.921	0.909
Powers	(1, 0.4, 0.2)	500	0.757	0.648	0.645	0.626	0.579	0.464
		1000	0.959	0.958	0.952	0.947	0.930	0.895
	(1, 0.2, 0.4)	500	0.807	0.704	0.716	0.710	0.659	0.574
		1000	0.985	0.978	0.980	0.979	0.969	0.935

Table 8. Empirical sizes and powers for NB-INGARCH(1,1) models when $p = 0.01$ and $\kappa = 0.5$.

| | $\theta = (d, a, b)$ | n | $\widehat{T}_n^{score}$ | $\widehat{T}_n^\alpha$ with α | | | | |
				$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
	(1, 0.2, 0.2)	500	0.258	0.069	0.069	0.070	0.076	0.080
		1000	0.292	0.061	0.061	0.057	0.058	0.068
Sizes	(1, 0.2, 0.4)	500	0.177	0.048	0.048	0.052	0.057	0.058
		1000	0.236	0.072	0.079	0.081	0.073	0.074
	(1, 0.2, 0.7)	500	0.095	0.048	0.054	0.058	0.060	0.055
		1000	0.097	0.049	0.050	0.050	0.050	0.051
	$\theta' = (d', a', b')$	n	$\theta = (d, a, b) = (1, 0.2, 0.2)$ changes to $\theta' = (d', a', b')$					
	(1.5, 0.2, 0.2)	500	0.840	0.771	0.768	0.740	0.688	0.599
		1000	0.923	0.924	0.932	0.926	0.925	0.897
Powers	(1, 0.4, 0.2)	500	0.808	0.704	0.709	0.673	0.634	0.536
		1000	0.938	0.946	0.946	0.943	0.935	0.898
	(1, 0.2, 0.4)	500	0.842	0.723	0.740	0.735	0.696	0.586
		1000	0.997	0.989	0.984	0.977	0.972	0.923

Table 9. Empirical sizes and powers for NB-INGARCH(1,1) models when $p = 0.03$ and $\kappa = 0.6$.

| | $\theta = (d, a, b)$ | n | $\widehat{T}_n^{score}$ | $\widehat{T}_n^\alpha$ with α | | | | |
				$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
	(1, 0.2, 0.2)	500	0.289	0.079	0.077	0.076	0.086	0.069
		1000	0.328	0.060	0.068	0.068	0.077	0.075
Sizes	(1, 0.2, 0.4)	500	0.228	0.051	0.054	0.051	0.052	0.047
		1000	0.246	0.054	0.064	0.066	0.064	0.059
	(1, 0.2, 0.7)	500	0.090	0.035	0.040	0.040	0.044	0.036
		1000	0.108	0.058	0.053	0.052	0.050	0.040
	$\theta' = (d', a', b')$	n	$\theta = (d, a, b) = (1, 0.2, 0.2)$ changes to $\theta' = (d', a', b')$					
	(1.5, 0.2, 0.2)	500	0.818	0.685	0.705	0.702	0.675	0.582
		1000	0.925	0.892	0.900	0.899	0.905	0.909
Powers	(1, 0.4, 0.2)	500	0.806	0.637	0.666	0.664	0.627	0.522
		1000	0.938	0.927	0.926	0.922	0.913	0.896
	(1, 0.2, 0.4)	500	0.870	0.690	0.734	0.731	0.704	0.604
		1000	0.990	0.976	0.978	0.974	0.969	0.931

Table 10. Empirical sizes and powers for NB-INGARCH(1,1) models when $p = 0.03$ and $\kappa = 0.5$.

| | $\theta = (d, a, b)$ | n | $\widehat{T}_n^{score}$ | $\widehat{T}_n^{\alpha}$ with α | | | | |
				$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1$
	(1, 0.2, 0.2)	500	0.469	0.085	0.088	0.100	0.102	0.096
		1000	0.563	0.075	0.088	0.097	0.105	0.100
Sizes	(1, 0.2, 0.4)	500	0.506	0.068	0.071	0.076	0.081	0.072
		1000	0.532	0.089	0.096	0.101	0.089	0.078
	(1, 0.2, 0.7)	500	0.188	0.054	0.066	0.072	0.066	0.061
		1000	0.207	0.053	0.051	0.064	0.069	0.059
	$\theta' = (d', a', b')$	n	$\theta = (d, a, b) = (1, 0.2, 0.2)$ changes to $\theta' = (d', a', b')$					
	(1.5, 0.2, 0.2)	500	0.879	0.749	0.784	0.797	0.758	0.687
		1000	0.930	0.880	0.889	0.893	0.889	0.886
Powers	(1, 0.4, 0.2)	500	0.867	0.698	0.766	0.756	0.734	0.636
		1000	0.948	0.891	0.900	0.906	0.906	0.889
	(1, 0.2, 0.4)	500	0.927	0.735	0.770	0.770	0.743	0.639
		1000	0.995	0.977	0.984	0.981	0.971	0.944

3.2. Real Data Analysis

In this section, we demonstrate the validity of $\widehat{T}_n^{\alpha}$ using a real data analysis. To this end, we analyze the return times of extreme events related to GS stock, which are constructed based on the daily log-returns for the period of 5 May 1999 to 15 March 2012. Davis and Liu [12] and Kim and Lee [22] previously investigated this data set in their works on geometric INGARCH(1,1) models (i.e., NB-INGARCH(1,1) models with $r = 1$).

We first compute the hitting times, $\tau_1, \tau_2, \ldots$, for which the log-returns of GS stock fall outside the 0.05 and 0.95 quantiles of the data. The return times of these extreme events are calculated as $Y_t = \tau_t - \tau_{t-1}$. Figure 1 plots Y_t, $t = 1, \ldots, 323$. The figure shows that the data include large observations; for example, a sample variance of 1106 with a sample mean of 10.01 indicates the existence of aberrant observations.
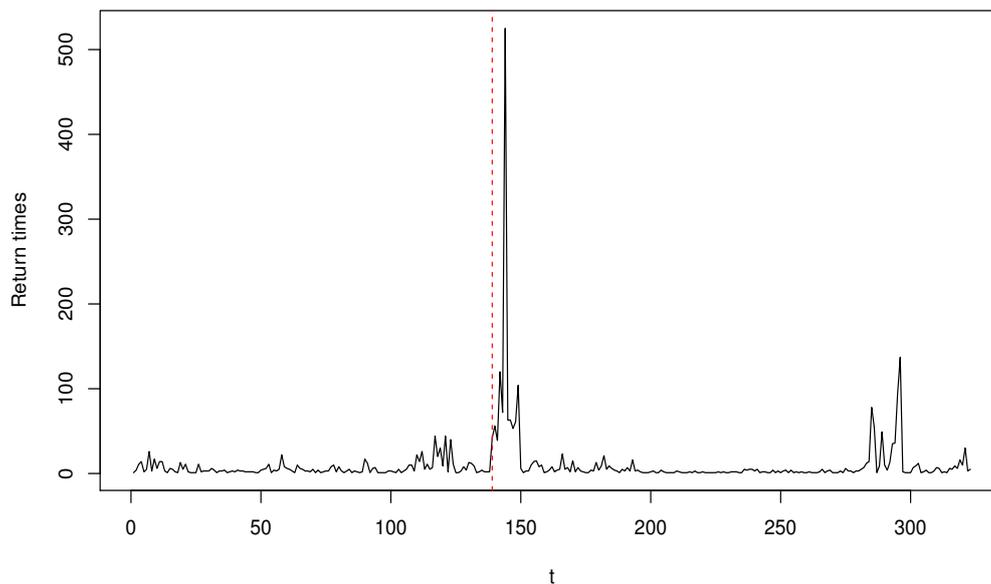
Figure 1. Plot of the return times of extreme events for Goldman Sachs Group (GS) stock.

Since $Y_t \geq 1$, we consider a geometric distribution that counts the total number of trials, rather than the number of failures, to fit the following geometric INGARCH(1,1) models to the data:

$$Y_t | \mathcal{F}_{t-1} \sim \text{Geo}(p_t), \quad X_t = \frac{1}{p_t} = d + aX_{t-1} + bY_{t-1},$$

where $\widetilde{X}_1$ is set as the sample mean of the data. Kim and Lee [22] showed that the optimal α for the MDPDE is 0.25, using the criterion provided in Remark 1. The results for the parameter estimation are summarized in Table 11 for $\alpha = 0$ (CMLE) and 0.25 (MDPDE with optimal α); figures in parentheses denote the standard errors of the corresponding estimates. We observe that, compared with the CMLE, the MDPDE with $\alpha = 0.25$ is quite different and has smaller standard errors.

Table 11. Parameter estimates for geometric INGARCH(1,1) models.

α	$\hat{d}$	$\hat{a}$	$\hat{b}$	$\widehat{\text{AMSE}}$
0(CMLE)	0.526(0.406)	0.490(0.175)	0.483(0.156)	0.623
0.25	0.432(0.242)	0.518(0.129)	0.418(0.115)	0.398

Next, we use $\widehat{T}_n^{score}$ and $\widehat{T}_n^{0.25}$ ($\widehat{T}_n^{\alpha}$ with $\alpha = 0.25$) to perform a parameter change test at the nominal level of 0.05 (the corresponding critical value is 3.004). Let $\widehat{T}_n^{score} = \max_{1 \le k \le n} SCORE_{k,n}$ and $\widehat{T}_n^{0.25} = \max_{1 \le k \le n} DPD_{k,n}$. The left and right panels of Figure 2 display $SCORE_{k,n}$ and $DPD_{k,n}$, respectively. For most k, $DPD_{k,n}$ appears to be smaller than $SCORE_{k,n}$, which is definitely attributed to the robustness of the MDPDE and DPD. We obtain $\widehat{T}_n^{score} = 5.136$, which suggests the existence of a parameter change. In Figures 1 and 2, the red, vertical, dashed line represents the location of a change when $\widehat{T}_n^{score}$ is applied. However, this result is not so reliable because $\widehat{T}_n^{score}$ can signal a change point affected by outliers as seen in the previous section, and the change point is truly detected at the occurrence time of an outlier in this case. In contrast, $\widehat{T}_n^{0.25}$ yields a value of 1.219, indicating that no change point exists. This result clearly demonstrates that outliers can severely affect parameter estimates and change point tests by mistakenly identifying a change point. Our findings confirm that the DPD-based change point test provides a functional and robust alternative to the score-based CUSUM test in the presence of outliers.
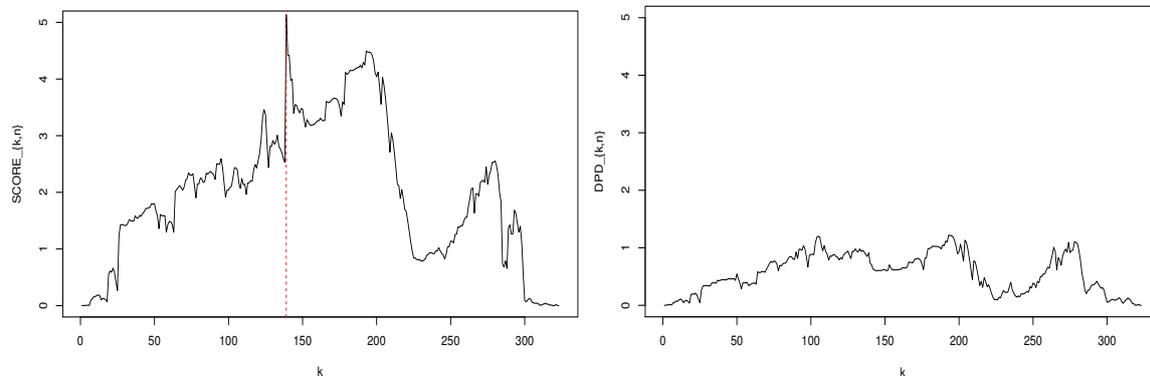
Figure 2. Plots of $SCORE_{k,n}$ and $DPD_{k,n}$.

4. Conclusions

In this study, we developed a DPD-based robust change point test for general integer-valued time series models with a conditional distribution that belongs to the one-parameter exponential family. We provided regularity conditions under which the proposed test converges weakly to the function of a Brownian bridge. The simulation study showed that the DPD-based test produces reasonable sizes and powers regardless of the existence of outliers, whereas the score-based CUSUM test suffers from severe size distortions when the data are contaminated by outliers. In the real data analysis using the return times of extreme events related to GS stock, the score-based CUSUM test supported the existence a parameter change, due to the influence of outliers, while the DPD-based test did not detect a change point because of its robust property. This result confirms the validity of the proposed test as a robust test in practice. It is noteworthy that the DPD-based test can be feasibly extended to other parametric models as far as the asymptotic properties of the MDPDE for the models are validated. We leave the issue of extension to other models as our future study.

Appendix A

In this appendix, we prove Theorem 1 for $\alpha > 0$; refer to Lee and Lee [14] for the case of $\alpha = 0$. The following properties of the probability mass function of the non-negative integer-valued exponential family are useful for proving Lemma A1. For all $y \in \mathbb{N}_0$ and $\eta \in \mathbb{R}$:

(E1) $0 < p(y|\eta) < 1$,
(E2) $\sum_{y=0}^{\infty} p(y|\eta) = 1$,
(E3) $\sum_{y=0}^{\infty} y\, p(y|\eta) = B(\eta)$,
(E4) $\sum_{y=0}^{\infty} y^2 p(y|\eta) = B'(\eta) + B(\eta)^2$,
(E5) $\sum_{y=0}^{\infty} y^3 p(y|\eta) = B''(\eta) + 3B'(\eta)B(\eta) + B(\eta)^3$.

Throughout this section, we denote $L_{\alpha,n}(\theta) = n^{-1} \sum_{t=1}^{n} l_{\alpha,t}(\theta)$ and employ the notation $\eta_t = \eta_t(\theta)$, $\tilde{\eta}_t = \tilde{\eta}_t(\theta)$, and $\eta_t^0 = \eta_t(\theta_0)$ for brevity. Furthermore, if we define two functions $h_\alpha(\eta)$ and $m_\alpha(\eta)$ as

$$
h_\alpha(\eta) = \sum_{y=0}^{\infty} p(y|\eta)^{1+\alpha} \frac{y - B(\eta)}{B'(\eta)} - p(Y_t|\eta)^\alpha \frac{Y_t - B(\eta)}{B'(\eta)},
$$

$$
m_\alpha(\eta) = \sum_{y=0}^{\infty} p(y|\eta)^{1+\alpha} \left[(1+\alpha) \left(\frac{y - B(\eta)}{B'(\eta)} \right)^2 - \frac{B''(\eta)}{B'(\eta)^2} \frac{y - B(\eta)}{B'(\eta)} - \frac{1}{B'(\eta)} \right]
$$

$$
- p(Y_t|\eta)^\alpha \left[\alpha \left(\frac{Y_t - B(\eta)}{B'(\eta)} \right)^2 - \frac{B''(\eta)}{B'(\eta)^2} \frac{Y_t - B(\eta)}{B'(\eta)} - \frac{1}{B'(\eta)} \right],
$$

we obtain

$$
\frac{\partial l_{\alpha,t}(\theta)}{\partial \theta} = (1+\alpha) h_\alpha(\eta_t) \frac{\partial X_t(\theta)}{\partial \theta},
$$

$$
\frac{\partial^2 l_{\alpha,t}(\theta)}{\partial \theta \partial \theta^T} = (1+\alpha) \left(h_\alpha(\eta_t) \frac{\partial^2 X_t(\theta)}{\partial \theta \partial \theta^T} + m_\alpha(\eta_t) \frac{\partial X_t(\theta)}{\partial \theta} \frac{\partial X_t(\theta)}{\partial \theta^T} \right).
$$

Lemma A1. *Suppose that conditions* **(A3)**, **(A6)**, *and* **(A11)** *hold. Then, we have*

$$|h_\alpha(\eta_t)| \leq \frac{1}{\underline{c}}(Y_t + 3X_t(\theta)),$$

$$|h_\alpha(\tilde{\eta}_t)| \leq \frac{1}{\underline{c}}(Y_t + 3X_t(\theta) + 3|X_t(\theta) - \widetilde{X}_t(\theta)|),$$

$$|m_\alpha(\eta_t)| \leq \frac{\alpha}{\underline{c}^2}Y_t^2 + \frac{K}{\underline{c}^{1/2}}Y_t + \frac{\alpha}{\underline{c}^2}X_t(\theta)^2 + \frac{3K}{\underline{c}^{1/2}}X_t(\theta) + \frac{3+\alpha}{\underline{c}},$$

$$|h_\alpha(\eta_t) - h_\alpha(\tilde{\eta}_t)| \leq \left[\frac{\alpha}{\underline{c}^2}Y_t^2 + \frac{K}{\underline{c}^{1/2}}Y_t + \frac{2\alpha}{\underline{c}^2}\left(X_t(\theta)^2 + |X_t(\theta) - \widetilde{X}_t(\theta)|^2\right)\right.$$
$$\left. + \frac{3K}{\underline{c}^{1/2}}\left(X_t(\theta) + |X_t(\theta) - \widetilde{X}_t(\theta)|\right) + \frac{3+\alpha}{\underline{c}}\right]|X_t(\theta) - \widetilde{X}_t(\theta)|,$$

$$|m_\alpha(\tilde{\eta}_t)| \leq \frac{\alpha}{\underline{c}^2}Y_t^2 + \frac{K}{\underline{c}^{1/2}}Y_t + \frac{2\alpha}{\underline{c}^2}\left(X_t(\theta)^2 + |X_t(\theta) - \widetilde{X}_t(\theta)|^2\right)$$
$$+ \frac{3K}{\underline{c}^{1/2}}\left(X_t(\theta) + |X_t(\theta) - \widetilde{X}_t(\theta)|\right) + \frac{3+\alpha}{\underline{c}},$$

$$|m_\alpha(\eta_t) - m_\alpha(\tilde{\eta}_t)| \leq \left[\frac{\alpha^2}{\underline{c}^3}Y_t^3 + \frac{3\alpha K}{\underline{c}^{3/2}}Y_t^2 + \left(\frac{3\alpha}{\underline{c}^2} + M + 3K^2\right)Y_t\right.$$
$$+ \frac{4(3\alpha^2 + 4\alpha + 2)}{\underline{c}^3}\left(X_t(\theta)^3 + |X_t(\theta) - \widetilde{X}_t(\theta)|^3\right)$$
$$+ \frac{6\alpha K}{\underline{c}^{3/2}}\left(X_t(\theta)^2 + |X_t(\theta) - \widetilde{X}_t(\theta)|^2\right)$$
$$+ 3\left(\frac{\alpha^2 + 5\alpha + 3}{\underline{c}^2} + M + 3K^2\right)\left(X_t(\theta) + |X_t(\theta) - \widetilde{X}_t(\theta)|\right)$$
$$\left. + \frac{(\alpha^2 + 5\alpha + 8)K}{\underline{c}^{1/2}}\right]|X_t(\theta) - \widetilde{X}_t(\theta)|.$$

Proof. The proofs for the first four parts of the lemma can be found in Lemma 4 of Kim and Lee [22]. The fifth part is obtained directly from the third part, together with the fact that $\widetilde{X}_t(\theta) \leq |X_t(\theta) - \widetilde{X}_t(\theta)| + X_t(\theta)$.

By the MVT, (E1)–(E5), **(A3)**, **(A6)**, and **(A11)**, we have

$$
|m_\alpha(\eta_t) - m_\alpha(\tilde{\eta}_t)|
$$

$$
= \left| \frac{\partial m_\alpha(B^{-1}(X_t^*(\theta)))}{\partial X_t(\theta)} \right| |X_t(\theta) - \widetilde{X}_t(\theta)|
$$

$$
= \left| \frac{\partial m_\alpha(\eta_t^*)}{\partial \eta_t} \frac{1}{B'(\eta_t^*)} \right| |X_t(\theta) - \widetilde{X}_t(\theta)|
$$

$$
= \frac{1}{B'(\eta_t^*)} \left| \sum_{y=0}^\infty p(y|\eta_t^*)^{1+\alpha} \left[(1+\alpha)^2 \frac{1}{B'(\eta_t^*)^2} (y - B(\eta_t^*))^3 - 3(1+\alpha) \frac{B''(\eta_t^*)}{B'(\eta_t^*)^3} (y - B(\eta_t^*))^2 \right. \right.
$$

$$
\left. + \left(-3(1+\alpha) \frac{1}{B'(\eta_t^*)} - \frac{B^{(3)}(\eta_t^*)}{B'(\eta_t^*)^3} + 3 \frac{B''(\eta_t^*)^2}{B'(\eta_t^*)^4} \right) (y - B(\eta_t^*)) + 2 \frac{B''(\eta_t^*)}{B'(\eta_t^*)^2} \right]
$$

$$
- p(Y_t|\eta_t^*)^\alpha \left[\alpha^2 \frac{1}{B'(\eta_t^*)^2} (Y_t - B(\eta_t^*))^3 - 3\alpha \frac{B''(\eta_t^*)}{B'(\eta_t^*)^3} (Y_t - B(\eta_t^*))^2 \right.
$$

$$
\left. \left. + \left(-3\alpha \frac{1}{B'(\eta_t^*)} - \frac{B^{(3)}(\eta_t^*)}{B'(\eta_t^*)^3} + 3 \frac{B''(\eta_t^*)^2}{B'(\eta_t^*)^4} \right) (Y_t - B(\eta_t^*)) + 2 \frac{B''(\eta_t^*)}{B'(\eta_t^*)^2} \right] \right| |X_t(\theta) - \widetilde{X}_t(\theta)|
$$

$$
\leq \left[(1+\alpha)^2 \frac{1}{B'(\eta_t^*)^3} \left(B''(\eta_t^*) + 3B'(\eta_t^*)B(\eta_t^*) + B(\eta_t^*)^3 + B(\eta_t^*)^3 \right) + 3(1+\alpha) \left| \frac{B''(\eta_t^*)}{B'(\eta_t^*)^3} \right| \right.
$$

$$
+ \left(3(1+\alpha) \frac{1}{B'(\eta_t^*)^2} + \left| \frac{B^{(3)}(\eta_t^*)}{B'(\eta_t^*)^4} \right| + 3 \frac{B''(\eta_t^*)^2}{B'(\eta_t^*)^5} \right) (B(\eta_t^*) + B(\eta_t^*)) + 2 \left| \frac{B''(\eta_t^*)}{B'(\eta_t^*)^3} \right|
$$

$$
+ \alpha^2 \frac{1}{B'(\eta_t^*)^3} (Y_t^3 + B(\eta_t^*)^3) + 3\alpha \left| \frac{B''(\eta_t^*)}{B'(\eta_t^*)^4} \right| (Y_t^2 + B(\eta_t^*)^2)
$$

$$
+ \left. \left(3\alpha \frac{1}{B'(\eta_t^*)^2} + \left| \frac{B^{(3)}(\eta_t^*)}{B'(\eta_t^*)^4} \right| + 3 \frac{B''(\eta_t^*)^2}{B'(\eta_t^*)^5} \right) (Y_t + B(\eta_t^*)) + 2 \left| \frac{B''(\eta_t^*)}{B'(\eta_t^*)^3} \right| \right] |X_t(\theta) - \widetilde{X}_t(\theta)|
$$

$$
\leq \left[\frac{\alpha^2}{\underline{c}^3} Y_t^3 + \frac{3\alpha K}{\underline{c}^{3/2}} Y_t^2 + \left(\frac{3\alpha}{\underline{c}^2} + M + 3K^2 \right) Y_t + \frac{3\alpha^2 + 4\alpha + 2}{\underline{c}^3} B(\eta_t^*)^3 + \frac{3\alpha K}{\underline{c}^{3/2}} B(\eta_t^*)^2 \right.
$$

$$
+ \left. \left(\frac{3\alpha^2 + 15\alpha + 9}{\underline{c}^2} + 3M + 9K^2 \right) B(\eta_t^*) + \frac{(\alpha^2 + 5\alpha + 8)K}{\underline{c}^{1/2}} \right] |X_t(\theta) - \widetilde{X}_t(\theta)|,
$$

where $X_t^*(\theta)$ is an intermediate point between $X_t(\theta)$ and $\widetilde{X}_t(\theta)$, and $\eta_t^* = B^{-1}(X_t^*(\theta))$. Note that since B^{-1} is strictly increasing, η_t^* lies between $B^{-1}(X_t(\theta)) = \eta_t$ and $B^{-1}(\widetilde{X}_t(\theta)) = \tilde{\eta}_t$. Then, because $B(\eta_t^*) \leq B(\eta_t) + |B(\eta_t) - B(\tilde{\eta}_t)|$, the last part of the lemma is established. $\square$

Lemma A2. *Suppose that conditions **(A0)–(A11)** hold. Then, under H_0, we have as $n \to \infty$,*

$$
\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \Theta} \left\| \frac{\partial^2 l_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 \tilde{l}_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} \right\| = o(1) \ \ a.s.
$$

and

$$
\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \Theta} \left\| \frac{\partial l_{\alpha,t}(\theta)}{\partial\theta} \frac{\partial l_{\alpha,t}(\theta)}{\partial\theta^T} - \frac{\partial \tilde{l}_{\alpha,t}(\theta)}{\partial\theta} \frac{\partial \tilde{l}_{\alpha,t}(\theta)}{\partial\theta^T} \right\| = o(1) \ \ a.s.
$$

Proof. It is sufficient to show that as $t \to \infty$,

$$
\sup_{\theta \in \Theta} \left\| \frac{\partial^2 l_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 \tilde{l}_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} \right\| = o(1) \ \ a.s.
$$

and

$$\sup_{\theta\in\Theta}\left\|\frac{\partial l_{\alpha,t}(\theta)}{\partial\theta}\frac{\partial l_{\alpha,t}(\theta)}{\partial\theta^T}-\frac{\partial\tilde{l}_{\alpha,t}(\theta)}{\partial\theta}\frac{\partial\tilde{l}_{\alpha,t}(\theta)}{\partial\theta^T}\right\|=o(1)\ \text{a.s.}$$

Note that we can write

$$\frac{1}{1+\alpha}\sup_{\theta\in\Theta}\left\|\frac{\partial^2 l_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T}-\frac{\partial^2\tilde{l}_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T}\right\|$$

$$\leq\sup_{\theta\in\Theta}\left\|h_\alpha(\tilde{\eta}_t)\left(\frac{\partial^2 X_t(\theta)}{\partial\theta\partial\theta^T}-\frac{\partial^2\widetilde{X}_t(\theta)}{\partial\theta\partial\theta^T}\right)\right\|+\sup_{\theta\in\Theta}\left\|(h_\alpha(\eta_t)-h_\alpha(\tilde{\eta}_t))\frac{\partial^2 X_t(\theta)}{\partial\theta\partial\theta^T}\right\|$$

$$+\sup_{\theta\in\Theta}\left\|(m_\alpha(\eta_t)-m_\alpha(\tilde{\eta}_t))\frac{\partial X_t(\theta)}{\partial\theta}\frac{\partial X_t(\theta)}{\partial\theta^T}\right\|+\sup_{\theta\in\Theta}\left\|m_\alpha(\tilde{\eta}_t)\frac{\partial X_t(\theta)}{\partial\theta}\left(\frac{\partial X_t(\theta)}{\partial\theta^T}-\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta^T}\right)\right\|$$

$$+\sup_{\theta\in\Theta}\left\|m_\alpha(\tilde{\eta}_t)\left(\frac{\partial X_t(\theta)}{\partial\theta}-\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta}\right)\left(\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta^T}-\frac{\partial X_t(\theta)}{\partial\theta^T}\right)\right\|$$

$$+\sup_{\theta\in\Theta}\left\|m_\alpha(\tilde{\eta}_t)\left(\frac{\partial X_t(\theta)}{\partial\theta}-\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta}\right)\frac{\partial X_t(\theta)}{\partial\theta^T}\right\|$$

$$\leq\sup_{\theta\in\Theta}|h_\alpha(\tilde{\eta}_t)|\sup_{\theta\in\Theta}\left\|\frac{\partial^2 X_t(\theta)}{\partial\theta\partial\theta^T}-\frac{\partial^2\widetilde{X}_t(\theta)}{\partial\theta\partial\theta^T}\right\|+\sup_{\theta\in\Theta}|h_\alpha(\eta_t)-h_\alpha(\tilde{\eta}_t)|\sup_{\theta\in\Theta}\left\|\frac{\partial^2 X_t(\theta)}{\partial\theta\partial\theta^T}\right\|$$

$$+\sup_{\theta\in\Theta}|m_\alpha(\eta_t)-m_\alpha(\tilde{\eta}_t)|\left(\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}\right\|\right)^2+2\sup_{\theta\in\Theta}|m_\alpha(\tilde{\eta}_t)|\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}\right\|\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}-\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta}\right\|$$

$$+\sup_{\theta\in\Theta}|m_\alpha(\tilde{\eta}_t)|\left(\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}-\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta}\right\|\right)^2.$$

Using Lemma 2.1 of Straumann and Mikosch [34], together with Lemma A1, **(A2)**, **(A4)**, **(A7)**, **(A8)**, **(A10)**, and Lemma 1 of Kim and Lee [22], the right-hand side of the last inequality converges to 0 a.s. as $t\to\infty$. Hence, the first part of the lemma is verified.

Similarly, we have

$$\frac{1}{(1+\alpha)^2}\sup_{\theta\in\Theta}\left\|\frac{\partial l_{\alpha,t}(\theta)}{\partial\theta}\frac{\partial l_{\alpha,t}(\theta)}{\partial\theta^T}-\frac{\partial\tilde{l}_{\alpha,t}(\theta)}{\partial\theta}\frac{\partial\tilde{l}_{\alpha,t}(\theta)}{\partial\theta^T}\right\|$$

$$\leq\sup_{\theta\in\Theta}\left\|(h_\alpha(\eta_t)^2-h_\alpha(\tilde{\eta}_t)^2)\frac{\partial X_t(\theta)}{\partial\theta}\frac{\partial X_t(\theta)}{\partial\theta^T}\right\|+\sup_{\theta\in\Theta}\left\|h_\alpha(\tilde{\eta}_t)^2\frac{\partial X_t(\theta)}{\partial\theta}\left(\frac{\partial X_t(\theta)}{\partial\theta^T}-\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta^T}\right)\right\|$$

$$+\sup_{\theta\in\Theta}\left\|h_\alpha(\tilde{\eta}_t)^2\left(\frac{\partial X_t(\theta)}{\partial\theta}-\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta}\right)\left(\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta^T}-\frac{\partial X_t(\theta)}{\partial\theta^T}\right)\right\|$$

$$+\sup_{\theta\in\Theta}\left\|h_\alpha(\tilde{\eta}_t)^2\left(\frac{\partial X_t(\theta)}{\partial\theta}-\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta}\right)\frac{\partial X_t(\theta)}{\partial\theta^T}\right\|$$

$$\leq\sup_{\theta\in\Theta}|h_\alpha(\eta_t)-h_\alpha(\tilde{\eta}_t)|\left(\sup_{\theta\in\Theta}|h_\alpha(\eta_t)|+\sup_{\theta\in\Theta}|h_\alpha(\tilde{\eta}_t)|\right)\left(\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}\right\|\right)^2$$

$$+2\sup_{\theta\in\Theta}|h_\alpha(\tilde{\eta}_t)^2|\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}\right\|\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}-\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta}\right\|$$

$$+\sup_{\theta\in\Theta}|h_\alpha(\tilde{\eta}_t)^2|\left(\sup_{\theta\in\Theta}\left\|\frac{\partial X_t(\theta)}{\partial\theta}-\frac{\partial\widetilde{X}_t(\theta)}{\partial\theta}\right\|\right)^2,$$

and the right-hand side of the last inequality also converges to 0 a.s. from Lemma 2.1 of Straumann and Mikosch [34]. Therefore, the lemma is asserted.　□

Lemma A3. *Suppose that conditions* **(A0)–(A11)** *hold. Then, under H_0, we have as $n \to \infty$,*

$$
K_\alpha^{-1/2} \frac{[ns]}{\sqrt{n}} \frac{\partial \widetilde{L}_{\alpha,[ns]}(\theta_0)}{\partial \theta} \xrightarrow{w} B_d(s),
$$

where B_d is a d-dimensional Brownian motion.

Proof. First, we show that K_α is nonsingular. Since $Var[h_\alpha(\eta_t^0)|\mathcal{F}_{t-1}] = Var[p(Y_t|\eta_t^0)^\alpha (Y_t - B(\eta_t^0))]/B'(\eta_t^0)$ $|\mathcal{F}_{t-1}] > 0$, we have $E(h_\alpha(\eta_t^0)^2|\mathcal{F}_{t-1}) > [E(h_\alpha(\eta_t^0)|\mathcal{F}_{t-1})]^2 = 0$. Hence, it holds that for $v \in \mathbb{R}^d / \{0\}$,

$$
v^T K_\alpha v = (1+\alpha)^2 E\left[h_\alpha(\eta_t^0)^2 \left(v^T \frac{\partial X_t(\theta_0)}{\partial \theta}\right)^2\right] = (1+\alpha)^2 E\left[E(h_\alpha(\eta_t^0)^2|\mathcal{F}_{t-1}) \left(v^T \frac{\partial X_t(\theta_0)}{\partial \theta}\right)^2\right] > 0,
$$

from **(A9)**, which implies that K_α is nonsingular.

Note that

$$
E\left(\frac{\partial l_{\alpha,t}(\theta_0)}{\partial \theta}\Big|\mathcal{F}_{t-1}\right) = (1+\alpha)\frac{\partial X_t(\theta_0)}{\partial \theta} E(h_\alpha(\eta_t^0)|\mathcal{F}_{t-1}) = 0,
$$

and K_α is finite from Lemma 5 of Kim and Lee [22]. Since $\partial l_{\alpha,t}(\theta_0)/\partial \theta$ is stationary and ergodic, it holds from the functional central limit theorem for martingales (cf. Section 18 in Billingsley [35]) that

$$
K_\alpha^{-1/2} \frac{[ns]}{\sqrt{n}} \frac{\partial L_{\alpha,[ns]}(\theta_0)}{\partial \theta} = K_\alpha^{-1/2} \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \frac{\partial l_{\alpha,t}(\theta_0)}{\partial \theta} \xrightarrow{w} B_d(s).
$$

Furthermore, we can show that

$$
\sup_{0 \le s \le 1} \frac{[ns]}{\sqrt{n}} \left\| \frac{\partial L_{\alpha,[ns]}(\theta_0)}{\partial \theta} - \frac{\partial \widetilde{L}_{\alpha,[ns]}(\theta_0)}{\partial \theta} \right\| \le \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\| \frac{\partial l_{\alpha,t}(\theta_0)}{\partial \theta} - \frac{\partial \widetilde{l}_{\alpha,t}(\theta_0)}{\partial \theta} \right\| = o(1) \text{ a.s.,}
$$

from Lemma 6 of Kim and Lee [22]. Hence, the lemma is verified. □

Lemma A4. *Suppose that conditions* **(A0)–(A11)** *hold. Then, under H_0, we have as $n \to \infty$,*

$$
\max_{1 \le k \le n} \frac{k}{n} \left\| \frac{\partial^2 \widetilde{L}_{\alpha,k}(\bar{\theta}_{\alpha,n,k})}{\partial \theta \partial \theta^T} + J_\alpha \right\| = o(1) \text{ a.s.,}
$$

where $\{\bar{\theta}_{\alpha,n,k}|1 \le k \le n, n \ge 1\}$ is any double array of Θ-valued random vectors satisfying $\|\bar{\theta}_{\alpha,n,k} - \theta_0\| \le \|\hat{\theta}_{\alpha,n} - \theta_0\|$.

Proof. From Lemma 5 of Kim and Lee [22], it holds that

$$
E\left(\sup_{\theta \in \Theta} \left\| \frac{\partial^2 l_{\alpha,t}(\theta)}{\partial \theta \partial \theta^T} - \frac{\partial^2 l_{\alpha,t}(\theta_0)}{\partial \theta \partial \theta^T} \right\|\right) < \infty.
$$

Since $\partial^2 l_{\alpha,t}(\theta)/\partial \theta \partial \theta^T$ is continuous in θ, for any $\epsilon > 0$, we can take a neighborhood $N_\epsilon(\theta_0)$, such that

$$
E\left(\sup_{\theta \in N_\epsilon(\theta_0)} \left\| \frac{\partial^2 l_{\alpha,t}(\theta)}{\partial \theta \partial \theta^T} - \frac{\partial^2 l_{\alpha,t}(\theta_0)}{\partial \theta \partial \theta^T} \right\|\right) < \epsilon \tag{A1}
$$

by decreasing the neighborhood to θ_0. Since $\hat{\theta}_{\alpha,n}$ converges to θ_0 a.s. by Proposition 1, we can write that for sufficiently large n,

$$
\max_{1 \leq k \leq n} \frac{k}{n} \left\| \frac{\partial^2 \widetilde{L}_{\alpha,k}(\bar{\theta}_{\alpha,n,k})}{\partial\theta\partial\theta^T} + J_\alpha \right\|
$$

$$
\leq \max_{1 \leq k \leq n} \frac{k}{n} \left\| \frac{\partial^2 \widetilde{L}_{\alpha,k}(\bar{\theta}_{\alpha,n,k})}{\partial\theta\partial\theta^T} - \frac{\partial^2 L_{\alpha,k}(\bar{\theta}_{\alpha,n,k})}{\partial\theta\partial\theta^T} \right\| + \max_{1 \leq k \leq n} \frac{k}{n} \left\| \frac{\partial^2 L_{\alpha,k}(\bar{\theta}_{\alpha,n,k})}{\partial\theta\partial\theta^T} - \frac{\partial^2 L_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T} \right\|
$$

$$
+ \max_{1 \leq k \leq n} \frac{k}{n} \left\| \frac{\partial^2 L_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T} + J_\alpha \right\|
$$

$$
\leq \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in N_\epsilon(\theta_0)} \left\| \frac{\partial^2 \tilde{l}_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 l_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} \right\| + \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in N_\epsilon(\theta_0)} \left\| \frac{\partial^2 l_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 l_{\alpha,t}(\theta_0)}{\partial\theta\partial\theta^T} \right\|
$$

$$
+ \max_{1 \leq k \leq n} \frac{k}{n} \left\| \frac{\partial^2 L_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T} + J_\alpha \right\|
$$

$$
:= I_n + II_n + III_n \quad \text{a.s.}
$$

By Lemma A2, $I_n = o(1)$ a.s. By using (A1) and the stationarity and ergodicity of $\partial^2 l_{\alpha,t}(\theta)/\partial\theta\partial\theta^T$, we have

$$
\lim_{n\to\infty} II_n = E\left(\sup_{\theta \in N_\epsilon(\theta_0)} \left\| \frac{\partial^2 l_{\alpha,t}(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 l_{\alpha,t}(\theta_0)}{\partial\theta\partial\theta^T} \right\| \right) < \epsilon \quad \text{a.s.}
$$

Finally, since $\left\| \partial^2 L_{\alpha,n}(\theta_0)/\partial\theta\partial\theta^T + J_\alpha \right\|$ converges to 0 a.s., we can show that

$$
\max_{1 \leq k \leq \sqrt{n}} \frac{k}{n} \left\| \frac{\partial^2 L_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T} + J_\alpha \right\| \leq \frac{1}{\sqrt{n}} \sup_{1 \leq k} \left\| \frac{\partial^2 L_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T} + J_\alpha \right\| = o(1) \quad \text{a.s.,}
$$

and

$$
\max_{\sqrt{n} \leq k \leq n} \frac{k}{n} \left\| \frac{\partial^2 L_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T} + J_\alpha \right\| \leq \max_{\sqrt{n} \leq k \leq n} \left\| \frac{\partial^2 L_{\alpha,k}(\theta_0)}{\partial\theta\partial\theta^T} + J_\alpha \right\| = o(1) \quad \text{a.s.,}
$$

which assert $III_n = o(1)$ a.s. Therefore, the lemma is established. □

Proof of Theorem 1. First, we show that

$$
\frac{[ns]}{\sqrt{n}} \frac{\partial \widetilde{L}_{\alpha,[ns]}(\theta_0)}{\partial\theta} + \frac{[ns]}{n} \frac{\partial^2 \widetilde{L}_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial\theta\partial\theta^T} J_\alpha^{-1} \sqrt{n} \frac{\partial \widetilde{L}_{\alpha,n}(\theta_0)}{\partial\theta} \xrightarrow{w} K_\alpha^{1/2} B_d^o(s). \tag{A2}
$$

From Lemma A3, we have

$$
\frac{[ns]}{\sqrt{n}} \frac{\partial \widetilde{L}_{\alpha,[ns]}(\theta_0)}{\partial\theta} - \frac{[ns]}{n} \sqrt{n} \frac{\partial \widetilde{L}_{\alpha,n}(\theta_0)}{\partial\theta} \xrightarrow{w} K_\alpha^{1/2} B_d^o(s).
$$

Since $\sqrt{n} \partial \widetilde{L}_{\alpha,n}(\theta_0)/\partial\theta = O_p(1)$ by Lemma A3 with $s = 1$, using Lemma A4, it holds that

$$
\sup_{0 \leq s \leq 1} \frac{[ns]}{n} \left\| \frac{\partial^2 \widetilde{L}_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial\theta\partial\theta^T} J_\alpha^{-1} \sqrt{n} \frac{\partial \widetilde{L}_{\alpha,n}(\theta_0)}{\partial\theta} + \sqrt{n} \frac{\partial \widetilde{L}_{\alpha,n}(\theta_0)}{\partial\theta} \right\|
$$

$$
\leq \left\| J_\alpha^{-1} \sqrt{n} \frac{\partial \widetilde{L}_{\alpha,n}(\theta_0)}{\partial\theta} \right\| \max_{1 \leq k \leq n} \frac{k}{n} \left\| \frac{\partial^2 \widetilde{L}_{\alpha,k}(\theta^*_{\alpha,n,k})}{\partial\theta\partial\theta^T} + J_\alpha \right\|
$$

$$
= o_p(1),
$$

where $\theta^*_{\alpha,n,k}$ denotes that corresponding to $\theta^*_{\alpha,n,s}$ when $[ns] = k$. Hence, (A2) is verified.

Next, from Lemma A4, we have

$$\sup_{0 \leq s \leq 1} \frac{[ns]}{n} \left\| \frac{\partial^2 \widetilde{L}_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial\theta\partial\theta^T} \right\| \leq \max_{1 \leq k \leq n} \frac{k}{n} \left\| \frac{\partial^2 \widetilde{L}_{\alpha,k}(\theta^*_{\alpha,n,k})}{\partial\theta\partial\theta^T} + J_\alpha \right\| + \|J_\alpha\| = O_p(1)$$

and

$$\left\| \frac{\partial^2 \widetilde{L}_{\alpha,n}(\theta^*_{\alpha,n,1})}{\partial\theta\partial\theta^T} + J_\alpha \right\| \leq \max_{1 \leq k \leq n} \frac{k}{n} \left\| \frac{\partial^2 \widetilde{L}_{\alpha,k}(\theta^*_{\alpha,n,k})}{\partial\theta\partial\theta^T} + J_\alpha \right\| = o(1) \text{ a.s.}$$

Then, since $\sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0) = O_p(1)$ by Proposition 1, we have

$$\sup_{0 \leq s \leq 1} \frac{[ns]}{n} \left\| \frac{\partial^2 \widetilde{L}_{\alpha,[ns]}(\theta^*_{\alpha,n,s})}{\partial\theta\partial\theta^T} \left(\frac{\partial^2 \widetilde{L}_{\alpha,n}(\theta^*_{\alpha,n,1})}{\partial\theta\partial\theta^T} + J_\alpha \right) \sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0) \right\| = o_p(1). \tag{A3}$$

Therefore, from (5), (A2), and (A3), the theorem is validated. □

Lemma A5. *Suppose that conditions* **(A0)–(A11)** *hold. Then, under H_0, we have as $n \to \infty$,*

$$\frac{1}{n} \sum_{t=1}^n \frac{\partial \tilde{l}_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta} \frac{\partial \tilde{l}_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta^T} \xrightarrow{a.s.} K_\alpha.$$

Proof. In a similar way to Lemma A4, from Lemma 5 of Kim and Lee [22], we can also take a neighborhood $N_\epsilon(\theta_0)$, such that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in N_\epsilon(\theta_0)} \left\| \frac{\partial l_{\alpha,t}(\theta)}{\partial\theta} \frac{\partial l_{\alpha,t}(\theta)}{\partial\theta^T} - \frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta} \frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta^T} \right\|$$

$$= E\left(\sup_{\theta \in N_\epsilon(\theta_0)} \left\| \frac{\partial l_{\alpha,t}(\theta)}{\partial\theta} \frac{\partial l_{\alpha,t}(\theta)}{\partial\theta^T} - \frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta} \frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta^T} \right\| \right) < \epsilon \text{ a.s.} \tag{A4}$$

Note that we can write

$$\left\| \frac{1}{n} \sum_{t=1}^n \frac{\partial \tilde{l}_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta} \frac{\partial \tilde{l}_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta^T} - E\left(\frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta} \frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta^T} \right) \right\|$$

$$\leq \left\| \frac{1}{n} \sum_{t=1}^n \frac{\partial \tilde{l}_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta} \frac{\partial \tilde{l}_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta^T} - \frac{1}{n} \sum_{t=1}^n \frac{\partial l_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta} \frac{\partial l_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta^T} \right\|$$

$$+ \left\| \frac{1}{n} \sum_{t=1}^n \frac{\partial l_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta} \frac{\partial l_{\alpha,t}(\hat{\theta}_{\alpha,n})}{\partial\theta^T} - \frac{1}{n} \sum_{t=1}^n \frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta} \frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta^T} \right\|$$

$$+ \left\| \frac{1}{n} \sum_{t=1}^n \frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta} \frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta^T} - E\left(\frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta} \frac{\partial l_{\alpha,t}(\theta_0)}{\partial\theta^T} \right) \right\|$$

$$:= I_n + II_n + III_n.$$

By Lemma A2,

$$I_n \leq \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \Theta} \left\| \frac{\partial \tilde{l}_{\alpha,t}(\theta)}{\partial\theta} \frac{\partial \tilde{l}_{\alpha,t}(\theta)}{\partial\theta^T} - \frac{\partial l_{\alpha,t}(\theta)}{\partial\theta} \frac{\partial l_{\alpha,t}(\theta)}{\partial\theta^T} \right\| = o(1) \text{ a.s.}$$

Since $\hat{\theta}_{\alpha,n}$ converges to θ_0 a.s. by Proposition 1, from (A4), we have

$$\lim_{n \to \infty} II_n \leq \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in N_\epsilon(\theta_0)} \left\| \frac{\partial l_{\alpha,t}(\theta)}{\partial \theta} \frac{\partial l_{\alpha,t}(\theta)}{\partial \theta^T} - \frac{\partial l_{\alpha,t}(\theta_0)}{\partial \theta} \frac{\partial l_{\alpha,t}(\theta_0)}{\partial \theta^T} \right\| < \epsilon \quad \text{a.s.}$$

Finally, by the ergodic theorem, $III_n = o(1)$ a.s. Therefore, the lemma is established. $\square$

References

1. McKenzie, E. Some simple models for discrete variate time series. *J. Am. Water Resour. Assoc.* **1985**, *21*, 645–650. [CrossRef]
2. Al-Osh, M.A.; Alzaid, A.A. First order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **1987**, *8*, 261–275. [CrossRef]
3. Ferland, R.; Latour, A.; Oraichi, D. Integer-valued GARCH processes. *J. Time Ser. Anal.* **2006**, *27*, 923–942. [CrossRef]
4. Engle, R.F. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **1982**, *50*, 987–1007. [CrossRef]
5. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econom.* **1986**, *31*, 307–327. [CrossRef]
6. Fokianos, K.; Rahbek, A.; Tjøstheim, D. Poisson autoregression. *J. Am. Stat. Assoc.* **2009**, *104*, 1430–1439. [CrossRef]
7. Davis, R.A.; Wu, R. A negative binomial model for time series of counts. *Biometrika* **2009**, *96*, 735–749. [CrossRef]
8. Christou, V.; Fokianos, K. Quasi-likelihood inference for negative binomial time series models. *J. Time Ser. Anal.* **2014**, *35*, 55–78. [CrossRef]
9. Zhu, F. Modeling overdispersed or underdispersed count data with generalized poisson integer-valued garch models. *J. Math. Anal. Appl.* **2012**, *389*, 58–71. [CrossRef]
10. Zhu, F. Zero-inflated Poisson and negative binomial integer-valued GARCH models. *J. Stat. Plan. Infer.* **2012**, *142*, 826–839. [CrossRef]
11. Lee, S.; Lee, Y.; Chen, C.W.S. Parameter change test for zero-inflated generalized Poisson autoregressive models. *Statistics* **2016**, *50*, 540–557. [CrossRef]
12. Davis, R.A.; Liu, H. Theory and inference for a class of observation-driven models with application to time series of counts. *Stat. Sin.* **2016**, *26*, 1673–1707.
13. Diop, M.L.; Kengne, W. Testing parameter change in general integer-valued time series. *J. Time Ser. Anal.* **2017**, *38*, 880–894. [CrossRef]
14. Lee, Y.; Lee, S. CUSUM test for general nonlinear integer-valued GARCH models: Comparison study. *Ann. Inst. Stat. Math.* **2019**, *71*, 1033–1057. [CrossRef]
15. Kang, J.; Lee, S. Parameter change test for Poisson autoregressive models. *Scand. J. Stat.* **2014**, *41*, 1136–1152. [CrossRef]
16. Basu, A.; Harris, I.R.; Hjort, N.L.; Jones, M.C. Robust and efficient estimation by minimizing a density power divergence. *Biometrika* **1998**, *85*, 549–559. [CrossRef]
17. Ghosh, A.; Basu, A. Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electron. J. Stat.* **2013**, *7*, 2420–2456. [CrossRef]
18. Lee, S.; Song, J. Minimum density power divergence estimator for GARCH models. *Test* **2009**, *18*, 316–341. [CrossRef]
19. Kim, B.; Lee, S. Robust estimation for the covariance matrix of multivariate time series based on normal mixtures. *Comput. Stat. Data Anal.* **2013**, *57*, 125–140. [CrossRef]
20. Kang, J.; Lee, S. Minimum density power divergence estimator for Poisson autoregressive models. *Comput. Stat. Data Anal.* **2014**, *80*, 44–56. [CrossRef]
21. Kim, B.; Lee, S. Robust estimation for zero-inflated Poisson autoregressive models based on density power divergence. *J. Stat. Comput. Simul.* **2017**, *87*, 2981–2996. [CrossRef]
22. Kim, B.; Lee, S. Robust estimation for general integer-valued time series models. *Ann. Inst. Stat. Math.* **2020**, in press.

23. Kang, J.; Song, J. Robust parameter change test for Poisson autoregressive models. *Stat. Probab. Lett.* **2015**, *104*, 14–21. [CrossRef]

24. Song, J.; Kang, J. Test for parameter change in the presence of outliers: The density power divergence based approach. *arXiv* **2019**, arXiv:1907.00004.

25. Kang, J.; Song, J. A robust approach for testing parameter change in Poisson autoregressive models. *arXiv* **2019**, arXiv:1908.11466.

26. Batsidis, A.; Horváth, L.; Martín, N.; Pardo, L.; Zografos, K. Change-point detection in multinomial data using phi-divergence test statistics. *J. Multivar. Anal.* **2013**, *118*, 53–66. [CrossRef]

27. Batsidis, A.; Martín, N.; Pardo, L.; Zografos, K. ϕ-divergence based procedure for parametric change point problems. *Methodol. Comput. Appl. Probab.* **2016**, *18*, 21–35. [CrossRef]

28. Martín, N.; Pardo, L. Comment on: Extensions of some classical methods in change point analysis. *Test* **2014**, *23*, 279–282. [CrossRef]

29. Lehmann, E.; Casella, G. *Theory of Point Estimation*, 2nd ed.; Springer: New York, NY, USA, 1998.

30. Warwick, J. A data-based method for selecting tuning parameters in minimum distance estimators. *Comput. Stat. Data Anal.* **2005**, *48*, 571–585. [CrossRef]

31. Warwick, J.; Jones, M.C. Choosing a robustness tuning parameter. *J. Stat. Comput. Simul.* **2005**, *75*, 581–588. [CrossRef]

32. Lee, S.; Ha, J.; Na, O.; Na, S. The cusum test for parameter change in time series models. *Scand. J. Stat.* **2003**, *30*, 781–796. [CrossRef]

33. Fried, R.; Agueusop, I.; Bornkamp, B.; Fokianos, K.; Fruth, J.; Ickstadt, K. Retrospective Bayesian outlier detection in INGARCH series. *Stat. Comput.* **2015**, *25*, 365–374. [CrossRef]

34. Straumann, D.; Mikosch, T. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Ann. Stat.* **2006**, *34*, 2449–2495. [CrossRef]

35. Billingsley, P. *Convergence of Probability Measures*, 2nd ed.; Wiley: New York, NY, USA, 1999.